MACHINE TRANSLATION: A CRITICAL LOOK AT THE PERFORMANCE OF RULE-BASED AND STATISTICAL MACHINE TRANSLATION

Brita Banitz[1]
[1]Universidad de las Américas Puebla, San Andrés Cholula, México

Abstract: The essay provides a critical assessment of the performance of two distinct machine translation systems, Systran and Google Translate. First, a brief overview of both rule-based and statistical machine translation systems is provided followed by a discussion concerning the issues involved in the automatic and human evaluation of machine translation outputs. Finally, the German translations of Mark Twain's *The Awful German Language* translated by Systran and Google Translate are being critically evaluated highlighting some of the linguistic challenges faced by each translation system.
Keywords: Rule-based machine translation. Statistical machine translation. Evaluation of machine translation output

1. Introduction

1.1 Defining machine translation

In today's globalized world, the need for instant translation is constantly growing, a demand human translators cannot meet fast enough (Quah 57). Machine translation (MT), defined by Somers as "a range of computer-based activities involving translation" (Somers 428), is therefore considered a "cost-effective alternative to human translators" (Quah 57).

The goal of MT is, according to Hutchins and Somers, the production of useful automatic translations within specific contexts, requiring the least amount of changes to the output in order to make it acceptable to users (Hutchins and Somers 2). But the early history of MT was driven by an unrealistic expectation of creating computer programs capable of high-quality fully automatic translation and the infamous ALPAC report of 1966, which argued that "MT was slower, less accurate, and twice as expensive as human translation" (Somers 428), brought MT research to a standstill in the USA. However, research in other countries continued thus leading to the realization that high-quality fully-automatic translation was not feasible and that systems producing acceptable output, often based on restricted texts, were preferable (Somers 429).

Quah distinguishes between three generations of MT architectures: The first generation (1960s to 1980s) was based on direct translation, the second generation (1980s to present) consists of rule-based systems such as the transfer and interlingua systems, and the third generation (1990s to present) includes corpus-based systems that are either statistical based or example based (Quah 68). While direct translation systems employed a "word-for-word translation … with no clear built-in linguistic component" (Quah 60), the rule-based and corpus-based systems are far more complex and will be dealt with in more detail below.

1.2 Objectives

The purpose of this essay is to provide an overview of two different approaches to MT, rule-based and statistical MT, and to critically analyze the performance of each based on the translation of a short text translated by Systran and Google Translate. Systran is a well-known rule-based system freely available online at http://www.systranet.com/translate. Google, on the other hand, is a statistical MT system which is based on a large corpus of bilingual aligned texts. The free online translator can be accessed at https://translate.google.com.

As the source text for the translation, the first 24 sentences, 687 words, of the English text *The awful German language* by Mark Twain was used whereas the German outputs of Systran and Google Translate served as the target texts for the present analysis. In addition, Schneider's human translation into German served as a reference translation to evaluate the MT output.

In section 2 below, a brief overview of both rule-based and statistical machine translation is given followed by section 3 which presents some of the issues related to the automatic and human evaluation of the outputs provided by MT systems. In this section, the performance of both Systran and Google Translate as well as the linguistic challenges faced by both MT systems are discussed in greater detail.

2. Approaches to MT

Currently, the two most common MT systems are rule-based MT (RBMT) and statistical MT (SMT) (Costa-Jussà *et al*. 247). Both approaches are dealt with next.

2.1 Rule-based MT

According to Quah, "rule-based approaches involve the application of morphological, syntactic and/or semantic rules to the analysis of a source-language text and synthesis of a target-language text" (Quah 70-71) requiring "linguistic knowledge of *both* the source and the target languages as well as the differences between them" (Douglas, Arnold *et al*. 66, *emphasis in original*). Rule-based systems are further divided into transfer and interlingua systems (Hutchins; Somers). Interlingua systems work with an abstract intermediate representation of the source text out of which the target text is generated "without 'looking back' to the original text" (Hutchins; Somers 73). However, in practice, "designing a general-purpose interlingua is tantamount to designing a complete

model of the real world" (Forcada 219) limiting this approach to translation within specific domains only (*ibid*.).

