GOOGLE TRANSLATE ACCURACY EVALUATION

Darija Lunić[1]*,
Nataša Stanišić[1],
Angelina Njeguš[2],
Ivana Đerić[1]

[1]Faculty of Business,
 Singidunum University,
 Belgrade, Serbia

[2]Faculty of Informatics and Computing,
 Singidunum University,
 Belgrade, Serbia

Abstract:

Google Translate (GT), as the most used language translation application, has been significantly improved when artificial neural network (ANN) based Neural Machine Translate System was introduced. The purpose of this paper is to evaluate the translation accuracy of the GT system when used for education purposes. The text in the English language is translated into two languages, by professional human translators and by GT. It is IT related text and refers to computer networks. Students' comprehension of the given text is evaluated regarding language fluency, adequacy, meaning and severity.

Keywords:

Google Translate (GT), machine translate (MT), accuracy, human interpreter, education.

Correspondence:

Darija Lunić

e-mail:
dlunic@singidunum.ac.rs

1. INTRODUCTION

Automatic translation to designated languages offered to website users is the trend that is visible for the last couple of years. Google Translate (GT) is one of the most frequently used tools of the MT system. The services provided by GT are free of charge, and in line with official data from Google, GT is used by over 500 million users [1]. Due to its necessity and rapid progress in the development of machine translation, it is expected to process not only texts but also audio and video files.

GT architecture has changed and improved over time (Figure 1). At the beginning, MT was rule based (*Rule-based Machine Translation*) which worked on the morphology, syntax and semantic of both languages. Then, *Statistical Machine Translation* (*SMT*) was developed, based on two statistical models, language model and translation model.
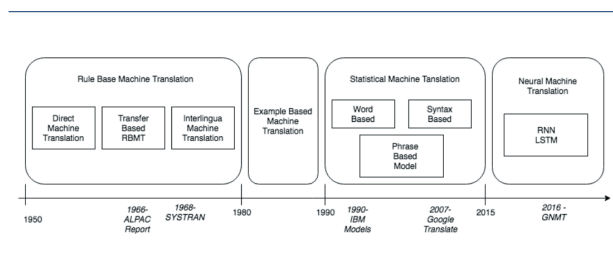
Fig. 1. GT architecture development over time [2]

One of the flaws of this architecture is that massive parallel corpuses are created. Today, GT architecture is based on artificial neural networks, thus the model is called Neural Machine Translation (NMT). This model uses Recurrent Neural Networks algorithm. This Google NMT model has some limitations, namely: lexical, format, expertise and accuracy [3].

Since Google NMT (GNMT) algorithm has been recently introduced, the question that arises is GT accuracy. Khoong et al. [4] have assessed the use of GT to translate emergency department discharge instructions into Spanish and Chinese. Results showed that 92% sentences were accurately translated into Spanish, and 81% into Chinese. Aiken in 2019 [5] showed that translations between English and German, African, Portuguese, Spanish, Danish, Greek, Polish, Hungarian, Finnish and Chinese tend to be the most accurate.

Yamada [6] in his study in 2019, investigated college students' post-editing potential when using Google SMT, and Google NMT of the same source text. NMT showed better results in terms of its final product, which contained fewer errors. However, there were no significant difference in cognitive effort or error correction rate. Guldal and Isisag [7] analysed the translation errors of GT outputs conducted from Turkish to English. The errors were classified into: Lexial Errors, Morphological, Syntactic, Semantic and Pragmatic Errors. The results show that eventhough GT provides faster translations, there is still a need for human assistance.

Jabak [8] conducted the assessment of Arabic-English translation produced by GT. Results showed that GT made lexical and syntactic errors, therefore GT cannot be used as a valid translation tool for Arabic-English translation. Rensburg et al. [9], assessed GT in a South African higher education environments. One document was translated from African into English and another form English into African. The results showed that a user would not have to spend much time correcting translations.

