Fine-grained human evaluation of NMT applied to literary text: case study of a French-to-Croatian translation

Marta Petrak,* Mia Uremović,* Bogdanka Pavelin Lešić*

* Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb
mpetrak@ffzg.hr
uremovic.mia@gmail.com
bpavelin@ffzg.hr

Abstract

Even though neural machine translation (NMT) has demonstrated phenomenal results and has shown to be more successful than previous MT systems, there is not a large number of works dealing with its application to literary text. This results from the fact that literary texts are deemed to be more complex than others because they involve more specific elements such as idiomatic expressions, metaphor, a specific author's style, etc. Regardless of this fact, there is a growing body of research dealing with NMT applied to literary texts, and this case study is one of them. The goal of the present paper is to conduct an in-depth, fine-grained evaluation of a novel translated by Google Translate (GT) in order to reach detailed insights into NMT performance on literary text. In addition, the paper aims to include for the first time, to the best of our knowledge, the French-Croatian language combination.

1. Introduction

Numerous studies have demonstrated that neural machine translation (NMT) outperforms previous MT systems (e.g. Bentivogli et al., 2016; Burchardt et al., 2017; Klubička et al., 2018; Hansen, 2021). This has been demonstrated for a number of various text types, among which literary texts are the least represented due to their specificities such as lexical richness, metaphorical and idiomatic elements (e.g. Toral and Way, 2018). Literary translation is also usually considered to be more complex than technical translation because it includes elements such as writer's individual style (Hadley, 2020).

Due to these facts, literary texts are still perceived to be "the greatest challenge for MT" (Toral and Way, 2018). Some more pessimistic authors even claim that "there is no prospect of machines being useful at (assisting with) the translation of [literary texts]" (Toral and Way, 2018). While the use of machine translation followed by the post-editing phase is a widespread practice generally speaking, it has not yet become a permanent fixture in literary translation (Besacier, 2014).

In spite of this fact, there has been a growing interest in applying MT to literature, which can be seen, for example, in the fact that there is a workshop on computational linguistics for literature organised by ACL since 2012[1]. Moreover, the French-speaking world has seen the creation of an observatory for MT (*Observatoire de la traduction automatique*) by the ATLAS[2] association in December 2018 to follow the development of MT application to literary text[3].

Even though studies that analyse the application of MT to literary text are less numerous than those applying MT to other types of text, they are not inexistant. Hansen's (2021) paper brings a detailed and up-to-date overview of the works dealing with MT of literary texts. The first literary text translated by MT was done by Besacier (2014), and it comprised an essay translated from English to French. A number of languages have already been covered by various studies of MT to literary text, among which Slavic (e.g. Slovene, Kuzman et al., 2019), Romance (e.g. Catalan, Toral and Way, 2018, French, Besacier, 2014), Hansen, 2021), Germanic (English, in a number of papers; German, Matusov, 2019); Scottish Gaelic and Irish, Ó Murchú, 2019), etc.

2. Goal of the paper

The goal of this case study is to go beyond the overall performance of NMT on literary text and to provide an extensive, in-depth human analysis of its results. In order to do so, we will, firstly, produce a MT of a French novel and, secondly, compare that translation with a human translation of the same text. The human translation will be done by a student in translation from French into Croatian as part of her Master's thesis, and the analysis will be carried out by two human evaluators, the student and an experienced professional translator.

In addition to providing an in-depth analysis of the translation of a literary text done by MT, our case study is the first one to pair, to the best of our knowledge, a large Romance language, French, with Croatian[4], a smaller scale language rich in morphology.

The rest of the paper is structured as follows: in Section 3 we describe the methodology used. Section 4 is the central part of the paper, as it sums up the results of our analysis combined with a number of specific

[1] Cf. e.g. https://aclanthology.org/events/clfl-2020/.
[2] ATLAS stands for *Association pour la promotion de la traduction littéraire* (Association for the promotion of literary translation), https://www.atlas-citl.org/.
[3] https://www.atlas-citl.org/lobservatoire-de-la-traduction-automatique/
[4] Croatian is the official language of the Republic of Croatia and of the EU., but is also spoken in Bosnia and Herzegovina, Montenegro, etc. It has approximately 5.6 million native speakers worldwide. Cf. https://www.european-language-grid.eu/ncc/ncc-croatia/.

examples from the corpus. In Section 5 we bring some concluding remarks and recommend some further steps.

