User expectations towards machine translation: A case study

Barbara Heinisch
Centre for Translation Studies
University of Vienna
Austria
barbara.heinisch@univie.ac.at

Vesna Lušicky
Centre for Translation Studies
University of Vienna
Austria
vesna.lusicky@univie.ac.at

Abstract

Neural machine translation (NMT) systems have emerged as powerful platforms for providing fluent translations in a variety of languages and domains. The widespread adoption of NMT has heightened the need for studying the results and impact of these systems. Although acceptance of machine translation has been analyzed, the expectations of users towards NMT have not received much attention yet. This paper investigates the expectations of novice translators enrolled on a postgraduate program in specialized translation. In addition, it examines the confirmation or disconfirmation of expectations towards machine translation (MT) output among this user group. A three-step mixed-method approach was applied: a quantitative questionnaire and two recurrent (pre-trial and post-trial) evaluations of raw MT outputs. The evaluations consisted of the identification and classification of errors in NMT output according to the Multidimensional Quality Metrics. The respondents expected the MT output to be of rather low quality, but the quality of NMT output was not as high as the participants expected. Compared to the expected frequency of error types in the MT output, the reported frequency differed significantly. This paper argues that the users' experience and expectations have an impact on the use and evaluation of machine translation.

1 Introduction

Language technology applications have become a ubiquitous service used by various user groups to overcome language barriers. While certain types of technology, such as translation memory systems, are specialized tools used by translators only, machine translation (MT) systems are also used by non-translators. If the exposure of MT users was somewhat limited to gist translation in the past, users are increasingly implementing MT in professional and other scenarios. The acceptance of MT tools and services is attested by the high number of users of generic online MT services (Way, 2018). Based on their prior experiences, users develop and form expectations towards MT.

Expectations are beliefs about attributes or performance of a product or service in the future (Olson et al., 1979). Users' expectations may have an influence on the intended use and evaluation of MT. Expectations also provide the frame of reference for satisfaction (Higgs et al., 2005). Satisfaction with a service is crucial when introducing or evaluating MT. Expectations are dynamic constructs, a synergy of users' pre-trial perceptions and beliefs about performance or attributes of a product or a service. Although there is some ambiguity regarding the definition and operationalization of expectations, the service quality literature differentiates several categories of expectations, most frequently: forecast, normative, ideal and minimum tolerable. The four categories cover different dimensions of expectations: forecast describes users' perception of what will occur; normative describes users' perception of what should occur; ideal describes the highest level attainable in a category; and minimum tolerable describes the minimum baseline for normative and ideal

© 2019 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC BY-ND.

(Higgs et al., 2005). Users' expectations and the type of expectations depend on internal and external cues, such as users' prior experience and information on products.

Users' (quality) expectations towards MT output and resulting implications for MT use are an under-explored topic in MT research (Way, 2018). So far, expectations were addressed in relation to the estimation of the quality of post-editing effort (Specia et al., 2009). Way (2018) gives an overview of what level of quality can be expected from MT. Existing research recognizes the critical role played by adoption (Cadwell et al., 2018) and acceptance of MT (Moorkens & Way, 2016; Koskinen & Ruokonen, 2017). Gaspari et al. (2015) also attempted to map the expectations, requirements and needs of the translation industry concerning translation quality and MT.

With the widespread application of neural machine translation (NMT) as the MT approach of choice in generic as well as specialized MT systems, the question of pre-trial user expectations should be addressed, especially user expectations based on previous use and information obtained on the service. They may have implications for the users' intended purpose of MT use and their satisfaction with the service. The notion of expectations should also be considered in human evaluation of MT output: the types of expectations and a potential negative bias may influence the results of human evaluations of MT output.

There is a growing body of literature that recognizes the importance of quality assessment of MT output. For MT developers, scale and robustness are major concerns, but end-users are also interested in receiving good-enough or high-quality translations (Way, 2018). The concept of fitness-for-purpose of translation has been widely recognized, but the assessment methods vary in operationalization and theoretical framework. The quality of MT output is either assessed automatically or by humans. First, automatic evaluation is usually based on evaluation metrics such as BLEU (Papineni et al., 2002), NIST, WNMf or METEOR (Anastasiou & Gupta, 2011). Metrics such as BLEU compare the MT output string with a human translation which is seen as "gold standard". However, these metrics ignore the source sentence as a reference and the fact that there might be more than one correct translation (Way, 2018). Second, human evaluation (also) requires the use of evaluation criteria (a brief overview of evaluation criteria provide Fiederer & O'Brien (2009)). When comparing raw MT output with human translations, the purpose of MT, e.g. whether MT is used to get the gist of a text or for publication purposes, is usually not taken into account. Only the latter would usually require post-editing.