Chen et al. [10] evaluated the accuracy of GT when translating health information from English to Spanish, and English to Chinese. Results showed that Google provided accurate translation for simple sentences.

The goal of is this paper is to evaluate to what extent the quality of translation of Google Translate is satisfactory in order to be applicable in learning professional language, in our case, in the area of IT.

The authors took the text from Wikipedia on computer networks that was originally in English, that was subsequently translated to Italian and German by human translator and GT. The interpreter evaluates the text translated by GT on the basis of Trujillo scale which includes intelligibility and accuracy. The intelligibility scale is a is five-degree, and the accuracy scale is a seven-degree scale. Further research will include students who will be given both translations, based on which they will do a task regarding reading comprehension.

Based on the questions about the text responded by students we will measure the level of understanding of one and the other text, based on which it will be concluded whether GT can be used in learning foreign languages in the field of informatics and computer science.

It is assumed that the level of comprehension of the text translated with the use of Google Translate will be surprisingly high and that it will not deviate significantly from the text translated by human interpreter. The authors of this paper rely on research confirming multiple utility of Google Translate, especially in relation to costs compared to a professional translator [11].

It is expected that despite a significant number of errors regarding syntax and style, the level of understanding of out text will be surprisingly high. It is assumed this is due to the structure of the sentences of special purpose language.

1. MACHINE TRANSLATION

Machine translation (MT) entails automatic translation using a computer program that is designed to translate text from one language (source) into another (target) language without the aid of a human [12] [13].

MT dates since the World War II when a connection was noticed between the translation and cryptography during decoding of the messages in computers [14]. During the 60ies, MT was criticized due to its inaccuracy and slow performance, but in the late sixties it became success, especially in Canada where it is used today in translating weather reports [14].

With the onset of World Wide Web, MT blooms later on, when it becomes available online as Babel Fish [15]. After this period a number of tools emerged for machine translation, including Google Translate that is easily available to users and free of charge.

Google explains that according to the previous translated text by human translators, the GT looks for patterns in millions of documents and makes intelligent guesses to create an appropriate translation [16].

Google launched Google Translate in 2009 enabling users of webites to translate, instantly, the content of the website into more than sixty languages. A new feature for the Website Translator was launched in 2012. In this way, the website administrators were enabled to edit translations and users were allowed to propose a better translation. The suggested translation could have been accepted or rejected by the web administrator [17]. With the aim of improving the quality of the translation it is possible to post-edit the MT output. Even though it facilitates the process of translation, a number of studies show that MT is efficient in case when it is corrected by the human factor [18].

Traditional phrase-based translation systems have executed their tasks by breaking the source sentences to parts, and then, translating them phrase by phrase. This manner of translation is nothing similar to human translation. A man would read the source sentence first, understand it, and then translate. This is the manner of translation imitated by Neural Machine Translation, the NMT.

In 2016, Google implemented the Google Neural Machine Translation model (GNMT). This GNMT system loads the source sentence first in order to produce the vector of "thought", that is, the sequence of numbers representing the meaning of the sentence and the decoder that is processing the vector of the sentence and translates it to target language. GNMT is based on encoder-decoder architecture of the system using Recurrent Neural Network (RNN) (Fig. 2).
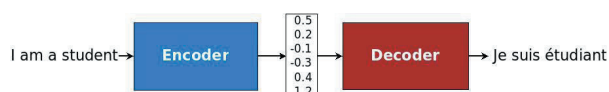

Fig. 2. GNMT system encoder-decoder architecture [19].

Based on artificial neural network, it was supposed to improve the quality of translation to great extent. GNMT (Neural Machine Translation System) has improved the quality of the two most commonly used language pairs – Spanish-English and French-English. The accuracy of translation improved by 85%. In 2017, Google launched a survey among GT users. They are asked to evaluate three translating options: machine statistical, neural and human. The results were a surprise – it turned out that the translation relying on neural networks was almost perfect in certain language pairs [16].