3. Methodology

In order to conduct our analysis, we have chosen a novel, which is "arguably the most popular type of literary text" (Toral and Way, 2018). Our corpus comprises the first eight chapters of the novel *La traduction est une histoire d'amour* (*Translation is a Love Affair*) written by Jacques Poulin, a contemporary Canadian author. It comprises a total of 8,347 words. The original text, written in French, is first translated by GT, and subsequently by a human translator. The MT is analysed in detail by two evaluators, after which the two translations are compared.

Hansen (2021) argues that evaluation of texts produced by MT still remains a major obstacle. More precisely, if BLEU (Papineni et al., 2002) is the most widely used automatic metric, it has to be taken with caution in case of literary texts (*ibid*.). Papineni et al. (2002) argue that human evaluations of MT are "extensive" and therefore usually more fine-grained than automatic ones, but the authors also point to their expensiveness.

In our case study, we present a quantitative and qualitative analysis of errors. We base our methodology on the one developed by Pavlović (2016). Pavlović (*ibid.*) also argues that in the literature there is not a single classification of translation errors that all authors would agree upon, so she makes her own classification based upon extant ones by a number of previous authors and some specificities of the corpus. Her study (2016) included only non-literary texts, newspaper reports, public opinion reports and EU legal documents (opinions and decisions), a total of 3,406 words. Still, Pavlović's (2016) methodology was developed with the goal of comparing MT done by GT and human translation, and it takes into account some specificities of the Croatian language such as a rather free word order, abundance of inflection and morphological complexity. It should be emphasized that Pavlović's (2016) study was conducted before GT used NMT for Croatian, which is available today [5] and is the technology used for the analysis presented in this case study.

The analysis of errors conducted for this paper follows that given by Pavlović (2016), with only minor alterations. For example, the sub-category (D.c), 'numbers', is not present in the machine translation of the chosen text and is hence not part of this analysis.

4. Results and analysis

4.1. Fine-grained human evaluation

Our analysis has demonstrated that GT has provided a very satisfactory translation generally speaking, and some of its solutions were even better than the ones provided by the human translation in the cases where there was a possible choice between a general word and its more suitable or literary synonym.

Below we first bring a table with a general presentation of errors found in the MT.

Error category	%
Morphosyntax	55.3
Lexicon	32.1
Spelling	7
Other	5.6

Table 1: Classification of general error types produced by MT.

Table 1 demonstrates that morphosyntactic errors visibly make the most frequent error type in our corpus, i.e. more than half of the total number of errors. These are followed by errors in lexical choice. In Table 2 (below) we bring a detailed list of error types found in our corpus.

Error type	%
C.a. congruence	39.3
B.a. lexical choice	18.8
C.c. word order / order of phrase constituents	10.9
B.c. idiomatic expressions	7.5
B.b. term or title	5.8
C.b. verbal forms / tenses	5.2
A.a. punctuation	4.5
A.b. capital letters	2
D.a. not translated	2
D.b. omissions	1.9
D.d. format, etc.	1.6

[5] Cf. https://translate.google.com/intl/hr/about/languages/.

A.c. other spelling errors	0,5
D.c. numbers	0

Table 2: Detailed breakdown of error types found in the corpus.

4.1.1. Morphosyntactic errors

According to our analysis, the most common errors done by GT are morphosyntactic errors, more specifically congruence errors, representing 39.3%. This type of errors most frequently have to do with grammatical gender. Here is an example:

original	GT	human translation
La meilleure ***traductrice*** *du Québec*	*Najbolji* ***prevodilac*** *u Quebecu*	*Najbolja* ***prevoditeljica*** *u Québecu.*

Table 3: Example of congruence error.

In the above example, *traductrice* 'female translator' is translated by GT as *prevoditelj* 'male translator' even though both French and Croatian are marked for gender, and even though there is a ready-made solution in Croatian, *prevoditeljica* 'female translator'. The problem here is probably the fact that GT uses English as a sort of pivot or intermediate language (e.g. Ljubas, 2018) when translating between French and Croatian[6], that do not share as large a corpus of texts as they do with English individually.

This is a frequent error produced by GT in the corpus, i.e. not marking whatever has to do with the narrator, who is a woman, as female, but leaving male nouns, adjectives etc., which we also attribute to translating via English: e.g. *Je raccrochai* is translated as *Spustio* (masc.) *sam slušalicu* instead of *Spustila* (fem.) *sam slušalicu / Poklopila* (fem.) *sam*.