A series of error typologies have been developed to assess the quality of machine-translated content. The Multidimensional Quality Metrics (MQM) error typology (Lommel et al., 2014) has been increasingly used and expanded for the evaluation of NMT (Klubička et al., 2018). The MQM framework provides a comprehensive typology of quality issues. This error typology contains standardized names and definitions of errors and has the flexibility of several assessment layers and their granularity. The MQM issues are organized in eight major dimensions: Accuracy, fluency, terminology, locale convention, style, verity, design, and internationalization (Lommel et al., 2014).

By the nature of design, the assessment of the quality of MT output is a post-trial evaluation and does not consider pre-trial expectations.

2 Research design and method

The research reported in this paper has several objectives. First, the research investigates the expectations of a group of postgraduate specialized translation students towards MT. This paper explores how previous experience with MT influences their expectations towards the overall quality of and error types found in MT output. Second, it seeks to examine the confirmation or disconfirmation of these expectations by an evaluation of two MT outputs.

This study makes a contribution to research on expectations towards MT by demonstrating that experience and expectations influence the use of MT systems and the evaluation of MT output. We applied a mixed-method approach, combining a quantitative questionnaire as well as MT output evaluation, i.e. error identification, error classification and correction of MT output.

2.1 Questionnaire

A questionnaire consisting of three parts with closed and open questions was distributed among the user group. The first part was designed to ascertain the respondents' translation experience, working languages (A, B and C language (AIIC, 2018)) and professional experience.

The second part of the questionnaire addressed the respondents' prior experience in MT use, including the frequency of and reasons for MT use. The participants were asked to state whether they use MT for professional, study or private purposes, which MT systems they use and for which

types of text. This part also elicited information on the respondents' forecast, normative and ideal expectations towards MT. The participants were asked to rank the quality-related issues and their frequency they would expect in MT output according to the MQM. All respondents had to state the most frequent errors they expect in MT output.

The third part of the questionnaire elicited information on the quality expectations and expected errors when using an MT system for two different texts. The students were asked to read the English source text. Afterwards they had to state their expectations towards the quality of the related MT output utilizing a five-point grade system (excellent, good, satisfactory, sufficient, useless). They had to rank the expected errors in the MT output according to the MQM. Second, they had to download a spreadsheet containing the MQM and TAUS Dynamic Quality Framework (DQF) (Görög, 2014). They compared the source and target text and identified (and corrected) errors in the MT output. Each error was assigned to an MQM error (sub)category and an error severity level on a five-point scale in the spreadsheet. The completed spreadsheets served as basis for the third step, which consisted in ranking the error types found in the MT output according to their frequency. By using the TAUS DQF and MQM for the error identification and classification task, we could compare their expectations with the evaluation result.

The questionnaire was circulated in early 2019. 79 students enrolled on a master's program in translation and focusing on specialized translation were recruited for this study. 32 individuals were excluded from the study because English was none of their working languages or they did not complete all the tasks.

2.2 Evaluation of MT output

The objective of the participants' evaluation of MT output in the third part of the questionnaire was to collect the error issues detected in raw MT output by the respondents. The evaluation was used for contrastive analysis of users' expectations towards error issues in MT output and the actual errors detected. It helped analyze the confirmation or disconfirmation of expectations.

The quality of the raw MT output was evaluated by the students based on the MQM error typology and the TAUS DQF. Prior to evaluation, they were familiarized with both frameworks.

The students were given two English source texts and their German MT outputs. The MT outputs used for evaluation were excerpts from British newspaper articles on a topic related to Austria. They comprised about 200 words each and were translated from English to German with the EU Council Presidency Translator (2019) platform. The study participants were provided with the source texts and the raw MT output as well as the MQM and TAUS DQF spreadsheet for both texts. The sentences in German were evaluated at the segment level in accordance with the MQM.

