EVALUATING MACHINE TRANSLATION QUALITY: A CASE STUDY OF A TRANSLATION OF A VERBATIM TRANSCRIPTION FROM SLOVAK INTO GERMAN

Jozef Štefčík

Department of Translation Studies
Faculty of Arts
Constantine the Philosopher University in Nitra, Slovakia
jstefcik@ukf.sk

The paper addresses the issue of using online statistical machine translation tools for the translation of specific text types and the problems of their translatability when using automated translation systems. An attempt is made to analyze and evaluate the machine translation of a verbatim transcription from Slovak into German.

WHY IS MACHINE TRANSLATION BECOMING MORE IMPORTANT?

Over the last decade, machine translation (MT) has become an important research topic in both the academic sphere and commercial sector. Translation is essential for international trade, not only because of the need to translate commercial contracts and legislation, but also for the use of instructions and manuals for imported products. Another aspect of the commercial interest in MT involves the cost of human translation and translation services, particularly in view of the fact that translation is highly demanding work that requires an exceptional degree of professionalism and considerable amount of time. A translator can usually translate no more than 9–10 pages a day (depending on the subject and complexity), which sometimes leads to delays, for example, in marketing a new product. This is also a major issue for EU institutions owing to the need to translate legislation into the 24 official languages of the European Union, along with numerous other documents. The cost and speed of translation are major factors in determining the social, political and economic importance of MT in many other areas of human activity.

For these reasons, IT specialists and researchers are continuing to develop MT from common online systems (Google Translate, Bing Translate, etc.) through to the sophisticated systems designed for the specific needs of companies (e.g. DGT uses the Moses system). There is also a philosophical dimension to MT in the endeavour to

automate human thinking and linguistic production. "From an academic perspective, MT is interesting because it allows the application and testing various hypotheses in linguistics, computer science and artificial intelligence" (Munková 2013).

Thus, MT has achieved a considerable advantage in speed and quality in recent years. Its usage is changing the status of professional translators and creating a new role for them as pre-editors or post-editors of MT. The very notion of translation is changing. Currently, however, MT is used with caution where a high degree of quality is required, but it is also a recognized fact that certain text types allow MT to be more successfully applied than others.

TEXT ANALYSIS AND TEXT TYPES FOR THE PURPOSES OF MT

Most text types are more or less hybrid in form. Translators must, therefore, decide whether the relevant source text can be machine-translated or not. Trained and qualified translators have an advantage over untrained translators because they can perform macro-and micro-stylistic translation-relevant text analysis faster and more easily, which enables them to overcome time constraints. They can quickly decide which texts or text segments can be translated by people or by MT, or in which segments the re-edited text would require only minimal changes.

Some text types have proven to be relatively reliable when translated by MT: technical, legal, marketing and management, tourism and catering, manuals, instructions, EU directives, regulations, insurance contracts and the like. For this reason we believe it would be most useful to analyse the following text types in the most frequent language pairs and how MT renders them: manuals, reference books, scientific reports, records, certificates, balance sheets and other financial statements.

Here is a brief example of a machine-translated text into English:

1) *In large organizations dealing with large numbers of customers it are essential for the effective operation of various departments and business processes did the latest customer information is available". (In: "Managing Customer data, 2007 Global Industries")*

The translation contains mistakes, but the content is understandable, and a translator can easily post-edit such a text. The German variant of the MT of the same text via Google Translate (analytical language) is as follows:

2) *In großen Organisationen, die sich mit einer großen Anzahl von Kunden _____, ist es wichtig für das reibungslose Funktionieren der verschiedenen Abteilungen und Geschäftsprozesse, dass die neuesten Informationen über die Kunden zur Verfügung steht. "*

A few words are missing in this translation: the verb *to deal with* and the adjective *available* are not translated at all (… *von Kunden <u>beschäftigen</u>* (…) *zur Verfügung <u>stehen</u>* accordingly), but after post-editing the text would be understandable. One might expect even better results if, prior to translation, each text were subjected to a text analysis based on the following points (Byrne 2012, 90):

1) topic;
2) text category;
3) text function;
4) target audience;
5) purpose of the text (how will the text be used);
6) distinguishing features.
7) potential problems in translation

Translation-relevant text analysis gives us more information about the difficulties to be expected when translating and the external tools that can help translators to produce better quality within a limited timeframe. After textual macro-analysis, the translator might consider microanalysis, which can be based on several steps. Arnold (1994) proposed three steps in the process of MT:

1) Preparation of input (pre-editing): intra-linguistic transfer. The translator makes the source text "translatable" by simplifying it on a linguistic level;
2) Translation using a translation system – an inter-linguistic transfer. This is a typical translation process, which consists of three common steps: analysis, transfer, synthesis.
3) Revision of translations – an intra-linguistic transfer. In this phase, the post-editing, revising and proofreading of the translated text ("raw translation") are important.