In other words, it can be said generally that our analysis has demonstrated that GT had no problems, for example, with the Croatian rich nominal case system and general subject-verb or noun-adjective agreement. This is in line with findings from the literature that neural systems have been found to make fewer morphological, lexical and word-order errors (e.g. Burchardt, 2017). What was a problem, however, in the category of morphosyntactic errors is recognizing the narrator as a female, and consequently translating all her attributes and making all the agreements in the feminine gender. This is a feature of the text that extends beyond sentence level and permeates the entire discourse of the novel. In some French sentences, this difference between masculine and feminine gender cannot be seen, for example in the present tense or in the past tense (*passé compose*) formed with the auxiliary verb to have (*avoir*). In Croatian, the same goes for the present tense, but the past tense always shows agreement with the subject in gender. The large number of errors in this category undoubtedly stems from the use of English as a pivot language.

4.1.2. Lexical errors

The next most represented category are lexical errors (32.1%), listed in the table below.

original	GT	human translation
Eh bien, c'était le ***portrait tout craché*** *de ma mère.*	*Pa, to je bila* ***pljuvačka slika*** *moje majke.*	*E pa to je* ***pljunuti portret*** *moje majke.*
Les ***ouaouarons****, affolés, …*	*Uplašeni* ***bikovi žabe*** *…*	***Žabe*** *su se preneražene …*
Je suis sur la route parce que ma ***maîtresse*** *ne peut plus s'occuper de moi, (…)*	*Na putu sam jer se moja* ***ljubavnica*** *više ne može brinuti o meni*	*Na ulici sam jer se moja* ***vlasnica*** *više ne može brinuti o meni, (…)*
Ma mère et ma grand-mère ***reposaient*** *derrière l'église …*	*Moja majka i baka* ***odmarale su se*** *iza crkve …*	*Moja majka i baka* ***bile su pokopane*** *iza crkve …*
… dans l'herbe jonchée de ***feuilles mortes****.*	*… u travi posutoj* ***mrtvim lišćem****.*	*… travi prekrivenoj* ***suhim lišćem****.*
J'étais très heureuse, ***presque sur un nuage****, (…)*	*Bio sam vrlo sretan, skoro* ***na devetom oblaku****, (…)*	*Bila sam sretna,* ***gotovo u sedmom nebu****, (…)*
Les maudites ***algues****…*	*Proklete morske* ***alge****…*	*Proklete* ***alge****…*

Table 3: Examples of lexical choice errors.

[6] This has been claimed generally as a feature of GT that it uses when translating between any pair of languages. A Google spokesperson has admitted that Google Translate uses English for „bridging“ between languages with fewer resources. See https://algorithmwatch.org/en/google-translate-gender-bias/; cf. https://www.circuitmagazine.org/chroniques-126/sur-le-vif-126/google-uses-english-as-a-pivot-language.

Errors is this category concern the following: 1) single-word polysemy, 2) idiomatic expressions, 3) calques from English.

With respect to single-word polysemy, GT has, for instance, erroneously translated *maîtresse* ‘owner’ (of a cat) as ‘lover’. It also translated *reposaient* ‘rested’ as *odmarale su se* ‘were having a rest’ instead of *bile su pokopane*, which is used in the context of the dead buried in a graveyard. Furthermore, it translated *algues* as *morske alge* ‘sea algae’, which is an incorrect specification stemming from the fact that algae are usually related to the sea, but algae in the story, however, come from a pond.

As for idiomatic expressions (7% of total errors), GT rendered *le portrait craché* ‘spitting image’ as **pljuvačka slika* instead of *pljunuti portret*. It clearly calqued the expression *être sur un nuage* ‘be on cloud nine’ on English and translate dit as **biti na devetom oblaku*, which does not exist in Croatian, and should be translated as *na sedmom nebu* ‘lit. on seventh sky’. The noun phrase *feuilles mortes* is litterally translated as **mrtvo lišće* instead of *suho lišće* ’lit. dry leaves’, etc.

There are several instances of calquing from English, such as in the example of *ouaouarons*, animals known in English as American bullfrogs, which are litterally translated as *bikovi žabe* ‘bulls-frogs’, and for which we would suggest the translation *žabe* due to the fact that the particular species is irrelevant to the plot.

