A Human-machine Cooperation Protocol for Machine Translation Output Edit Annotation

Felipe de Almeida Costa
Adriana S. Pagano
Thiago Castro Ferreira
Wagner Meira Jr.

revista.tradumàtica
tecnologies de la traducció

Costa, Felipe Almeida
Federal University of Minas Gerais
felipealco@dcc.ufmg.br
ORCID: 0000-0002-5273-6621

Pagano, Adriana S.
Federal University of Minas Gerais
apagano@ufmg.br
ORCID: 0000-0002-3150-3503

Ferreira, Thiago Castro
Federal University of Minas Gerais
Thiagocf05@ufmg.br
ORCID: 0000-0003-0200-3646

Meira Jr., Wagner
Federal University of Minas Gerais
meira@dcc.ufmg.br;
ORCID: 0000-0002-2614-2723

Abstract

We report on a study exploring automatic edit annotation in a post-editing corpus with a new method for computing edit types. We examine edit type association with quality scores assigned to the machine translation output and the post-edited texts. Finally, we account for shortcomings in our method and point out edit types worth leveraging.

Keywords: machine translation; human post-editing; automatic error analysis; human-machine cooperation

Resum

Presentem un estudi que explora la detecció automàtica d'errors en un corpus de postedició amb un mètode inèdit per calcular tipus d'edició. Examinem la seva associació amb les puntuacions de qualitat assignades a la producció de traducció automàtica i als textos posteditats. Finalment, expliquem les deficiències del nostre mètode i assenyalem els tipus d'edició que val la pena aprofitar.

Paraules clau: traducció automàtica; postedició humana; anàlisi automàtica d'errors; cooperació persona-ordinador

Resumen

Presentamos un estudio que explora la detección automática de errores en un corpus de posedición con un método novedoso para computar los diferentes tipos de corrección. Examinamos su asociación con la puntuación asignada a la calidad de los resultados de la traducción automática y de los textos poseditados. Por último, analizamos algunos defectos de nuestro método y destacamos los tipos de correcciones que conviene aprovechar.

Rebuda: 18 de febrer de 2021 / Acceptació: 16 de juliol de 2021 / Publicació avançada: 15 de desembre de 2021

Revista Tradumàtica. Número 19 (desembre 2021).
Cita recomanada: FELIPE DE ALMEIDA COSTA; ADRIANA S. PAGANO; WAGNER MEIRA, JR.; CASTRO FERREIRA, THIAGO; (2021). «A Human-machine Cooperation Protocol for Machine Translation Output Edit Annotation». Revista Tradumàtica. Tecnologies de la Traducció, 19, 148-170.
<https://doi.org/10.5565/rev/tradumatica.292>

ISSN 1578-7559

Palabras clave: traducción automática; posedición humana; análisis automático de errores; cooperación persona-máquina

1. Introduction

As technology became more accessible and part of daily tasks, the general idea was that machines would eventually replace humans for a significant number of tasks. However, as machines became more efficient and smarter, other concerns arose such as shared labour, privacy, transparency and ethics. In this context, human-machine cooperation has now become a very valuable feature of computational systems. Having the human as part of the system makes it more trustworthy and more efficient since there are some human aspects that cannot be reproduced by machines.

In the field of translation, one form of human-machine cooperation is post-editing (PE), which entails humans analysing and correcting machine translation (MT) errors. Delving deeper into analysing MT errors is useful, not only to evaluate an MT system but also to improve it by focusing on its shortcomings. When performed by humans, error analysis is a time-consuming task and subject to varying levels of agreement when there is more than one annotator. In contrast, automatic error analysis may initially be faster and more consistent, but prone to misclassifying low frequency errors which give poor feedback for machine learning, or fail to classify certain errors, especially if error typologies to be classified are very fined grained.