These steps of pre-editing and post-editing are shown in Table 1 (intra-language translation) and Table 2 (inter-language translation) below.

Table 1. A sample of pre-editing of the text for MT (Kit, Pan, Webster 2002, 44).

Source text	Pre-Editing of the Source text
Let the water run hot at the sink and then pull the connector from the recess in the back of the dishwasher. Upon the completion of the above task, lift the connector to the faucet by pressing down the thumb release.	1. *Turn on the faucet at the sink until the water runs hot.* 2. *Pull the connector from the recess in the back of the dishwasher.* 3. *Press down on the thumb release and lift the connector onto the faucet.*

Table 2. A sample of post-editing of the same text after MT translation into German.

Target text – Translation into German	Target text of the pre-edited version
Lassen Sie das Wasser laufen heiß in der Wanne und dann ziehen Sie den Stecker aus der Aussparung in der Rückseite der Spülmaschine. Nach Abschluss der oben genannten Aufgabe, heben Sie den Anschluss an den Wasserhahn durch Drücken des Daumen Release.	*1. Drehen Sie den Wasserhahn an der Spüle, bis das Wasser heiß ist.* *2. Ziehen Sie den Stecker aus der Aussparung in der Rückseite der Spülmaschine.* *3. Drücken Sie auf den Daumen Release und heben Sie den Stecker auf den Wasserhahn.*

It is clear that the pre-edited text provides a better output. It could be edited much faster than the text on the left, which has a more complicated and longer sentence structure.

TRANSLATION OF A SPECIFIC TEXT TYPE: WRITTEN RECORDS

Records belong to a commonly used text type. Written records (e.g. transcripts, minutes, protocols, and all kinds of written documents containing factual material) document reported speech from events that have already taken place (with the exception of a memory record) and rely on descriptions and observations. There are different types of written records:

- Based on results (streamlined, structured summaries that keep essential and non-essential information apart),
- Based on history (chronological, most realistic representation of the time sequence, the essential and the non-essential equally documented),
- Based on purpose (retrieved from memory or from recordings).

Each written record should be accurate (precise language), objective (no personal opinions), non-judgmental (no observations or judgments), and positive (FLT).

The source text of our written record selected for analysis was stylistically dense and contained administrative, narrative and technical features; the goal was to describe what was discussed in the communication act objectively and in detail. The analysed source text of the record had several functions – it reported, instructed, and advocated at the same time. The text contained many verbs and subordinate clauses. The text also had many time references, names and technical terms that referred to specific areas and operations.

The main problems in translating any written record is the information within the text itself, represented by second or third parties, which may prove difficult to decode. The authors of the records ‑ tend to use originally reformulated phrases from

speeches that were recorded on audio and transcribed without the necessary attention to context. Since external translators often lack the relevant context, they are mostly unable to get the message across. We therefore want to determine the extent to which the studied hybrid text can generate linguistically appropriate output when prepared for revision by using the online system of statistical MT, *Google Translate*.

TRANSLATION QUALITY ASSESSMENT

There have been many discussions over the last decade about the evaluation of translation quality, or Translation Quality Assessment (hereinafter referred to as TQA), which is equally important for both the professional quality control of content and for study programmes that prepare would-be translators for the profession. Although there is a widely acknowledged need to establish the general criteria for assessing the quality of translations or to come up with a definition of what might be considered "good, satisfactory or accepted" translation, there is still no universal definition of translation quality or even generally accepted methods of evaluation. Although there are national and international standards of translation (ATA, Sical, etc.), they are not widely accepted as objective criteria for assessing the quality of translation. Basically, when assessing the quality of a translated text, we tend to stick to the following criteria:

- linguistic correctness,
- fidelity to the source text,
- readability of the target text,
- equivalence,
- transfer of the meaning (appropriateness, shifts…) (Ackaert et al. 2013).