3 Results

3.1 Profile of the respondents

Of the final cohort of 47 respondents, 8 already worked as professional translators and 39 were novice translators. The majority (68%) of the respondents worked with German as A language, ahead of Italian (11%) and Russian, Hungarian, Polish, English and French. More than half of the participants (60%) stated that English was their B language, with German, Russian, Croatian and Japanese being the B language of the remaining respondents. The C languages were quite diverse, ranging from English (38%), French, Spanish, Slovakian, Italian, German to Greek and Romanian. Six respondents stated that they do not work with a C language. When asked about their translation experience, the majority (79%) indicated that they had translated more than 15 texts during their studies. The 8 students (17%) who had already worked as professional translators were active in the fields of engineering, social sciences and humanities.

3.2 Experience in MT use

About 62% of the respondents already had experience in MT use. Almost all of them (93%) reported that they use MT as part of their studies. More than two-thirds (69%) indicated that they use MT for private purposes and 31% of the respondents for professional purposes. When asked about the frequency of MT use in a professional, private or study context, 41% of the students indicated that they use MT for study purposes on a weekly basis and the remainder several times a year (19%) or several times a month (15%). For private purposes, they commented to use MT several times a year (31%), on a weekly basis (21%), on a daily basis (3%) or never (14%). For professional purposes, the respondents indicated that they never use MT (55%) or they use it several times a month (17%), on a daily basis (14%), several times a year or on a weekly basis (7% each).

Those experienced in MT use translated documents, e.g. reports or files (79%), ahead of websites (34%) or correspondence, e.g. e-mails (24%). Most of them reported that they use MT for translations from German into English and vice versa. They listed DeepL (69%) and Google Translate (59%) when asked about the MT system of choice. Another system mentioned was eTranslation. Among the MT systems which the respondents already tested but did not use frequently were Google Translate, the Facebook translator, Bing, Yandex and Babel.

The reasons for using MT included saving time (69%), getting the gist of a text (66%), consulting a reference (55%), avoiding repetitive work (31%), avoiding typing (21%) and avoiding research (3%).

3.3 Expectations towards MT quality

The participants expected MT to provide a raw translation, i.e. a first draft they can post-edit (53%) or a gist translation (38%) when using MT for study purposes. Only 5 respondents (11%) would want MT to provide immediately usable translations in a study context. For professional and private purposes, 21 respondents (45%) expected MT output to produce texts which can be used immediately without post-editing, i.e. they expected a final translation. For professional purposes, 15 respondents (32%) reported that they would use MT output as a draft translation. For private purposes, 24 respondents (51%) would use MT output only as a gist translation. This means that draft translations were more important in a study context, whereas gist purposes (to understand the meaning of the text) and final translations were more relevant in a private context.

When asked to rank their general expectations towards working with an MT system, 81% of the respondents ranked fast translation first. Proper functioning and intuitive use of the MT system ranked second among 60% of the respondents, whereas intuitive use still ranked third among 28% of the respondents. On ranks 4 to 6 the respondents predominately listed translation of different file formats, status feedback and accessibility of the MT system.

In response to the question about the expected quality-related issues in MT output, nearly a third (30%) of those surveyed ranked accuracy first while nearly one quarter (23%) ranked fluency first. Just over a third of those who responded ranked accuracy second, while approximately a fifth (21%) ranked fluency second. Terminology (30%) and style (23%) were the two main aspects on the third rank while locale conventions and style (23% each) had the highest number of responses on the fourth rank. Design and verity were mentioned predominantly on ranks 6 and 7.

3.4 Expectations towards error types and their (dis)confirmation

After having read the first source text (ST1), the respondents rated the expected quality of the related MT output (O1) with a grade ranging from excellent to useless. Almost half (49%) of the respondents expected the quality of the MT output to be sufficient, while 40% of those surveyed expected satisfactory MT output. Only a small number of the participants expected good quality (4%) or useless translations (6%). After having read O1 and after having identified, categorized and corrected the errors in the raw MT output, the participants rated the quality of O1 as follows: Sufficient (40%), useless (28%), satisfactory (23%) and good (9%). Thus, the number of useless grades increased significantly while the number of satisfactory and sufficient grades decreased.