In this article we introduce an efficient protocol for classifying edit operations using human-machine cooperation. By efficient, we mean applying automatic methods and using human resources only when necessary. Our approach is directly related to Error Analysis, however, we have chosen to refer to it as Edit Evaluation, because not all edits in our corpus can be characterised as errors, as is the case of dialectal preferences and meaning explicitation. Computationally, there is no distinction between an error and an edit, therefore methods for detecting errors and edits are interchangeable. In order to comprise both errors and editor choices we have used the umbrella term “edits”: we refer to edit operations as edit types and analyse them as implicit translation errors (Popović, 2018). We performed a semi-automatic experiment in a post-editing corpus previously created by us (Costa et al., 2020) and then developed a set of rules to be followed by an automated system. A very important part of this protocol is measuring how much human effort each edit type requires so that we can assign subtasks to humans efficiently and avoid repetitive and unnecessary work (Chatterjee et al., 2019).

The English source texts were extracted from the WebNLG corpus (Gardent et al., 2017) and automatically translated into Portuguese using a Neural Machine Translation (NMT) system (DeepL). Post-edits were then carried out in an experiment using native speakers of Brazilian Portuguese, and the machine-translated and post-edited texts were then submitted for human evaluation. We started our analysis by defining categories of edit types, drawing on Popović’s automatic error classification (Popović, 2011). We

computed the distribution of edit types automatically by means of a new method we refer to as "Editing Brackets", then we computed the association of each edit type with the quality score assigned to the machine translation output and the post-edited texts. Finally, we focused on our results to check the accuracy of our method and account for shortcomings in our classification, pointing out edit types that required more fine-grained human inspection.

The remainder of this article is organized as follows: section 2 briefly provides the motivation behind our study; section 3 details our method; section 4 presents the analysis carried out; section 5 presents a discussion of our results; and section 6 summarises our study and main findings, followed by our suggestions to expand further. The source texts, automatic and post-edited translations, as well as the findings of our analysis, are publicly available in our repository[1].

2. Motivation

"Error analysis" is a topic in machine translation research encompassing detection, classification and explanation of shortcomings in the text yielded by a translation model. Subsumed under this heading are tasks investigating not only wordings that are deemed unacceptable in the target language (strictly speaking "errors"), but also acceptable wordings introduced by post-editors, which are nevertheless considered as indicators of linguistics aspects worthy of improvement from a human assessment perspective. While some studies pursue linguistically-based taxonomies of errors, others focus on error detection by mapping edits found when contrasting the original machine output and the text obtained after a post-editing task. Whatever the purpose, error analysis is a fundamental step, not only to assess and improve machine translation systems, but also to examine ways of enhancing human-machine cooperation through post-editing machine translation output (Popović, 2018). By considering each edit operation as an actual correction of the machine's failure to produce an output closely matching human translation, we can obtain feedback data to improve the computational model that was used to carry out the translation in the first place. While human error analysis provides deeper insights into linguistic shortcomings in the machine's performance, it is not very feasible for large datasets; automatic error analysis, on the other hand, allows for processing larger amounts of text, but the results yielded still have to be manually inspected to explain post-edits not easily accounted for as basic language problems.

Regarding assessment in the machine translation workflow, tools have been reported to tap and evaluate human post-editing. One such tool is PET (Aziz, Souza & Specia, 2012), which allows for both collecting post-editing data and computing post-editing effort based on several distinct metrics. The tool allows for the extraction and labelling of edits, even though it was not conceived of as targeting error analysis or edit type classification as such. Among the few tools developed specifically for error analysis, Popović (2011) mentions Hjerson, a tool to automatically detect and classify post-edits. Drawing on Vilar et al. (2006), the author develops a methodology to sort post-edits into

[1] https://github.com/felipealco/webnlg-pt/

five categories, namely: inflectional error; reordering error; missing word; extra word; and incorrect lexical choice (Popović, 2011; Popović & Ney, 2011). Popović et al. (2014) further expand error analysis by examining distinct types of post-editing operations and their association, in particular, with post-editing temporal and cognitive effort. On the one hand, temporal effort was measured as time spent on editing each sentence, while cognitive effort was measured on the basis of quality level scores assigned by a human annotator to machine translated sentences in terms of the amount of editing the annotator considered necessary to improve them. The same five sources of error used in Popović (2011) were explored, namely: word form, word order, missing words, extra words, and lexical choice. Her study found that word order and lexical choice were the edit types which implied the greatest cognitive effort, with the number of reordering edits being inversely proportional to the machine translation quality level. In other words, reordering edits were low for high quality translations and high for low quality translations. The authors also analysed reordering distance (number of word positions by which a particular word is shifted) and found that short range reordering prevails in high-quality translations.