These general criteria may be categorized into several subgroups or attributed to the following translation-errors, according to which professional translators and translation instructors may assess translated texts: *omissions, negative shifts of meaning, register, punctuation, spelling, grammar, style and vocabulary* (cf. also Vilar et al. 2006).

These are the basic aspects to consider in evaluating translations. However, translations cannot be restricted to the learning environment, or to one purpose or target group. We may generally claim that translation quality should be perceived from the point of view of its recipients, their needs, and their knowledge of the subject matter, etc. This could be regarded as a final text approach based on human or manual evaluation. On the other hand, the evaluation of MT requires a different approach, since the output (translated text) cannot be regarded as a final text in need of revision. The evaluation of MT can be done manually or with the help of certain software, i.e. it may be performed automatically.

EVALUATING MT

The evaluation of MT plays a key role in the field of MT and there have certainly been many attempts to evaluate it (Hutchins and Somers 1992, Arnold et al. 1994, Callison-Burch et al. 2006, to name but a few). In order to make evaluation more efficient, experts have begun to think about automatic evaluation methods without human intervention. There have been projected series of automated quality assessment metrics to achieve the effectiveness in the evaluation process. Automatic evaluation metrics were used for additional human evaluation, while providing high efficiency and consistency at a relatively low cost. Most of these are based on measuring similarities between automated translations and referential–human–translation. Automatic evaluation metrics can be based on statistical principles (n-grams or editing distance) or on deep linguistic structures (morphological, syntactic or semantic information).

The evaluation of MT or of MT systems is an essential area of research not only in an attempt to determine the effectiveness of existing MT systems, but also to optimize their performance. Progress in MT relies on the quality of newly developed MT systems, the goal of their evaluation being to demonstrate a greater effectiveness than that of existing systems. However, here we stumble on the question of translation quality (texts generated by MT systems), and more exact methods of quality evaluation criteria. Is it possible to claim that this particular translation is the only correct translation of the original and there is no other correct translation? How to evaluate the quality of two translations that are not identical but both represent the original? Or two translations that are only partially correct? It is a very difficult task which is affected by several factors. It primarily depends on the recipient (for whom is the evaluation of MT performed?) and its further use (for what purpose is the evaluation of MT used?). Style in translation can be crucial in some applications and irrelevant in others. Therefore we avoid using the term "good" translation, but rather "appropriate" or "adequate" translation.

The evaluation of MT can be approached in two ways: *Glass Box evaluation* or *Black Box evaluation*. Glass Box measures the quality of the evaluation system based on system characteristics. It focuses on the linguistic coverage of the system and the theory used in natural language processing. This type of evaluation would be relevant for scientists and developers. Scientists would need it in terms of confirming or rejecting hypotheses. Developers attempt to figure out if the MT system works the same way it was projected to in order to determine its limits. Black Box measures the

quality evaluation system based on the generated translation. This type of evaluation is intended for recipients and translators. Translators need to know whether the use of the system will improve their productivity from the point of view of quantity and post-editing. The recipient is interested in the cost, speed and readability of the translation.

This is why we also focus on the Black Box evaluation in our study, which uses internal (intrinsic) and external (extrinsic) methods to evaluate the accuracy and applicability of MT. Internal methods (evaluation scales, trial order, error analysis, etc.) subjectively assess the quality of MT based on the comparison (hypothesis) with a referential translation which is considered to be the "gold standard". Evaluators subjectively evaluate the main characteristics of reliable translation quality, such as the adequacy and fluency of MT text. External methods, called the task-oriented methods (post-editing or reading with comprehension) are focused on efficiency and text usability with regard to a specific task. Automatic internal metrics do not require human intervention. It represents a significant breakthrough in the field of MT in terms of the quality of automatic evaluation and MT system optimization. Automatic metrics (accuracy, coverage, WER, PER, BLEU and others), which determine the quality of translation or errors through comparison, calculate the similarity between hypothesis and reference translation (and a given set of reference translations) and provide a relatively rapid feedback on the quality of the system as well as the newly-created text generated by the MT system.

With the introduction of IT-technologies as a tool for computer-assisted translation memories a new era of translation came into existence. At present, translatology is considered an interdisciplinary science working in conjunction with other disciplines. The idea of interdisciplinary science is based on the hypothesis that a natural language can use a variety of symbols; it can be fully analysed, controlled and mathematically encoded. With MT, this primary hypothesis of interdisciplinary studies was extended following the results of experimentation with natural language processing. This is based on the fact that language is so rich and complex that it cannot be completely analysed and split into a set of rules that can be subsequently encoded as computer program algorithms.