2. RESEARCH METHODOLOGY

Choice of sentences

It was decided to focus on IT related text. The authors selected the text from Wikipedia on computer networks. The original English text contained 10 sentences translated by the translator and GT. An example of chosen sentence in its original form and translated by a professional translator and GT are shown in Table 1.

Table 1. An example of chosen sentence and its translations

Original sentence	„A computer network or data network is a telecommunications network which allows computers to exchange data"
Translated by professional translator in German	„Ein Computernetz oder Datennetz ist ein Telekommunikationsnetz, das Computerdaten austauschen können."
Translated by GT in German	"Ein Computernetzwerk oder Datennetzwerk ist ein Telekommuni-kationsnetzwerk, mit dem Computer Daten austauschen können."
Translated by professional translator in Italian	"Una rete di computer o rete di informazioni è una rete di telecomunicazioni che permette lo scambio di dati informativi."
Translated by GT in Italian	"Una rete di computer o una rete di dati è una rete di telecomunicazioni che consente ai computer di scambiare dati."

Language selection

Even though GT offers a wide spectrum of translations from English to other languages the authors of this paper focused on two languages, Italian and German due to a large number of students of the faculty for IT that select these languages.

Method for Machine Tanslation Evaluation

A common evaluation technique is used, where four expert translators gave ratings based on Trujillo's scale of intelligibility as well as accuracy. Research steps are drawn in the figure 3.
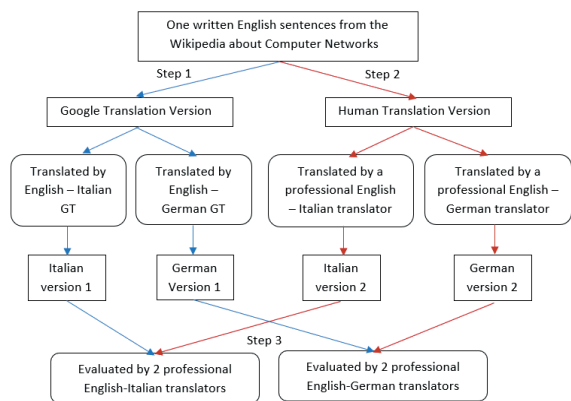

Fig. 3. Research steps

When evaluating the text translated using Google Translation the human factor is the only valid one to judge the quality of the text translated. Translated text is evaluated based on intelligibility and accuracy, using appropriate scales [20], [21], [12,13].

Intelligibility is the level of fluency and grammar accuracy of the text translated using the MT system, as well as the text translated by the human factor [20]. Also, it represents a level of comprehension of the text [12,13].

It was noticed that the text translated by GT did not decrease in the quality on the scale of intelligibility as well as accuracy, but there were changes in the style and syntax of the sentence.

After the examination in the form of a test based on the translated text, which will include students in terms of measuring the intelligibility and accuracy of both translations, it will be concluded which text has a higher score. The further step will be to examine which language pair has a greater degree of intelligibility and accuracy. Future research will be based on expanding the number of languages (including Spanish, Russian, French and Arabic), on the basis of which of the mentioned languages has the largest database and thus the largest score.

The test contains 15 multiple-choice questions. An example of the test questions in German and Italian language is shown in Table 2.

Table 2. An example of the test questions in both languages

Test question in German	Ein Computernetz oder Datennetz ist ein Telekommunikationsnetz, das Computer ermöglicht: _____.” a) Datenaustausch b) Verbindungen zwischen den Knoten c) Kommunikationsprotokol
Test question in Italian	Una rete di computer o rete di informazioni è una rete di telecomunicazioni che permette ai computer _____.” a) lo scambio di dati informativi b) nodi di rete c) protocollo di comunicazione

Intelligibility

The term intelligibility includes clarity, fluency and sometimes, ease of reading [16]. Regarding intelligibility the style may or may not be considered [21]. Intelligibility is an important factor of evaluating translation quality due to the fact that even when a text is accurately translated but the spirit of the target language is neglected, certain level of intelligibility is lost [1]. In order to avoid subjectivity when evaluating a translation by one of the assessors, objective evaluation may be received using statistical methods, through the combination of assessments of several different assessors [20]. Arnold et el. suggest at least four assessors and point out they should be related to the area [21].