Consequently, the more common approach to rule-based MT are transfer systems (218). According to Somers, transfer-based systems analyze a source text sentence by sentence identifying the part of speech of each word and its possible meanings (Somers 433). If the source language is morphologically rich, language specific morphological rules are used to analyze the source text. Language-specific syntactic rules are then applied to identify the syntactic categories of the words contained in the sentence. Finally, the system determines the target word and generates the target sentence closely following the structure of the source sentence (*ibid*.) further subjecting the target sentence to a "simple *morphological generation* routine" (Hutchins; Somers 134, *emphasis in original*) in order to apply target language-specific morphological rules to the MT output.

2.2 Statistical MT

Statistical MT, on the other hand, is currently "the overwhelmingly predominant method in MT research" (Somers 434). Working with massive bilingual corpora, the system looks for the target sentence with the highest probability match. This is different from the example-based method in which the system searches for a previously translated sentence in an aligned corpus of translated source and target sentences, similar to using a translation memory (Forcada). Since both methods work with large corpora of parallel texts, they are commonly classified as corpus-based approaches to MT (*ibid*.).

Statistical MT systems are further divided into word-based and phrase-based models (Costa-Jussà *et al*.). Word-based models work with the assumption that for each individual word, the probability for how that word should be translated can be computed. However, more modern SMT systems use phrases as the unit of translation (251) where a phrase is defined as a "contiguous multiword sequence, without any linguistic motivation" (Koehn 148).

After a source text is segmented into phrases, these are subsequently compared to an aligned bilingual corpus and a statistical measure is used to compute the most probable target-language segment based on the information gathered from the system's translation model and target-language model (Quah 77). The translation model is responsible for calculating the degree to which each source-language word contained in the phrase corresponds to possible target-language words selecting the most probable lexical choice contained in the corpus (Somers). The target-language model, on the other hand, computes how likely it is that the target segment is considered legitimate, again based on the data contained in the bilingual corpus (*ibid*.). As a final step, the target text is produced with the newly translated segments (Quah). In the next section, the evaluation of MT output is discussed in greater detail.

3. Evaluation of MT output

As Douglas *et al*. point out, "the evaluation of MT systems is a complex task" (157) since the adequacy of a system's output largely depends on the purpose of the translation (Forcada; Somers). Therefore, there is "no *golden standard* against which a translation can be assessed" (Kalyani *et al*. 54, *emphasis in original*). In the following section, an overview of both the automatic and human evaluation of MT output is provided and a critical discussion of the Systran and Google translations of the source text mentioned above is offered.

For the analysis of the MT output, the first 24 sentences of the source text were entered into Systran, translated, and copied into a Word document. The same procedure was followed for Google. Next, all of the sentences were aligned along with the corresponding reference translation and subjected to automatic and human evaluation.

3.1 Automatic evaluation

The automatic evaluation of MT output has "become the norm" (Somers 438) since it is faster and more cost efficient (Kalyani *et al.*), more objective (Quah), allows for a large number of outputs to be evaluated (Somers), and provides useful and immediate feedback during system development (Forcada). According to Somers, the most widely used automatic evaluation metric is BLEU. It compares the MT output, segmented into four-word sequences, to a human reference translation in terms of lexical precision and assigns a score of 0 for the worst translation and a score of 1 for the best translation (Costa-Jussà *et al.* 257). However, the system is limited to a relatively small sequence of words, "penalizes valid translations that differ substantially in choice of target words or structures" (Somers 438), does not efficiently evaluate the MT output of free word order languages such as Hindi (Kalyani *et al.* 57), and greatly underestimates the quality of non-statistical system output compared to human raters (Callison-Burch; Osborne; Koehn), a shortcoming that also applies to other automatic evaluation engines such as METEOR and Precision and Recall (Callison-Burch *et al.*). As a consequence, other measures have been proposed.

One such measure is the TER score suggested by Snover, Dorr, Schwartz, Micciulla, and Makhoul. The TER score, or Translation Error Rate, "measures the number of edits required to change a system output" (Costa-Jussà *et al.* 257) to match a human reference translation as closely as possible (Snover *et al.*). According to Snover *et al.*, insertions, deletions, substitutions, and changes in word order count as edits (Definition of translation edit rate, para. 2). Yet, while the measure does give some indication as to how close the MT output is to a human translation, two important shortcomings have to be pointed out. First, the TER score does not necessarily reflect the acceptance or adequacy of the MT output (para. 6) and second, the measure directly depends on the quality of the reference translation since any deviation from the human translation will be penalized. Nonetheless, the TER score offers a

"more intuitive measure of 'goodness' of MT output" (*Introduction*, para. 2) and can be easily calculated using the Levenshtein distance calculator, a free measurement tool available online at http://planetcalc.com/1721 .