4.1.3. Other errors

In the category of capital letters, GT had difficulties rendering street names, which appeared in the text several times. Examples such as *609, rue Richelieu* were rendered by GT as *609, ulica Richelieu*, where all the individual elements are correctly translated, but the street name as a whole should be written as *Ulica Richelieu 609*, which is a conventional way of writing street names in Croatian.

Another interesting error concerns proper names. Let us cite two examples: *Marine* and *Chaloupe*. Marine, the name of the main character and narrator, is sometimes translated by GT as *marinac* ‘Marine, i.e. member of an elite US fighting corps’. In addition to the same form, the English word is always capitalised, so that could be another reason for such a translation. *Chaloupe*, on the other hand, is the name of the cat that appears several times in the text. It is derived from the common noun *chaloupe* denoting a type of boat. GT translated the noun as *čamac* ‘boat’, making it a common noun and even leaving out the capital letter.

Bentivogli et al. (2016) and Toral and Sánchez Cartagena (2017) found that NMT improves notably on reordering and inflection than PBMT. In the case of Poulin’s novel translated and analysed in this paper, there were generally very few problems with inflection, and word / constituent order represented only 10% of all the errors. What our analysis seems to point to is the fact that using English as a pivot language is the source of a large number of errors, and that using language-pair specific language corpora could arguably give better results in translating between two languages of which neither is English. This would also probably have a positive effect on the translation of culturally specific elements such as spelling and writing of toponyms (e.g. street names). Furthermore, our analysis also demonstrates that more improvement should be done in the detection and translation of polysemy and idiomatic expressions.

4.2. BLEU evaluation

In addition to a fine-grained human translation, BLEU score was also calculated using the interactive BLEU score evaluator[7] available via the Tilde platform. BLEU score is based on the correspondence of the MT output and the reference human translation.

Overall cumulative BLEU score for the literary text analysed in our case study was 5.49, which would suggest very poor MT quality. As a reference, BLEU scores of 30 to 40 are considered to be “understandable to good translations”, while those of 40 to 50 are “high quality translations”[8]. Here is the breakdown of the BLEU score:

Type	1-gram	2-gram	3-gram	4-gram
Individual	21.92	5.86	2.79	2.54
Cumulative	21.92	11.33	7.10	5.49

Table 4: Results of automatic BLEU evaluation.

In other available case studies dealing with MT of a literary text, BLEU scores show significant variation. In the case of a translation of a literary essay from English into French (Besacier and Schwartz, 2015), BLEU score was around 30. In another case study dealing with English literary texts translated into Slovene, BLEU scores varied from 1.73 to 30 depending on the texts on which the MT model was trained (Kuzman et al., 2019). Toral and Way (2018) obtained BLEU scores of around 30 for English-to-Catalan translations of 12 English

[7] https://www.letsmt.eu/Bleu.aspx.
[8] https://cloud.google.com/translate/automl/docs/evaluate

novels by PBSMT and NMT systems, where NMT outperformed PBSMT.

Unlike the results obtained by Kuzman et al. (2019) in their study of a literary translation from English into Slovene, a language genetically very close to Croatian, where "there were no sentences that would not need postediting", in our case study there were a number of sentences entirely correctly rendered by GT, i.e. that would be publication ready.

In any case, it should be borne in mind that BLEU automatic evaluation metric was calculated with respect to a single human translation, and that it cannot represent the "real quality" of MT output. In that sense, Hansen (2022) notes, for instance, that two MT models used in his case study had a similar BLEU score in spite of the fact that the first one produced correctly translated words in incomprehensible sentences, while the second one generated correct sentences with words that semantically did not correspond the lexical field of the translated literary text. This is one of the reasons why we would not entirely agree that the translation provided by GT analysed in this paper is irrelevant or "useless", as it would be classified due to its BLEU score inferior to 10 (cf. footnote n° 8).

In addition, it should be noted that some authors claim that morphological richness of the Croatian language could raise problems for BLEU evaluation due to the fact that each Croatian noun has approximately 10 different word forms, which are considered by BLEU to be 10 different words, and not 10 different word forms of a single lemma (cf. Seljan et al., 2012). This could result in lower BLEU scores.

5. Conclusion

This case study is a contribution to a growing number of papers dealing with applying (N)MT to literary text, which has been thought of until only recently as a domain that could not be translated by MT. Various authors have, however, demonstrated the usefulness of using MT in literary translation. Some (e.g. Besacier and Schwartz, 2015) even argue that MT of literary text may even be of interest for all participants of the translation chain from editors, through readers to authors and translators.