There have been many suggestions of how to measure quality, some focusing on target specific syntactic constructions, others assessing various sentences as a whole on the N-point scale or on automated translations with a reference. However, these methods have been mainly tested on major languages (English and other world languages). There have been very few attempts to focus on inflectional languages such

as Slovak, with its specific morphological richness. In the present paper we focus on the evaluation of MT in the language pair Slovak (Source) vs. German (Target).

In our evaluation we use the following metrics:

1. F-measure (Precision and Recall)

These are the easiest automatic evaluation metrics and are often used in natural language processing. They are based on a word congruency hypothesis with the words in the referential translation, regardless of the word position in a sentence. They have a mutually opposite relationship, which means that increasing accuracy scores may lead to the reduction of the coverage score and vice versa.

A disadvantage of accuracy metrics is where the hypothesis falls short in terms of the number of words, but acquires a high accuracy score (but low coverage). The reverse also applies: we can get a hypothesis containing all possible words, which increases the likelihood that some of these words will also appear in the referential translation, but the hypothesis is too long, with a high score coverage, but low accuracy score.

The question is how to solve this problem. We do not want to generate a sentence (hypothesis) that would include misspelled words, but we do not want to have any omissions as well. Therefore, experts from the field of MT come up with the F-measure, also known as the F-score. It is actually the harmonic mean of the two metrics (accuracy and coverage).

2. WER (Word-error rate)

WER metrics or error word rate was first used in the evaluation of statistical MT. It belongs to the first generation metrics of automatic MT evaluation systems. WER rate was taken from the field of speech recognition and is based on the edit distance taking into account the word order. The edit distance is represented by the Levenshtein distance, which is defined as a minimum number of single-character edits (insertion, removal and substitution) needed to achieve a conformity of two sequences (sentences).

3. PER (Position-independent Error Rate)

This is occasionally used in the evaluation of MT. It is similar to metrics coverage by using the same denominator – the length of the referential translation or the number of words in the reference. As the name suggests, it is a certain degree of error rate. It does not measure congruence, but the mismatch. It takes into account defective and superfluous words which must be removed in long translations.

RESEARCH METHOD OF AUTOMATED TRANSLATION EVALUATION

Case study: Translation of a record from Slovak into German by using the online system of MT

In our experiment we used an automated evaluation with metrics of a written record ("verbatim transcript of a meeting") which was translated from the Slovak language into German without pre-editing. It was a 12-page text documenting a working meeting within the framework of an EU-project involving cross-border regional cooperation between Slovakia and Austria (RECOM).

There were two approaches used during the translation process: MT with the "statistical system of MT" and the classical computer-assisted translation with electronic dictionaries. There were two reasons for choosing the online system of MT in our case study:

1. Easy access for any user working online.
2. Online system of statistical MT is the only system that can translate from a large number of language pairs (even from and to synthetic analytic languages).

The translation (referential translation – RT) in our case study has been revised by a German native speaker. Afterwards, both outputs were compared by a software program *MT evaluator* (this program was developed in close cooperation between the Institute for Computer Science and the Institute of Translation studies at Constantine the Philosopher University in Nitra). In the qualitative analysis of the translated text using statistical MT, we have taken two important criteria into consideration:

1. F-measure – transfer of the lexically most appropriate words and phrases (adequate, faithful reproduction).
2. WER – syntactical word order appropriateness (linguistic correctness and readability).

We have also partially dealt with some shortcomings in the original text which contributed to some mistranslations. The following are problem areas that we have encountered:

1. A lack of input quality from a linguistic perspective. The first difficulty for the translator was some unintelligible text segments, formulations, or sentences in the source text. It was a literal transcript of the audio-discourse that included implicit or unclearly formulated information. The incomprehensibility

occurred because of artificial ad-hoc language that was very difficult to understand in the source text without background-knowledge, which is why it was not easy to transfer the text into the target language without consulting the producer of the text.