The expected errors and their frequency in O1 were primarily related to fluency (38% on the first rank), accuracy (28% on the first rank, 32% on the second rank), style (23% on second rank) and terminology (21% on third rank). When compared to the errors reported, accuracy errors increased and fluency and verity errors decreased on rank 1, while fluency errors increased, and accuracy and terminology errors decreased on rank 2. Style errors increased slightly on rank 3 while locale convention errors increased on rank 4.

For the second source text (ST2), the students predominantly expected the MT output (O2) to be of sufficient quality (55%) or useless (26%). The other students reported that O2 would have satisfactory (13%) or good quality (6%). Compared to their expectations, they rated the actual translation to be of lower quality. The participants stated that O2 was useless (36%) or of sufficient quality (49%). This demonstrates that they expected the MT output to be of higher quality than later reported.

When asked about the expected error types in O2, well over half (64%) of the respondents ranked accuracy errors first and more than half (57%) ranked fluency errors second. Well under half of those surveyed (40%) ranked style errors third. After completing the MQM table, there was a significant increase in fluency errors and decrease of accuracy errors on rank 1 as well as a

significant increase in fluency errors on rank 2 and a slight increase in terminology errors. On rank 3, the students reported a higher number of accuracy errors and a smaller number of locale convention errors than expected.

Thus, both accuracy and fluency were the MQM error categories listed the most in all analyzed areas, i.e. the overall quality of MT output, the expected error types and the error types found. However, the data showed a slight shift of the accuracy and fluency categories between the expected and actual error types in both texts.

In summary, the majority of the participants expected the MT output to be of sufficient or inferior quality. Partly, the translations for both texts did not meet their expectations since they assessed the MT output of higher quality before and of lower quality after the evaluation.

There was a disconfirmation of the respondents' expectations towards the error types in MT output. For O1, the participants expected a higher frequency of fluency errors (on the first rank) before the evaluation. However, they reported a higher frequency of accuracy errors after the evaluation (62% on the first rank).

The expected error types in O2 mentioned by the students may be influenced by the outcome of the analysis of the error issues found in O1. As mentioned before, after having analyzed O1, the majority of the errors reported were related to accuracy (62% on the first rank). This is also reflected in the expected error issues reported for O2. Here, accuracy errors were expected by 64% of the respondents (on the first rank). There was a higher confirmation of their expectations towards the translation quality of O2. For O2, there were slightly more fluency and less accuracy errors (on the first rank) reported than expected.

This demonstrates that the participants in this study have rather low expectations towards the quality of the MT output. These expectations have been partly met, since the quality of both target texts translated with the MT system was reported to be lower than expected. This might also be the reason why the participants expected the second text to be of a slightly lower quality than the first one. This also means that there was a minor discrepancy between the pre-trial expectations and the errors found by the participants during evaluation. Moreover, this user group expected a higher frequency of some error types compared with the reported post-trial frequency.

4 Discussion

We focused on postgraduate translation students due to the documented competence profile of this user group. Their competence profile included translation, technological and revision competence (EMT, 2009). Therefore, we assumed that the students had a basic knowledge of MT systems, their advantages and disadvantages as well as post-editing. It was necessary to familiarize them with the rather complex MQM framework which required a certain amount of time.

Although this study is limited to a small number of participants, one NMT engine, the text type newspaper article and a certain language pair and direction, it revealed that participants use MT regularly or have used it at least once, especially freely available systems. DeepL was the most frequently used system among the translation students, ahead of Google Translate. We also saw that the users' previous experience with MT systems has an impact on future expectations towards similar systems. This is in accordance with Anastasiou & Gupta (2011), assuming that freely available, easily accessible MT which produces good-enough quality translations continues to be the MT system of choice for casual users who wish to translate websites or use MT for private purposes.

The expectations towards working with the MT system among the analyzed user group were that the system should work fast, function properly and can be used intuitively.