Our analysis has demonstrated that there was a total of 738 errors in the text produced by GT, largely falling into two groups: morphosyntactic (around 55%) and lexical choice (around 32%) errors. While the morphosyntactic errors largely concerned errors in congruence stemming probably from the usage of English as a pivot language between French and Croatian, the lexical choice errors had mostly to do with polysemy, idiomatic expressions and calques.

Let us now compare our results with those from other existing works on MT of literary texts involving either of the two languages from this case study, Croatian or French. Hansen (2022), who analysed English-to-French translations of fantasy books, observed that, generally speaking, the MT output was rather literal and it produced mostly lexical errors, as well as errors related to determiners and syntax. While Hansen (*ibid.*) does not provide further details, we can generally say that in our French-to-Croatian literary translation morphosyntactic errors were by 20% more present than lexical errors, which is different than what he found in the English-French language pair. Furthermore, Hansen (*ibid.*) was surprised to note that the specific vocabulary related to the fantasy series in question was respected almost entirely, which is probably due to the training of the MT model on texts written by the same author. This is one of the reasons why Hansen (2022) suggests that personalized MT systems should be introduced in literary translation for translating specific authors' styles.

In another paper, involving Slovene, a language closely related to Croatian, and analysing translation of literary texts from English, Kuzman et al. (2019) observe that "error analysis (…) revealed various punctuation errors, wrong translations of prepositions and conjunctions, inappropriate shifts in verb mood, wrong noun forms and co-reference changes". The authors emphasize the presence of numerous semantic errors, "especially in connection with idioms and ambiguous words". In this case, more detailed data is also lacking, but we can generally conclude that this study also differs from ours in that semantic errors are definitely not the leading error type in our French-to-Croatian translation. Interestingly, Kuzman et al. (2019) also found that GNMT assigned the wrong gender to the main character, just as happened in our case, as mentioned in 4.1.1.

We can conclude that in the French-to-Croatian GT of the novel analysed in this text, morphosyntactic errors (55.3%) are the most represented ones, followed by various lexical errors (32.1%). These results are somewhat different from what was observed in earlier extant studies dealing with MT of literary texts from English to French and English to Slovene.

Even though BLEU score was only 5.49, indicating very poor translation quality which should be deemed as useless, we believe that the GT output would be useful to some extent to translators translating Poulin's novel from scratch. Further analyses should be made however in order to analyse whether GT trained on French and Croatian corpora would amount to better results than GT that uses English as pivot. Furthermore, it should also be studied how much post-processing effort is needed to correct errors of GT in comparison to translation from scratch in the French-to-Croatian language combination.

6. References

Luisa Bentivogli, Arianna Bisazza, Mauro Cettolo, and Marcello Federico. 2016. Neural versus Phrase-Based Machine Translation Quality: a Case Study. In: J. Siu, K. Duh and X. Carreras, eds., *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 257–267, Association for Computational Linguistics, Austin, Texas.

Laurent Besacier and Lane Schwarz. 2015. Automated Translation of a Literary Work: A Pilot Study. In: A. Feldman, A. Kazantseva, S. Szpakowicz and C. Koolen, eds., *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 114–122. Association for Computational Linguistics, Denver, Colorado.

Laurent Besacier. 2014. Traduction automatisée d'une œuvre littéraire : une étude pilote. In: P. Blanche, F. Béchet and B. Bigi, eds., *Actes du 21ème Traitement Automatique des Langues Naturelles*, pages 389-94. Association pour le Traitement Automatique des Langues, Marseille. https://hal.inria.fr/hal-01003944

Marija Brkić, Sanja Seljan and Maja Matetić. 2011. Machine Translation Evaluation for Croatian-English and English-Croatian Language Pairs. In: B. Sharp, M. Zock, M. Carl, A. L. Jakobsen, eds., *Proceedings of the 8th International NLPCS Workshop: Human-Machine Interaction in Translation,* pages 93-104. Copenhagen: Copenhagen Business School.

Aljoscha Burchardt, Vivien Macketanz, Jon Dehdari, Georg Heigold, Jan-Thorsten Peter, and Philip Williams. 2017. A Linguistic Evaluation of Rule- Based, Phrase-Based, and Neural MT Engines. *The Prague Bulletin of Mathematical Linguistics*, 108:159–170.