Here are some examples from the source text:

3) „***Priestorové pokrytie*** *považuje za dostatočné...*" (literal translation as „räumliche Deckung" – this phrase does not exist either in the source text or in the target text)

4) „*….ak sa urobia opatrenia v jednej časti,* **nie je možné to uzavrieť tým, že tam to končí***...*"(unclear wording, too many demonstrative pronouns without factual reference)

5) „*... treba mať k dispozícii zoznam chránených druhov a živočíchov a* ***voči nim vykonávať prieskum***" (stylistically unclear – to make an investigation against sb.)

6) „*...Údaje majú zhŕňať celoročný aspekt…*" (the nominal expression is redundant)

7) „*.....v otázke migrácie živočíchov nie je dôležité* **okrem smernice** *na ochranu vtáctva zohľadniť aj existenciu Alpsko-karpatského koridoru pre vysokú zver...*" (the phrase in bold is superfluous and disturbs the cohesion, coherence, and thus the meaning of the whole statement)

8) „*...1x/rok je možné podať žiadosť o zmenu projektu a ešte stále je* **priestor na zareagovanie***...*" (a vague formulation – an opportunity for a response)

9) „*…Otázka premietnutia do finančných otázok je zatiaľ neznáma.*" (stylistically unclear and vague)

Suggestion: Since neither machine nor translator is capable of translating unclearly formulated sentences into the target language without background knowledge, the original texts should be adapted (pre-edited) before translation. Similarly, it appears essential to instruct the source text writers about how the texts should be written. The source text should be as unambiguous and concise as possible.

2. Variable output quality (MT and referential translation) from a linguistic perspective. In the first phase we translated the target text (translation product) without using *Google Translate* (human translation with the native German editor). Based on the referential translation we evaluated the statistical MT (Google translation) of the written record by measuring its quality. The following table shows the percentage results in the congruence between the referential translation and the statistical online MT (*Google Translate*) from the selected 22 text segments.

Table 3. BLEU 1, 2, 3, 4 – the number of consecutive correct word combinations (1 word phrase, 2 phrases, etc.) (in %)

Precision	Recall	F-measure	PER	WER	BLEU 1	BLEU 2	BLEU 3	BLEU 4
33,33	50	40	50,5	100	33,33	12,5	0	0
33,33	25	28,57	75	75	25	18,75	10,71	0
58,33	50	53,84	50	50	50	31,17	17,14	9,52
21,43	20	20,69	80	86,67	20	7,18	0	0
71,43	55,56	62,5	44,44	55,56	55,56	25,93	0	0
52	37,14	43,33	62,86	77,14	37,14	8,93	0	0
25	27,27	26,09	72,82	90,91	25	9,09	0	0
41,46	37,78	39,53	62,22	82,22	40	6,83	2,34	0
25	33,33	28,57	67	106,67	35	5,26	0	0
44,44	30,77	36,36	69,23	73,08	34,62	16,29	12,98	9,23
21,43	19,35	20,34	80,65	93,55	22,58	3,35	0	0
25	25	25	75	83,33	33,33	0	0	0
33,33	25	28,57	75	75	25	6,82	0	0
16,67	16,67	16,67	83,33	91,67	16,67	0	0	0
44,44	38,71	41,38	61,29	70,97	38,71	13,4	0	0
20	14,29	16,67	85,71	100	14,29	0	0	0
22,86	27,59	25	72,62	110,34	25,71	0	0	0
33,33	37,5	35,29	62,62	75	33,33	25	14,29	0
29,41	25	27,03	75	95	25	0	0	0
27,78	22,73	25	77,27	86,36	22,73	4,81	0	0
33,33	36,36	34,78	63,73	109,09	33,33	18,18	10	0
41,67	29,41	34,48	70,59	76,47	29,41	6,42	0	0
755	684,46	709,69	1516,88	1864,03	675,74	219,91	67,46	18,75
34,31818182	**31,11182**	**32,25864**	**68,94909**	**84,72864**	**30,71545**	**9,995909**	**3,066364**	**0,852273**

Analysis of the best and worst MT results according to the in the referential translation (RT) and MT (MT)

In the following lines we tried to pick out the 3 best and the 3 lowest results of MT from the 22 measured segments by using the metrics of F-measure (an average of precision with coverage).

The best results of F-measure (3. 5. 6. segment)

10) <u>Referential translation</u>: *Dann folgte die Ansprache des Bürgermeisters der Gemeinde Hohenau an der March, Herrn F.*

<u>MT</u>: *Gefolgt von einer Rede des Burgermeisters Hohenau an der March p F.*

Two words are missing in the translated sentence – *Gemeinde* – village, and *dann* – then. The main verb *folgen-folgte* has changed its form into a participle without changing the whole meaning.