Using the Levenshtein distance calculator, the TER score measure was applied to the translations of the source text provided by Systran and Google. The results of the automatic evaluation of the output are presented in Table 1 below.

Table 1: Results of the automatic evaluation of the output

Sentence	Length source sentence	Length target sentence Systran	TER score target sentence Systran	Length target sentence Google	TER score target sentence Google	Length reference translation sentence
1	25	25	93	22	56	21
2	6	6	**26**	7	**37**	8
3	29	28	81	28	33	28
4	27	51	53	27	17	27
5	43	41	122	38	86	40
6	18	17	66	18	56	14
7	20	19	**67**	19	**68**	18
8	63	60	226	52	195	46
9	19	19	**56**	20	**64**	17
10	14	18	73	19	60	20
11	9	9	28	6	20	9
12	42	41	173	43	156	42
13	30	29	73	25	28	24
14	27	23	51	22	42	20
15	15	14	42	13	26	15
16	12	12	42	11	39	14
17	13	13	**29**	12	**34**	16

18	27	29	119	26	97	20
19	28	26	85	23	62	25
20	17	16	38	15	36	14
21	83	74	219	72	182	66
22	44	44	150	44	104	42
23	31	25	103	26	101	29
24	45	44	198	42	155	49
total	687	683		630		624
average	28.6	28.5	92.2	26.3	73.1	26

Source: Autor

Table 1: TER scores

Table 1 above lists the word length of each of the source sentences, the length of target sentences translated by Systran, the length of the target sentences translated by Google, and the length of the sentences of the human reference translation. The TER scores for each of the sentences translated by Systran and Google are provided along with the overall word count of each of the texts, the average word length per sentence, and the average TER scores for the Systran and Google translations.

As the table indicates, the average word length of the source text sentences was 28.6 words per sentence, very similar to Systran's translation with an average sentence length of 28.5 words per sentence. Google's sentence length was slightly less with an average of 26.3 words per sentence, closer to the reference translation with an average of 26 words per sentence. Similarly, the total word count of the source text was 687 words, Systran's translation consisted of a total of 683 words, whereas Google's translation consisted of fewer words, a total of 630 words, again closer to the reference translation with a total of 624 words. The translation offered by Google is therefore more similar to the human translation in overall word count as well as the average number of words per sentence.

As far as the TER score is concerned, Systran's translation resulted in an average TER score of 92.2 whereas the average TER score for Google was 73.1 indicating that, in general, Google's output requires fewer edits to match the reference translation more closely. In fact, out of the 24 target sentences, only four obtained a higher TER score for the Google translation (marked in bold). It also appears clear that the longer the sentence, the higher the TER score in general. The obtained results suggest that automatic evaluation measures, at least the one used here, evaluate the SMT output more favorably than the RBMT output and that the translation by Systran requires more post-editing to be closer to the human reference translation.

3.2 Human evaluation

Although the human evaluation of MT output is costly, time-consuming, and rather subjective (Kalyani *et al.*), it does provide a more detailed analysis of the quality of the output depending on the rating criteria applied. From a set of target translations, the evaluator chooses the best translation option based on the provided reference translation (Farrús; Costa-Jussà; Popović). Although different rating scales do exist, the most common evaluation criteria suggested in the literature are fluency and adequacy (Quah). Fluency, also referred to as intelligibility (Douglas *et al.*), is concerned with both the grammatical correctness and word choice of the translation (Kalyani *et al.*) whereas adequacy, also called accuracy or fidelity (Douglas *et al.*), evaluates the degree to which the translation managed to represent the original meaning (Kalyani *et al.*). The rating scales suggested by Callison-Burch *et al.* (Implications) are, in my opinion, the most concrete suggested in the literature and were therefore used to assess the MT output provided by Systran and Google. Both scales are represented in Table 2 and Table 3 below:

Table 2: Fluency scale

Fluency
How do you judge the fluency of this translation?
5 = Flawless German 4 = Good German 3 = Non-native German 2 = Disfluent German 1 = Incomprehensible

Table 3: Adequacy scale

Adequacy
How much of the meaning expressed in the reference translation is also expressed in the hypothesis translation?
5 = All 4 = Most 3 = Much 2 = Little 1 = None

Source: Autor

After having applied both scales to the output provided by Systran and Google, the results were summarized in Table 4 below. The table lists the sentence by sentence fluency and adequacy scores for the source text translations along with the average score for each scale as well as the percentage of how often one system was chosen as better. As can be seen in the table, 75% of the fluency scores were better for Google whereas 25% were rated as equal to Systran. On the other hand, none of the sentences translated by Systran were rated better than Google with Google achieving an average fluency score of 3.6 compared to Systran's average fluency score of 2.5.