Popović et al. (2014) argue that errors detected by the machine can be probed semi-automatically to identify them from a linguistic perspective and explore what prevents "almost acceptable translations" from becoming fully acceptable and thus taking machine translation to the next level. Sharing the authors' view, our study pursues automatic edit type annotation, not only as an end in itself, but also as a first step towards a more insightful shared task between machine and human analysis. Edits ascribed to a particular language problem, such as punctuation, word form, and word order at low ranks (morpheme, word and phrase) are comparatively easy to classify, whereas edits at higher ranks (clause and clause complex) are more difficult to account for, not to mention inter-sentential edits, where processes such as co-reference operate. Other edits are even more challenging in that they yield texts which compete with perfectly adequate and acceptable machine translations. Our analysis begins with automatically detecting edits and then probing into their nature.

We have focused on an English-Brazilian Portuguese corpus. To the best of our knowledge, only one other study has explored error analysis characterization in this language pair: Caseli and Inacio (2020) report on a comparative study of MT errors generated by a Neural Machine Translation (NMT) system and a Phrase-based Statistical Machine Translation (PBSMT) system. The two systems were used to translate into English a subset of the Portuguese texts in the FAPESP parallel corpus, a corpus of news retrieved from a popular science magazine published by a research foundation of the State of São Paulo, Brazil (Aziz et al. 2011), and a comparison was made between MT and human translated texts in the corpus. The authors found lexical choice to be the most frequent error in both MT systems when compared with human translations. In addition, they found the NMT system to outperform the PBSMT when dealing with syntax and word order errors, whereas the PBSMT system performed better on additions, omissions and non-translated word edits.

Caseli and Inacio (2020) focused solely on machine translation and did not deal with post-editing. Moreover, they used a human annotator to detect and classify errors in the machine output, which they analysed both quantitatively and qualitatively. In contrast, our study targets post-editing as well as machine translation output and relies on fully automatic edit type detection to report on the quantitative results we obtained.

Our approach is very similar to methods used in edit distance metrics. One such method is Levenshtein's edit distance (Levenshtein, 1966), which, unlike ours, is limited to addition, deletion, and lexical substitution operations. The TER (Snover et al. 2006, 2009) metric also computes edit operations to calculate a similarity score. Besides the operations found in the Levenshtein method, TER also includes word order, also known as word shifting. Our approach introduces two additional edit types – punctuation and morphological substitution – and while TER treats punctuations as words, we consider them to be an edit category in itself. It is worth pointing out that our approach successfully manages to extract word order edits, a particularly challenging operation as it demands considering the whole sentence, rather than short range alterations, without increasing computational cost in terms of time. Regarding substitution, we manage to distinguish morphological and lexical substitutions by comparing token lemmas instead of simply contrasting tokens.

3. Methodology

3. 1 WebNLG Corpus

We extracted our source texts from WebNLG (Colin et al. 2016; Gardent et al., 2017a), a corpus consisting of pairs of instances of Resource Description Framework (RDF) triples and their verbalization in English. This corpus was built for the Natural Language Processing task of data-to-text generation, i.e., automatically verbalising meaning representation. Human participants were recruited from a crowdsourcing platform (Gardent et al., 2017a) and were asked to provide coherent descriptions of the triples by means of clauses, making sure no two verbalisations were identical. The final corpus consists of paired instances between a meaning representation and its English verbalizations. One such pair is shown in Figure 1.

Subject	Property	Object
Blockbuster_(comicsCharacter)	alternativeName	Mark Desmond

⬇

Verbalization 1	"Mark Desmond" is also known as the comic character Blockbuster.
Verbalization 2	The comic character, Blockbuster, has the alternative name, Mark Desmond.
Verbalization 3	The comic character Blockbuster is also known as Mark Desmond.