The majority of the respondents had considerable experience of MT use for study or private purposes. Almost half of the students (45% each) reported that they expect MT output, in professional and private contexts, to be useable immediately without any further editing. However, when they used MT as part of their studies, more than half of the respondents expected a raw translation they can post-edit rather than an immediately usable translation. Gist translations were more important in a private context. A possible explanation for this might be that the majority had already used MT output as a draft translation they post-edited. Based on our experience, translation students aim for producing high-quality translations. Therefore, they adapt the MT output to meet their ideal expectations. For private purposes, however, they seem to use MT output not as a pre-translation they can work on, but for languages they might not understand. Here, it might be more important to get the gist of the text rather than high accuracy and fluency. Thus, their expectations fall into the category of minimum tolerable. This finding

seems to be consistent with other research which found a dance of agency (Cadwell et al., 2018).

One interesting finding is that the students expected the MT output to be of rather low quality although they had used (general-purpose) MT before. This finding is contrary to previous studies which have suggested that those students expressing higher skepticism towards MT had the least exposure to it (Fulford, 2002) and that a negative attitude towards MT seems to be related to a lack of knowledge and (practical) experience (Gaspari, 2001). However, these studies focused on the students' opinions or attitudes, whereas this study addressed their previous MT experience in relation to their expectations as well as the confirmation or disconfirmation of their expectations. A possible explanation for the rather low expectations towards the quality of MT output is that students may be aware of the limitations of MT systems since they use it in their studies.

When we asked the students about their expectations towards the MT output quality, accuracy and fluency were ranked high. This suggests that accuracy and fluency made up translation quality for them. This finding was also reported by another study, where translators expected an MT engine to suggest correct translations, which may refer to correct target-language syntax as well as grammar and semantic equivalence to the source text (Lagoudaki, 2008).

With a small sample size and a focus on translation students (and not professional translators), caution must be applied, as the findings might not be generalizable to other user groups. However, MT-related tasks require other competences than the traditional profile of professional translators and additional competences than those acquired in translator training (Pym, 2013). Professional translators may also have limited practical exposure to MT and post-editing (Blagodarna 2018). In addition, a major issue with conceptualizing expectations is the sources of information or lack thereof used to form expectations: marketing communication by developers, mass media, training settings, word-of-mouth referrals, and prior experience with similar products. Service quality is not static but should be considered as a dynamic process (Boulding et al., 1993). Therefore, this study can only provide a small insight into user expectations of translation students at a certain point in time. In addition, students may not have identified all errors in the raw MT output. They may also lack critical evaluation of the MT output and would rather search for errors that human translators usually make (Sycz-Opoń & Gałuskina, 2017). Moreover, our analysis does not take account of intra-annotator or inter-annotator agreement when identifying and categorizing the errors of the MT output.

The aim of (neural) MT is to reach the fluency of human translations (Way, 2018). However, accuracy, e.g. whether the MT output imparts the meaning of the source text, seems to be a major concern of translation students for the texts analyzed in this study. NMT engines provide fluent and easily readable translations. However, these fluent translations may mislead users to think that the content is translated correctly, although the message may be completely wrong.

5 Conclusion

Translation should fulfil a specific purpose for the intended recipient in a certain context (Reiss & Vermeer, 1984). Therefore, this paper highlights the importance of paying attention to user expectations and not only to MT (quality) evaluation (by users). This article attempts to show that user expectations are crucial in translation, including processes in MT since they may help predict user interventions, such as pre- and post-editing. This paper argues that users' past experiences, expectations and (dis)confirmation of expectations frame human evaluation of MT. Therefore, users' expectations should be factored in when introducing MT services and novel approaches to MT.

Acknowledgment

This work has been partly funded by the European Union's Connecting Europe Facility under grant agreement no. INEA/CEF/ICT/A2016/1297953.

References

AIIC. 2018. *Working languages.* Retrieved from https://aiic.net/page/4004/what-are-working-languages-to-a-conference-interpreter/lang/1

Anastasiou, Dimitra and Rajat Gupta. 2011. Comparison of crowdsourcing translation with Machine Translation. *Journal of Information Science,* 37(6):637–659. https://doi.org/10.1177/0165551511418760

Blagodarna, Olena. 2018. Insights into post-editors'profiles and post-editing practices. *Tradumàtica: tecnologies de la traducció,* (16):35–51.