The length of the sentence did not seem to affect the fluency scores since regardless of length, Google's translation tended to receive a higher fluency score indicating that the grammaticality of the translation offered by Google was generally better than Systran's. This was an expected result because, as suggested by Costa-Jussà *et al.*, Systran's approach to translation is rule-based, translating each sentence word-for word, which tends to result in lower fluency scores.

Table 4: Fluency and adequacy scores

Sentence	Length target sentence Systran	Fluency score target sentence Systran	Adequacy score target sentence Systran	Length target sentence Google	Fluency score target sentence Google	Adequacy score target sentence Google
1	25	2	3	22	4	5
2	6	4	5	7	4	5
3	28	3	4	28	4	5
4	51	3	4	27	4	5
5	41	3	4	38	4	5
6	17	2	3	18	4	4
7	19	2	3	19	4	4
8	60	1	2	52	3	4
9	19	3	4	20	4	5
10	18	3	4	19	4	4
11	9	3	4	6	5	5
12	41	2	3	43	3	4
13	29	3	4	25	5	5
14	23	3	4	22	5	5
15	14	3	5	13	4	5
16	12	2	3	11	3	3
17	13	2	2	12	4	5
18	29	3	3	26	3	3
19	26	2	2	23	2	2
20	16	3	3	15	3	3
21	74	3	4	72	3	4
22	44	2	2	44	3	3
23	25	2	3	26	3	4
24	44	2	2	42	2	2

average	28.5	2.5	3.3	26.3	3.6	4.1
percent		0%	0%		75%	63%

Source: Autor

As far as the adequacy score is concerned, Google was also evaluated as better with 63% of the scores being higher than Systran's and 37% being equal. For shorter than average length sentences, Systran did receive a better result compared to its fluency score, which indicates that the content of the source sentence was represented better than its grammatical structure might suggest. Yet, Google still received a higher score overall in terms of adequacy representing the original meaning of the source sentence more faithfully than Systran. Therefore, even though the adequacy of Sytran's translation was rated slightly better than its fluency, Google was rated better overall for both criteria.

3.3 Linguistic challenges for MT systems

The fluency and adequacy measures discussed above, however, still do not provide any insight into the types of errors both systems committed. In order to gain a better understanding of the challenges faced by both Systran and Google, a linguistic error analysis of the systems' translations of the source text was performed, taking into consideration the following sub-categories within the classification suggested by Farrús *et al.*, p. 176-177 (see Table 5 below):

Table 5: Classification of linguistic errors

Classification	Category
Semantic errors	Homograph Polysemy
Lexical errors	Incorrect word Unknown word Missing target word Extra target word

Syntactic errors	Wrong word order Wrong preposition
Morphological errors	Grammatical case marker

Source: Farrús *et al*.

The most common error for both systems was *semantic* in nature. For Systran, this was an expected result since, according to Costa-Jussà *et al.*, RBMT systems follow a word-for-word translation methodology, resulting in output that "tends to be literal and lacks fluency" (Costa-Jussà *et al*. 252). Thus, a particular problem for these systems is, therefore, lexical ambiguity where "one word can be interpreted in more than one way" (Hutchins; Somers 85) as is the case with homographs and polysems. Homographs are words that are spelled the same way but that have different meanings. Systran, for example, incorrectly translated the word "sentence" in sentence 24 as "Strafe" [penalty] instead of "Satz" [sentence], whereas Google translated the homograph correctly. Yet while there were only three cases of homographs in the analyzed data (two by Systran and one by Google), most of the sentences, for both Systran and Google, had problems with polysemy.