Figure 1 - Semantic triples and verbalisations in the WebNLG corpus

Figure 1 shows an example of a WebNLG paired instance between a meaning representation and three verbalisations. In the representation, a property determines a relation between a subject and an object. The property "alternative name" establishes a relation between the subject "Blockbuster (comics Character)" and an object ,"Mark Desmond", another name given to the character

Blockbuster. Each verbalisation has a distinct wording, but all three share an analogous meaning.

The WebNLG corpus has a total of 42,901 English verbalizations (each being 95 characters long on average) covering topics such airports, artists, astronauts, athletes, buildings, celestial bodies, cities, comics characters, dishes, means of transportation, monuments, politicians, sports teams, universities and written works. Those verbalisations are freestanding texts which may include more than one sentence.

The fact that the WebNLG corpus is made up of meaning representations and their corresponding verbalisations allows us to investigate conceptual text complexity as estimated by the number of semantic relations that are implicated in a given text. In Costa, Ferreira, Pagano & Meira (in press), we explored the impact of the number of semantic relations in a text on post editing effort, with results pointing to increased post-editing effort as the number of relations increases.

3.2 Post-Editing Corpus

Our corpus was built by automatically translating the WebNLG's verbalisations from English to Portuguese, which is split into training, test and development sets. The test set contains 1,606 meaning representations and, taken together, these meaning representations are verbalised into 4,148 English texts with some having more than one verbalisation. We translated those texts into Portuguese using a generic neural machine translation system (DeepL).

After the automatic translation step, post-editing was carried out on a web interface designed for our study, publicly available in playground mode at http://dcc.ufmg.br/~felipealco/webnlg-pt. The interface displays the source text in English, a label for its domain category and the machine translation output. Participants were advised (1) to post-edit the texts to make them suitable for a Brazilian Portuguese speaking readership; (2) to translate proper nouns whenever an existing translation was available; and (3) to consult external sources such as online dictionaries and search engines whenever deemed necessary, on the condition that they do not pause the post-editing session while doing so.

Two post editing modes were available to participants: free mode, i.e., freely editing within a text box; and guided mode, i.e., selecting from a set of operations (insertion to the right, insertion to the left, delete, and update), which were defined based on neural programmer-interpreter approaches for Automatic Post-Editing (APE) (Vu and Haffari, 2018). The post-editing interface also allowed participants to skip a text whenever they felt unable to post-edit it. Our post editing tool measured the time spent on each verbalisation, pausing time, the operations and their order in guided

mode and a log of intermediate edits in free mode and skipped verbalizations. Figure 2 shows our interface in free-mode.

Figure 2- Web interface for post-editing in free-mode

Figure 3 shows our interface in guided mode.

Figure 3 - Web interface for post editing in guided mode

37 participants were recruited to post-edit the machine translated texts, all of them having Portuguese as their L1 and English as their L2. 33 of them reported upper-intermediate skills in English while 4 reported intermediate skills. No previous knowledge about the domain of the texts to be post-edited was required, nor was any previous training in translation assumed on their part. Each machine translated text was post-edited by two independent participants.

In order to evaluate the quality of the machine translated and human post-edited texts, a third participant was asked to assign a score to them on a scale (very poor-poor-average-good-very good) according to their judgment regarding three questions: (1) How analogous is the meaning in the target language (Brazilian Portuguese) compared to the source text (English)?; (2) How successful is the choice of words, grammar, and punctuation in the target language (Brazilian Portuguese)?; (3) How fluently does the text read in the target language (Brazilian Portuguese)?". Evaluators included both new participants recruited for this task, and some of the post-editors; steps were taken so that they could not evaluate texts they themselves had post-edited. We developed an interface for our evaluation system to display post-edited texts alongside the raw machine translated text. To avoid annotator bias, the texts were randomly sorted and presented as candidate texts with no labelling whatsoever as to their source (machine or human). Whenever two post-edited texts were identical to each other, or one or the two of them were identical to the machine translated text, only a single instance was displayed for the evaluator to rank.

To automatically analyse the post-edits, we developed an approach called "Editing Brackets", described in detail in Section 4. The patters produced by the editing brackets were then classified into six categories: word order, punctuation, addition, deletion, morphological inflection and lexical choice. Then the number of edit types was computed and pointwise mutual information (PMI) was used to find associations between edit types and quality scores of the texts. Subsequent to our automatic analysis, we checked the accuracy of our method to detect shortcomings in our classification.