11) <u>Referential translation</u>: *Herr Pr. hat den Anwesenden das Projekt BBMH vorgestellt.*

<u>MT</u>: *P Pr. fuhrte das vorliegende Projekt BBMH.*

In this short sentence the verb was substituted by its synonym (*vorstellen – führen)*, which causes a slight misunderstanding. *Vorstellen* means to introduce, *führen* means "to lead" or "to show someone round". Moreover, two different tenses are used (Perfekt in the referential translation and Präteritum in the MT) which are not stylistically compatible.

12) <u>Referential translation</u>: *Das Ziel ist es, anhand von Untersuchungen eine Studie zu erarbeiten und eine geeignete Variante für die Lösung der ganzjährigen Erschließung von Zufahrtswegen zu der Brücke über den Grenzübergang Moravský Sv Ján Hohenau.*

<u>MT</u>: *Das Ziel der Umfrage ist es eine Untersuchung durchzuführen und empfehlen eine geeignete Alternative zur Bereitstellung Ganzjahrig Zufahrten zur Brucke Grenzubergang mahrischen Vol John Hohenau.*

The MT generated one additional element (*Umfrage- survey*) that causes a shift in meaning. From the lexical perspective there is confusion of the words *Lösung* (solution) and *Bereitstellung* (provision). The adjectives were written in capital letters and the geographical names transferred into the target language, which is not acceptable. On the other hand, the MT system generated synonyms which do not alter the contextual meaning: the synonyms *durchführen – erarbeiten* (carry out research), *L*ösung – *Alternative*. The omitted prepositions – über *den Grenzübergang* (through the checkpoint) – disrupt the cohesionof the text.

The worst results of F-measure (11.14.16. segment)

13) <u>*Referential translation:* </u>*Es geht um die Termine, wann man die Untersuchungsergebnisse, Fokussierung, Bestimmung des Wasserspiegels und Empfehlungen aus der Sicht des Ökosystems u. ä. bekommen kann; Diese werden von der österreichischen Seite sichergestellt.*

<u>MT</u>: *Zu den Terminen wenn Sie die Ergebnisse einer Umfrage Fokus Ermittlung der maximalen Wasserstande und Empfehlungen in Bezug auf das Ökosystem erhalten und dergleichen die bietet AT Seite.*

This example sentence of MT demonstrates its potential for nonsense since the sentence elements do not maintain cohesion, perhaps due to incomprehensible or complex sentence structure in the Slovak source text. The machine-translated sentence reproduced only some words, such as *Ergebnisse, Empfehlungen and* Ökosystem, however without any logical linking. The word *Umfrage* is redundant and does not make any sense in the given context.

14) <u>Referential translation</u>: *Er hat die beauftragten Firmen aufgefordert, den Stand ihrer Arbeit zu präsentieren.*
<u>MT</u>: *Er nannte das Unternehmen verantwortlich fur die Prasentation der Fortgang der Arbeiten.*

In this sample sentence, in spite of its brevity, we can see the erroneous transfer of all lexical elements – *auffordern* vs. *nennen*, the adjective *verantwortlich* (responsible) is redundant and the verb *präsentieren* has been changed into a noun (*Präsentation*). *Fortgang der Arbeiten* could be regarded as lexically relevant, but in an incorrect morphological form. The synonyms *Firmen* and *Unternehmen* could be acceptable but they have been translated in the singular, even though in the referential text the plural is used.

15) <u>Referential translation</u>: *Das ganze Gebiet hält er für ausreichend.*
<u>MT</u>: Raumliche Abdeckung als ausreichend angesehen.

In this short sentence only the second part is correct. The first part stands for a non-existent lexical phrase which was caused by artificial source language formulation that needed additional explanation.

CONCLUSIONS

The aim of this article was to test the evaluation of MT for a specific text type – the written record. The analysis of the machine-translated text has demonstrated that written records are extremely complicated text types that are almost impossible to translate using automated translation systems. The main reason can be seen in the complex register and in the selected language style that uses various means of expression. Secondly, it is also important to take the sentence structure into consideration because machine-translated texts contain long and complex sentence structures. In addition, it should be noted that the lexis in automated translations provides diffuse equivalents which can be attributed to the various specialized areas and hybrid text functions of a written record. We have also found that the quality of the input plays an important role. This is why other types of written records should be tested. The results could be beneficial to both professional and amateur translators and enable them to organize and manage their work more effectively.