Margot Fonteyne, Arda Tezcan, and Lieve Macken. 2020.

Literary Machine Translation under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on Document Level. In: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Delreck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk and S. Piperidis, eds., *Proceedings of the 12th Conference on Language Resources and Evaluation*, pages 3790–3798, European Language Resources Association, Marseille.

James Hadley. 2020. Traduction automatique en littérature : l'ordinateur va-t-il nous voler notre travail. *Contrepoint*, 4:14–18. https://www.ceatl.eu/wp-content/uploads/2020/12/Contrepoint_2020_04_article_04.pdf

Damien Hansen. 2022. La traduction littéraire automatique : Adapter la machine à la traduction humaine individualisée. https://hal.archives-ouvertes.fr/hal-03583562/document

Damien Hansen. 2021. Les lettres et la machine : un état de l'art en traduction littéraire automatique. In: P. Denis, N. Grabar, A. Fraisse, R. Cardon, B. Jacquemin, E. Kergosien and A. Balvet, eds., *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles*, Vol. 1, pages 61–78. ATALA, Lille.

Filip Klubička, Antonio Toral and Víctor M. Sánchez- Cartagena. 2017. Fine-Grained Human Evaluation of Neural Versus Phrase-Based Machine Translation. *The Prague Bulletin of Mathematical Linguistics*, 108:121-132. https://arxiv.org/abs/1706.04389

Taja Kuzman, Špela Vintar and Mihael Arčan. 2019. Neural Machine Translation of Literary Texts from English to Slovene. In: J. Hadley, M. Popović, H. Afli, and A. Way, eds., *Proceedings of the Qualities of Literary Machine Translation*, pages 1–9, European Association for Machine Translation, Dublin. https://aclanthology.org/W19-7301

Rudy Loock. 2018. Traduction automatique et usage linguistique : une analyse de traductions anglais-français réunies en corpus. *Meta* 63(3):786–806. https://doi.org/10.7202/1060173ar

Sandra Ljubas. 2018. Prijelaz sa statističkog na neuronski model: usporedba strojnih prijevoda sa švedskoga na hrvatski jezik. *Hieronymus*, 5:72–79. https://www.bib.irb.hr/978980

Evgeny Matusov. 2019. The Challenges of Using Neural Machine Translation for Literature. In: J. Hadley, M. Popović, H. Afli, and A. Way, eds., *Proceedings of the Qualities of Literary Machine Translation, Machine Translation*, pages 10-19, European Association for Machine Translation, Dublin. https://aclanthology.org/W19-7302.pdf

Eoin P. Ó Murchú. 2019. Using Intergaelic to pre-translate and subsequently post-edit a sci-fi novel from Scottish Gaelic to Irish. In: J. Hadley, M. Popović, H. Afli & A. Way, eds., *Proceedings of the Qualities of Literary Machine Translation*, pages 20–25, European Association for Machine Translation, Dublin. https://aclanthology.org/W19-7303

Kishore Papineni, Salim Roukos, Todd Ward and Wei- Jing Zhu. 2016. Bleu: a Method for Automatic Evaluation of Machine Translation. In: P. Isabelle, E. Charniak and D. Lin, eds., *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318. Association for computational linguistics, Philadelphia, Pennsylvania, USA. doi.org/10.3115/1073083.1073135

Nataša Pavlović. 2017. Strojno i konvencionalno prevođenje s engleskoga na hrvatski: usporedba pogrešaka. In: D. Stolac and A. Vlastelić, eds., *Jezik kao predmet proučavanja i jezik kao predmet poučavanja*, pages 279–295, Srednja Europa, Zagreb.

Jacques Poulin. 2006. *La traduction est une histoire d'amour*. Leméac/Actes Sud, Montreal.

Sanja Seljan, Marija Brkić and Tomislav Vičić. 2012. BLEU Evaluation of Machine-Translated English-Croatian Legislation. In: N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis, eds., *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul, Turkey.

Antonio Toral and Andy Way. 2018. What Level of Quality Can Neural Machine Translation Attain on Literary Text? In: J. Moorkens, S. Castilho, F. Gaspari, S. Doherty, eds., *Translation Quality Assessment. Machine Translation: Technologies and Applications*, Vol 1, pages 263-287. Springer, Cham. https://doi.org/10.1007/978-3-319-91241-7