In this article, we focus on editing operations, leaving data such as task execution time, pausing time, and number of skipped texts to be addressed in future work.

3.3 Final Corpus

Our corpus is made up of 4,148 source English verbalizations, 4,148 machine translations of those verbalizations and 8,296 post-edited texts in Brazilian Portuguese (two post-edited texts per MT output).

For each post-edited text, as well as for each machine translated text, we obtained a quality score. Figure 4 shows an entry in our database.

From top to bottom in Figure 4, an instance in our corpus is organised by semantic category ("food"); size, referring to the number of RDF triples used in the original WebNLG experiment for the English verbalizations; the actual triple, with indication of the tokens in the text that are to be verbalised as "subject", "property" and "object"; its verbalization in English; its machine translation output in Portuguese and its two post-edited texts in Portuguese. The machine translated text as well as its post-edited versions have a quality score attached to them, displayed on the left column.

category		
food		
size		
1		
triple set		
object	**property**	**subject**
Onion	hasVariant	Amatriciana_sauce
English verbalization		
Onion is one variation of ingredients in Amatriciana sauce.		
machine translation output		
quality score	**text**	
medium	A cebola é uma variação de ingredientes no molho Amatriciana.	
post-editing		
quality score	**text**	
very good	A cebola é uma alternativa de ingrediente para o molho Amatriciana.	
medium	A cebola é uma variação de ingredientes no molho amatriciana.	

Figure 4 - Sample entry in post-editing corpus

For an overall assessment of quality scores contrasting machine output quality and post-edited quality, score percentages obtained in our experiment are displayed in Figure 5.

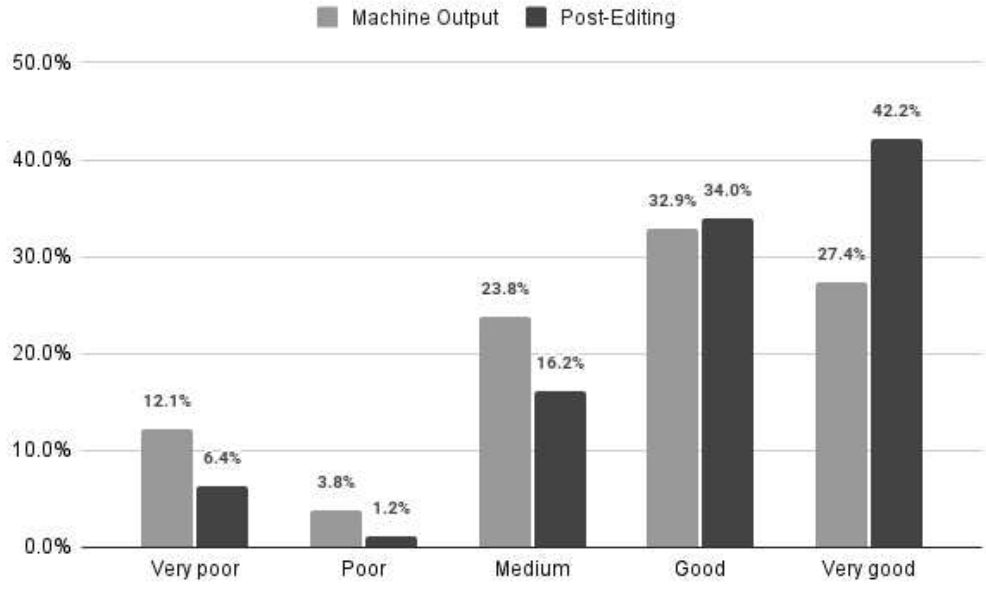

Figure 5 - Percentage of quality scores assigned to machine translation output and post-edited texts

Figure 5 shows that the quality of the machine output was consistently improved by post-editing for all quality ranks, evidenced by the decrease in very poor, poor and

medium scores and the increase in good and very good scores. Worth noticing is an almost 50% decrease in very poor scores, a 32% decrease in medium scores and a 54% increase in very good scores.