The final results have proved that MT is not suitable for certain text types because there are many anomalies at all linguistic levels between the referential translation (human translation without the intervention of computer-assisted translating) and MT. A particular disadvantage of the Google MT system for the Slovak-German language pair is that there are not many parallel texts for these languages included in the system as the majority of the text corpora are in English. This also applies to German technical texts that use numerous English terms, as well as borrowings from English. Our study confirmed that written reports (as well as other text types) should be pre-edited and re-written into shorter sentence segments. It is therefore necessary to carry out further research into MT of pre-edited records between various language pairs.

References

Anckaert Ph., Eyckmans J., Justens D., Segers W. 2013. Bon sens, Faux sens, Contresens et Non-sens sens dessus dessous. Pour une évaluation fidèle et valide de la compétence de traduction. *Le bon sens en traduction. Rivages Linguistiques.* Rennes: Presses Universitaires de Rennes, 79–93. (http://hdl.handle.net/1854/LU-4328090)

Arnold D. et. al. 2008. *Machine translation: an Introductory Guide.* London: Blackwells.

Byrne J. 2012. *Scientific and technical translation explained: A Nuts and Bolts Guide for Beginners. Manchester*: St. Jerome Publishing.

Callison-Burch et al. 2006. *Re-evaluating the role of BLEU in machine translation.* Proceedings of EACL, 249–256.

FLT – *Flexible Learning Toolboxes*: http://toolboxes.flexiblelearning.net.au/demosites/series14/14_03/content/or/learn_2.htm.

Hutchins W. J., Somers H. L. 1992. *An Introduction to Machine translation.* London: Academic Press.

Kit Chunyu, Pan Hayhua, Webster J.J. 2002. Example-based machine translation: A new paradigm. S.W. Chan (Ed.). Translation and innovation technology. Chinese University of Hong Kong Press, 57–78.

Munková D. 2013. *Prístupy k strojovému prekladu.* Nitra: Univerzita Konštantína Filozofa v Nitre.

Vilar, D., Xu, J., d'Haro, L. F., & Ney, H. 2006. Error analysis of statistical machine translation output. *Proceedings of LREC.* 697–702.

MAŠININIO VERTIMO KOKYBĖS VERTINIMAS: TRANSKRIBUOTO ŽODINIO DISKURSO VERTIMO IŠ SLOVAKŲ KALBOS Į VOKIEČIŲ KALBĄ ATVEJO ANALIZĖ

Jozef Štefčík

Santrauka

Šiame straipsnyje aptariama automatinio vertimo kokybės vertimo problema remiantis žodinio teksto vertimo iš slovakų kalbos į vokiečių kalbą pavyzdžiu. Autorius atliko eksperimentą, kurio tikslas – patikrinti susitikimų, posėdžių ir kitokių komunikacinių situacijų, per kurias žodžiu pasakomas turinys įrašomas į laikmenas (pvz., naudojant diktofoną ar vaizdo kamerą), o vėliau transkribuojamas kaip dokumentas (pvz., posėdžio protokolas), vertimo į kitą kalbą (konkrečiu atveju – vokiečių) kokybę. Vertinimui pasirinkti tikslumo, leksinės ir sintaksinės atitikties ir kiti kriterijai pagal Ackaerto ir kitų kartu su juo dirbusių mokslininkų (Ackaert et al. 2013) pasiūlytą metodiką. Vertimas atliktas naudojant statistiniu principu veikiančią automatinio vertimo sistemą *Google Translate*, vėliau lygintas su profesionalaus vertėjo darbu. Atliktas tyrimas patvirtino, kad kol kas kai kurių tekstų automatinis vertimas neįmanomas dėl daugelio neišspręstų problemų įvairiuose sistemos lygmenyse ir kad norint produktyviai panaudoti automatinio vertimo galimybes verčiant transkribuotus žodinio diskurso tekstus (kaip, beje, ir kitokio tipo rašytinius tekstus) tekstas turi būti tinkamai parengtas, t. y. atliktas jo išankstinis redagavimas (*pre-editing*), suskaidant tekstą į trumpesnius segmentus, kurie sistemoje būtų lengviau atpažįstami. Tinkamai parengus tekstą vertimo rezultatai buvo gerokai geresni.