Accuracy

Accuracy or "adherence" to the original text is a measure determining the level to which the translated text retains the meaning of the original text [20]. In order to receive a broader image of the translation quality, intelligibility and accuracy must be observed in pair. After the measured level of intelligibility, the accuracy must be measured [21]. When assessing accuracy, it is important that the assessors have necessary language skills and knowledge. If the issue of objectivity is considered, since some of the assessors are stricter than others, including a large number of assessors will result in a clearer image of accuracy level. Additionally, this may bring front an issue of one assessor understanding a sentence in different context than the others.

Scale selection

In order to evaluate the comprehension of MT system results, a number of scales has been developed. Intelligibility scales range from two to five level ones [21].

In these scales the sentences that are understood the best get the highest marks, and vice-versa. One of the renowned scales is the Trujillo five-point scale that includes clear, unequivocal descriptions for each scale point [20]. The scale measures how much an examiner agrees or disagrees with a specific statement. Scale points measure grammar, word usage and style in term of whether rewriting is needed or not. If there are some problems regarding scale points the examiner needs to look at the original sentences to interpet the meaning. The final score will depend on examiner's grades regarding abovementioned scale points [20] [1].

Trujillo's seven-point intelligibility scale [20] [1] measures the degree to which a native speaker understands the sentence and the degree to which the content of the source language sentence is clearly passed on. Word order, the use of tenses, subjects and predicates relationship are also graded by the examiner. The final score depends on the overall understanding by a native speaker and examiner.

Profile of evaluators

According to Arnold et al. [21] it is necessary to include at least four examiners in the examination. Thus, we selected four professional translator. Each of them had native command and knowledge of the target language. Also, they were required to have excellent language skills in order to understand the given instructions and the original English text in order to determine the intelligibility and accuracy scores. The research included expert translators who worked as foreign language lecturers at Singidunum University.

Before starting research, the examiners received detailed instructions regarding the intelligibility and accuracy of the translation, as well as explanations regarding the scales and the method of assessment. In order to achieve objectivity regarding the evaluation of the translated text, the examiners had a day off between the evaluation of the text translated by GT and the text translated by the translator. It was also required not to discuss or express an opinion regarding translations.

For the purpose of this paper the authors evaluated the intelligibility and accuracy of translations based on evaluations conducted by interpreters.

3. CONCLUSION

Based on the text taken from Wikipedia that has been translated from English to Italian and German with the aid of Google Translate and professional translators, multiple choice questions were prepared. We will determine a level of comprehension of both texts based on these responses. The expected result is that students understand the text translated with Google Translate better, regardless of whether it is German or Italian. This may be explained by the fact the text is from the IT area and the level of intelligibility is higher due to sentence structure and syntax. Technical language is, in its structure and style, significantly different that literary language that demands wider context when it comes to translation, so these results are expected.

Research will be performed for other languages (French, Russian, Spanish and Arabic) and we will measure the level of intelligibility and comprehension regarding language pairs.

Since translation by translators requires certain amount of time, knowledge, skills and effort; time and resource saving concerns are increasingly significant, hence, the use of MT expects to grow in a very short period of time. Regarding the development of MT, it is very likely that the grammatical accuracy of translations will be improved, especially in language pairs mostly used. If we consider increase in use of advance technologies in foreign language studies, it is up to the academic community to accept or reject such form of cooperation. MT will surely not replace the adoption of a foreign language in the close future but the method of communication using advance technologies will be of great assistance to the teachers and students alike. The human factor is of the most importance when it comes to skills of inter-cultural, inter-personal communication that will be difficult to master using technology.