The fact that after post-editing almost 8% of the texts could not attain a score above poor and little over 16% of the texts only attained a medium score, could be attributed to the lack of familiarity of post-editors with the domains of the texts, some of which involved highly technical terms as is the case, for instance, of means of transportation (Figure 9) and celestial bodies (Figure 11). Nonetheless, all post-edits were included in our analysis on the premise that they may contain useful edits for machine learning.

4.Analysis

We have named our method to automatically identify edit types "Editing Brackets''. This method highlights the differences between two texts by enclosing the differential part inside brackets, where a vertical bar or pipe symbol marks off the original token to the left side and the edited one to the right side. If there is an addition to the original text, the left side of the bar is empty. Similarly, if there is deletion, the right side of the vertical bar is empty. If both sides of the bar in the bracket are filled, the token on the left has been replaced by the token on the right. An example is shown in Figure 6.

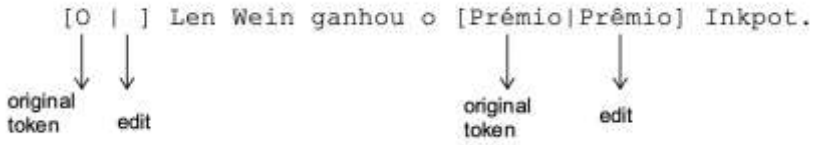

Figure 6 - Example of Editing Brackets method visualization

In the example in Figure 6, the machine output “O Len Wein ganhou o Prémio Inkpot” ("The Len Wein won the Inkpot Award") was post-edited as “Len Wein ganhou o Prêmio Inkpot” ("Len Wein won the Inkpot Award"). When this sentence is given as input to our method, two Editing Brackets are produced. The first one, [O|], indicates that the definite article “O” was in the machine translated text, but it was removed in the post-edited text. The second editing bracket, [Prémio|Prêmio], indicates a replacement operation where the word "Prémio" was replaced by "Prêmio".

We compute Editing Brackets in our corpus by using a modified version of the Longest Common Substring (LCS) algorithm, which involves comparing two strings and finding the longest sequence of items that are common to both (Gusfield 1997). We have adapted it to the token level instead of using it on a substring level, as its name suggests. With the LCS algorithm, we find the longest overlap between the tokens in the machine translated and the post-edited texts. If there is no intersection, this means all tokens in the texts are different and we place them inside the Editing Brackets. This amounts to saying that the texts are completely different from one another. If there is an intersection, we keep the overlapping tokens and recursively call the LCS algorithm for the remaining

text, starting from the extremities. Figure 7 shows an example of this process for the text displayed in Figure 6.

step 1	LCS("O Len Wein ganhou o Prémio Inkpot.", "Len Wein ganhou o Prêmio Inkpot.")
step 2	LCS("O", "") + "Len Wein ganhou o" + LCS("Prémio Inkpot.", "Prêmio Inkpot.")
step 3	"[O|] Len Wein ganhou o" + LCS("Prémio", "Prêmio") + "Inkpot."
step 4	"[O|] Len Wein ganhou o [Prémio|Prêmio] Inkpot."

Figure 7 - Steps in the Editing Brackets process

In step 1, the LCS algorithm found "Len Wein ganhou o" to be the longest overlap between the two texts. Thus, two other LCS algorithms were called to this text, one for each extremity. In step 2, no intersection was found for "O" and "", yielding "[O|]". Meanwhile at the other extremity, there was an intersection between "Prémio Inkpot." and "Prêmio Inkpot." This required a new LCS algorithm call. In step 3, the LCS algorithm made its last call as there was no further intersection between the remaining strings. The final result is presented in step 4.

4.1 Automatic Edit Detection

Besides providing a simplified visualization of the differential part of two texts, Editing Brackets can be computed to automatically detect edit types made to the original text. Patterns in the content of Editing Brackets allow us to group occurrences into edit types. For example, if the right side of the brackets is empty, as in [O|] in Figure 6 above, we assume there has been a deleting operation. However, a close analysis is needed to verify if there has in fact been a deletion. If, however, there is a change in the position of a token in a text, we will find two Editing Brackets: one for its deletion in its original position and another one for its addition in its new position in the text.