Boulding, William, Ajay Kalra, Richard Staelin, and Valarie A. Zeithaml. 1993. A Dynamic Process Model of Service Quality: From Expectations to

Behavioral Intentions. *Journal of Marketing Research*, 30(1):7–27. https://doi.org/10.2307/3172510

Cadwell, Patrick, Sharon O'Brien, and Carlos C. S. Teixeira. 2018. Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators. *Perspectives*, 26(3):301–321. https://doi.org/10.1080/0907676X.2017.1337210

EMT. 2009. *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved from https://ec.europa.eu/info/sites/info/files/emt_competences_translators_en.pdf

EU Council Presidency Translator. 2019. *EU Council Presidency Translator*. Retrieved from: https://translate2018.eu/

Fiederer, Rebecca and Sharon O'Brien. 2009. Quality and Machine Translation: A realistic objective? *JoSTrans*. (11):52–74. Retrieved from http://www.jostrans.org/issue11/art_fiederer_obrien.pdf

Fulford, Heather. 2002. Freelance translators and machine translation: An investigation of perceptions, uptake, experience and training needs. In *6th European Association of Machine Translation*, 117–122. Retrieved from http://www.mt-archive.info/EAMT-2002-Fulford.pdf

Gaspari, Federico. 2001. Teaching Machine Translation to Trainee Translators: A Survey of Their Knowledge and Opinions. In *Proceedings of the MT Summit VIII*, Santiago de Compostela, Spain, 35–44.

Gaspari, Federico, Hala Almaghout, and Stephen Doherty. 2015. A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, 23(3):333–358. https://doi.org/10.1080/0907676X.2014.979842

Görög, Attila. 2014. Dynamic Quality Framework: quantifying and benchmarking quality. *Tradumàtica: tecnologies de la traducció*, (12):443–454. https://doi.org/10.5565/rev/tradumatica.66

Higgs, Brownyn, Michael Jay Polonsky, and Mary Hollick. 2005. Measuring expectations: forecast vs. ideal expectations. Does it really matter? *Journal of Retailing and Consumer Services*, 12(1):49–64. https://doi.org/10.1016/j.jretconser.2004.02.002

Klubička, Filip, Antonio Toral, and Víctor M. Sánchez-Cartagena. 2018. Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian. *Machine Translation*, 32(3):195–215.

Lagoudaki, Elina 2008. The value of machine translation for the professional translator. *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas*, 262–269.

Lommel, Arle, Hans Uszkoreit, and Aljoscha Burchardt. 2014. Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica: tecnologies de la traducció*, (12):455–463.

Moorkens, Joss and Andy Way. 2016. Comparing Translator Acceptability of TM and SMT outputs. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, 141–151.

Olson, Jerry and Philip A. Dover. 1979. Disconfirmation of consumer expectations through product trial. *Journal of Applied Psychology* (64):179–189. https://doi.org/10.1037/0021-9010.64.2.179

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, 311–318. https://doi.org/10.3115/1073083.1073135

Pym, Anthony. 2013. Translation skill-sets in a machine-translation age. *Meta: Journal des traducteurs/Meta: Translators' Journal*, 58(3):487-503.

Reiss, Katharina and Hans J. Vermeer. 1984. *Grundlegung einer allgemeinen Translationstheorie*. *Linguistische Arbeiten: Vol. 147*. Tübingen, Max Niemeyer.

Koskinen, Kaisa and Minna Ruokonen. 2017. Love letters or hate mail? Translators' technology acceptance in the light of their emotional narratives. In Dorothy Kenny (ed.). *Human issues in translation technology*, Routledge, 26-42.

Specia, Lucia, Craig Saunders, Marco Turchi, Zhuoran Wang, and John Shawe-Taylor. 2009. Improving the confidence of machine translation quality estimates. *Proceedings of the Twelfth Machine Translation Summit (MT Summit XII)*, 136–143.

Sycz-Opoń, Joanna and Ksenia Gałuskina. 2017. Machine Translation in the Hands of Trainee Translators – an Empirical Study. *Studies in Logic, Grammar and Rhetoric*, 49(1):195–212. https://doi.org/10.1515/slgr-2017-0012

Way, Andy. 2018. *Quality expectations of machine translation*. Retrieved from http://arxiv.org/pdf/1803.08409v1