Polysems are words carrying several related meanings. One example included the verb "know" which was incorrectly translated with "kennen" [to know somebody/something] by Systran yet correctly rendered as "wissen" [to know something about somebody/something] by Google. As is the case with polysems, the choice of the correct target word depended on the context (Somers 431). Polysemy was the most common error for both systems. Out of the 24 target sentences, 18 sentences translated by Systran had issues with polysemy. For a statistical MT system like Google, however, this problem was not expected since it was not listed as a potential problem by Costa-Jussà *et al*. Out of the 24 source sentences, 13 had issues with polysemy, suggesting that the result is most likely a function of the type of source text chosen for this analysis. Since it is literary in nature, I believe it is open to more interpretation, thus

giving rise to more ambiguity compared to a source text employing controlled language, defined by Quah as featuring "pre-established vocabulary and sentence structures" (Costa-Jussà *et al*. 66).

As far as the *lexical* errors are concerned, all of the subcategories listed in Table 5 above were present, albeit not very often. One particular problem in this category, however, were the phrasal verbs in English. Systran, for example, rendered the expression "wash about" as "ungefähr … waschen" [roughly wash] instead of "hin und her schwemmen" [to and fro wash]. Google translated the expression as "herumspülen" [around wash] which is another viable option. Overall, phrasal verbs were translated incorrectly seven times by Systran and twice by Google, suggesting that Systran's rule-based approach does not easily recognize phrasal verbs in English because of its word-for-word analysis of the source text. Google, on the other hand, which is based on matching phrases in a parallel bilingual corpus, did not have a particular problem with rendering the English phrasal verbs correctly in German.

There were, on the other hand, only two cases of unknown words, both in the translations by Systran, where the source word was not translated, suggesting that, contrary to expectations (Costa-Jussà *et al*.), Google's SMT approach did not have any issues with words not present in its *corpus*. There were two cases of missing target words in the Google translations, both of them omitted verbs, compared to one case in Systran where a noun was missing. Finally, three instances of added target words were identified in Systran and two cases in Google, although there does not seem to be a clear pattern since the extra words were personal pronouns, prepositions, and one noun.

According to Costa-Jussà *et al*., *syntactic* problems arise when the source and target-languages have different word order rules, which can be a particular issue for SMT systems. RBMT systems, on the other hand, struggle with structural ambiguity, cases in which "there is more than one way of analyzing the underlying structure of a sentence" (Hutchins; Somers 88).

Yet, there were no cases of structural ambiguity in the data analyzed. Concerning word order errors, however, Systran had issues with the correct positioning of the finite verb in German, the infinitive with "zu" [to] construction (*e.g.*, "Ich ging häufig, die Sammlung von Kuriositäten … und einen Tag zu betrachten …" [I went frequently the collection of peculiarities and one day to see] instead of the syntactically correct "Ich ging häufig, die Sammlung von Kuriositäten zu betrachten und eines Tages …") [I went frequently the collection of peculiarities to see and one day], and separable verbs in German. In total, out of the 24 sentences translated, there were syntactic problems in 13 of them. Google, on the contrary, had only three cases of syntactic errors involving the position of verbs. Choosing the wrong preposition, on the other hand, did not prove to be an issue since there were only two cases found in the data, both of them in the Systran translations (*e.g.*, "für … zu jagen" [for to hunt] instead of the correct expression "um … zu jagen" [to to hunt]).

Finally, *morphological* problems arise if the target language features morphological rules different from the source language (Costa-Jussà *et al*.). In German, particularly problematic are the morphemes that mark grammatical case which appear to have been an issue for Systran. Of the six erroneous cases detected in the data, all were either preposition plus article (*e.g.*, "in den Schmiedeshop" [to the blacksmith shop] instead of the correct "in dem Schmiedeshop" [in the blacksmith shop]) or preposition plus pronoun sequences requiring the dative case in German. Google, on the other hand, translated these instances with the correct case markers. This was a surprising result since, according to Costa-Jussà *et al*., it is the SMT systems which tend to have issues with morphological rules.

Overall, it is interesting to note that of the 24 sentences translated by Google, seven presented no errors at all whereas all of the sentences translated by Systran had at least one linguistic error indicating that the Google translations were also better overall as far as the linguistic errors are concerned. In sum, it is evident that both the RBMT and SMT systems have their advantages but also face a number of challenges. According to Costa-Jussà *et al*., the primary

advantage of RBMT systems is that it is easy to perform error analysis since these systems are based on linguistic theories. SMT systems, on the other hand, do not require any linguistic knowledge and are therefore language independent. Among the chief disadvantages for RBMT systems are that since language specific rules and dictionaries are required, these systems are language dependent and can therefore not be used freely with new language pairs. The main disadvantage of SMT systems is that problems arise with language pairs that differ morphologically and syntactically (253).