REFERENCES

[1] A. Tobin, "Is Google Translate Good Enough for Commercial Websites?: A Machine Translation evaluation of text from English websites into four different languages," Computer Science, pp. 94-116, January 2015.

[2] C. Lungu, (2020) Tutorial on machine translation. Available at: http://www.clungu.com/machine%20translation/tutorial/Tutorial-on-Machine-Translation/ (13/09/2020).

[3] A. Balaoiu, (2020) Neural Machine Translation: an insight. Ad Verbum. Available at: https://www.ad-verbum.com/en/blog/nmt (13/09/2020).

[4] E.C. Khoong, E.Steinrook, C. Brown, A. Fernandez, "Assesing the Use of Google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions", JAMA Internal Medicine, 179(4): 580-582, 2019.

[5] M. Aiken, "An Updated Evaluation of Google Translate Accuracy". Studies in Linguistics and Literature, Vol. 3, No.3, 2019.

[6] M. Yamada, "The impact of Google Neural Machine Translation on Post-editing by student translators", Journal of Specialised Translation, 2020.

[7] B.K.Guldal i K.U.Isisag, "A comparative study on google translate: An error analysis of Turkish-to English translations in terms of the text typology of Katherina Reiss. RumeliDE Dil ve Edebiyat Araştırmaları Dergisi.

[8] O.O.Jabak, "Assessment of Arabic-English translation produced by Google Translate", International Journal of Linguistics, Literature and Translation (IJLLT), Vol. 2, Iss. 4, 2019. Available at: https://publication.ijllt.org/media/290334-assessment-of-arabic-english-translation-581a8141.pdf (18/09/2020)

[9] A.V. Rensburg, C. Snyman, and S. Lotz, "Applying Google Translate in a higher education environment: Translation products assessed. Journal Southern African Linguistics and Applied Language Studies, Vol. 30, Iss.4, 2012.

[10] X. Chen, S. Acosta, A.E. Barry, "Evaluating the Accuracy of Google Translate for Diabetes Education Material". JMIR Diabetes, Vol. 1, Iss. 1, e3, 2016.

[11] F. Sheppard, "Medical writing in English: The problem with Google Translate," La Presse Médicale, 40(6), pp. 565-566, 2011.

[12] W.J. Hutchins and H.L. Somers, An Introduction to Machine Translation. London: Academic Press, 1992.

[13] W.J.Hutchins, H.L.Somers (1992) An Introduction to Machine Translation. Academic Press, Cambridge, GB.

[14] P. Koehn, Statistical Machine translation, Cambridge: Cambridge University Press, 2010, pp.15

[15] J. Yang and E. Lange, "Going live on the internet," in Computers and Translation: A Translator's Guide, H. Somers, Eds. Amsterdam-Philadelphia: John Benjamins, 2003, pp. 191-200.

[16] http://www.isi.edu/natural-language/mteval/ [Accessed 10 September 2020]

[17] J. Chin, 2012. Now you can polish up Google's translation of your website. The Official Google Translate Blog [Online], 30 May. Available from:http://googletranslate.blogspot.jp/2012/05/now-you-can-polish-up-googles.html [Accessed 7 September 2020]

[18] M. Groves, and K. Mundt, "Friend or foe? Google Translate in language for academic purposes." English for Specific Purposes, 37, pp. 112-121, 2015.

[19] T. Luong, E. Bredo, R. Zhao (2019) Neural Machine Translation (seq2seq) Tutorial. Readme.md. available at: https://github.com/tensorflow/nmt (13/09/2020).

[20] A. Trujillo, Translation Engines: Techniques for Machine Translation. London: Springer-Verlag, 1999

[21] D. Arnold, L. Balkan, R. Humphreys, S. Meijer and L. Sadler, Machine Translation: An Introductory Guide. Oxford: NCC Blackwell, 1994