4. Conclusions

The essay provided an overview of rule-based and statistical MT, a discussion of different evaluation approaches to MT output, and an assessment of the source text translations offered by Systran and Google Translate. For the three evaluation measures used, the TER score for the automatic evaluation, the fluency and adequacy scores for the human evaluation, and the analysis of the linguistic errors, Google Translate fared better in all of them.

It is important to point out, however, that the evaluation of MT output is not without controversy. Automatic evaluation, for example, has been criticized for underestimating the quality of the RBMT output (Callison-Burch *et al*.), while measures such as the TER score are inconclusive as they do not provide any information regarding the acceptability of the translation to human users (Snover *et al*.). The human evaluation of the MT output, on the other hand, is not without its problems either, since raters may differ greatly in their judgment as to the quality of the translations, rendering the evaluation fairly subjective and unreliable (Kalyani *et al*.). This is an important limitation to the analysis presented in this essay since I was the only one evaluating the MT outputs.

A final caveat worth mentioning here is the type of source text used for this essay. Even though the Google translation was quite good in terms of its fluency, expressing most of the meanings of

the source text, grammatically it was still evaluated as somewhat non-native German. Therefore, considering that "general-purpose machine translation systems are still not suitable for certain types of text, especially creative texts" (Quah 66), MT should preferably be used for the purposes of assimilation, to understand the gist of a source text, and dissemination, to produce a machine translated target text for publication (Forcada), rather than the translation of literary texts similar to the one used here.

References

Callison-Burch, Chris; Osborne, Miles; Koehn, Philipp. "Re-evaluating the role of BLEU in machine translation research," *11th Conference of the European Chapter of the Association for Computational Linguistics*. 2006. Avaible to: http://homepages.inf.ed.ac.uk/pkoehn/publications/bleu2006.pdf. Accessed 13 February 2019.

Costa-Jussà, Marta R *et al*. "Study and comparison of rule-based and statistical Catalan-Spanish machine translation systems." *Computing and Informatics*, 31 (2012): 245-270.

Douglas, Arnold *et al*. *Machine translation:* An introductory guide. Oxford: Blackwell, 1994.

Farrús, Mireia *et al*. "Study and correlation analysis of linguistic, perceptual, and automatic machine translation evaluations." *Journal of the American Society for Information Science and Technology*, 63.1 (2012): 174-184.

Forcada, Mikel L. "Machine translation today." In: Gambier, Yves; Doorslaer, Luc (Ed). *Handbook of translation studies*. Amsterdam: John Benjamins, 2010. p. 215-223.

Hutchins, William John; Somers, Harold L. *An introduction to machine translation*. London: Academic Press, 1992.

Kalyani, Aditi *et al*. "Evaluation and ranking of machine translation output in Hindi language using precision and recall oriented metrics." *International Journal of Advanced Computer Research*, 4.14 (2014): 54-59.

Koehn, Phillip. *Statistical machine translation*. Cambridge: Cambridge University Press, 2010.

Schneider, Michael. "Die schreckliche deutsche Sprache." Avaible to: https://www.hmtm-hannover.de/uploads/media/Die_schreckliche_deutsche_Sprache_06.pdf. Accessed 13 February 2019. Acessed

Snover, Matthew, *et al*. "A study of translation edit rate with targeted human annotation." *Proceedings of association for machine translation in the Americas*. Vol. 200. nº. 6. 2006. Avaible to: https://www.cs.umd.edu/~snover/pub/amta06/ter_amta.pdf. Accessed 13 February 2019.

Somers, Harold L. "Machine translation: History, development, and limitations." In: Malmkjaer, Kirsten; Windle, Kevin (Ed). *The Oxford handbook of translation studies*. Oxford: Oxford University Press, 2011. p. 427-440.

Twain, Mark. "The awful German language." Avaible to: https://www.cs.utah.edu/~gback/awfgrmlg.html#x1. Accessed 14 May 2018.

Quah, Chiew Kin. *Translation and technology*. New York: Palgrave Macmillan, 2006.

Zydroń, Andrzej; Liu, Qun. "Measuring the benefits of using SMT." *MultiLingual*, 1/2 (2017): 63-66. Avaible to: http://dig.multilingual.com/2017-01-02/index.html?page=63. Accessed 13 February 2019.

Brita Banitz E-mail: brita.banitz@udlap.mx
ORCID: https://orcid.org/0000-0002-7254-0626