In the following subsections, we will explain the classes into which we have categorized edits and how we detected them using a filter for each edit type. In order to obtain a better performance of our lemmatiser, we first converted our texts to lowercase and proceeded to separate contractions in Portuguese, such as contractions of prepositions and articles. For instance, the contraction of preposition and female definite article "da" was separated into "de" ("of") and "a" (the").

The first step consists in computing Editing Brackets. After that, the text goes through a sequence of filters where each filter is responsible for detecting its corresponding edit type, assigning it to the Editing Bracket and removing it from the text. If there are further Editing Brackets in the text, the next filter is applied. All edit types in a sentence are considered to have been detected when there are no Editing Brackets left.

4.1.1 Filter 1: Word order

For the simplest case of word order edit, two Editing Brackets are yielded: one for the tokens removed and another one for their placement in the new position. Thus, the word or sequence of words that were placed in a new position in the text appear

twice in the text: once on the left side of a pair of Editing Brackets and another one on the right side of another pair. Figure 8 shows an example of this type of edit, which we call word order edit.

[a nacionalidade de|] karl kesel [é|tem nacionalidade] americana.

Figure 8 - Example of word order edit with two Editing Brackets

In the example in Figure 8, "nacionalidade" ("nationality") was moved from subject position at the beginning of the sentence to predicate position. Further changes within the brackets are a side effect of the word order edit and therefore counted as a single word order edit.

More complex word order adjustments may produce more than two Editing Brackets. An example of this is shown in Figure 9, where two text segments are combined into a new sentence added at the end of the text.

[a distância entre eixos de|] o abarth 1000 gt coupe [é de 2160 milímetros e|] é um coupé de duas portas. [|a distância entre os eixos de esse carro é de 2.160 milímetros.]

Figure 9 - Example of word order edit with three Editing Brackets

For Filter 1 in our method, we ignore cases where brackets contain only punctuation marks, as these are included in the next filter, dedicated to punctuation.

4.1.2 Filter 2: Punctuation

In addition to Popović's classes (2011), our study considered punctuation edit types. These covered any punctuation mark that is added, removed, moved, or replaced by a different one. Punctuation is a relevant issue in text production, as it is implicated in meaning production. We consider punctuation edits all Editing Brackets that contain punctuation marks solely, regardless of the operation carried out, that is, if there was an addition, a deletion, or a replacement implicated. For example, Figure 10 shows a sentence with one editing bracket containing a comma addition.

a área total de albany, oregon [|,] é de 45,97 km2.

Figure 10 Example of punctuation edit

4.1.3 Filter 3: Addition

The addition type is detected by checking if the left side in the bracket is empty, which means the text on the right was added without any further replacement of tokens. One such case is illustrated in Figure 11.

[|o corpo celeste (] 15788 [|) 1993 sb] foi descoberto por iwan p williams em 1993 [sb|] . seu período orbital é 7729430000, [periapsia|apside] de 3997100000000, e sua [|data de] época é 6 de março de 2006.

Figure 11 Example of addition edit

The example in Figure 11 shows two additions: "o corpo celeste" ("the celestial body") and "data de" ("date of"). Word order edits such as "sb" are not counted in this filter as they were computed in Filter 1. Word edits, such as "periapsia" by "apside", will be counted in Filter 5, explained in 4.1.5.

4.1.4 Filter 4: Deletion

Similar to addition, a deletion edit is also detected from an empty side in Editing Brackets. When a segment of text is removed from the original text, the right side in the brackets is empty, indicating that what was on the left side has been removed. An instance of deletion is shown in Figure 12.

josef klaus sucedeu [a|] alfons gorbach.

Figure 12 Example of word deletion edit

Figure 12 shows the deletion of preposition "a" in Portuguese, indicated by an empty space to the right of the vertical bar.

4.1.5 Filter 5: Word edit

Once all Editing Brackets with an empty side have been removed in the two previous filters (addition and deletion), the remaining Editing Brackets are the ones having tokens on both sides of the brackets. These are “word” edits. This edit type was inspected for patterns matching two subtypes by considering the lemma of the tokens inside the brackets. If the tokens on the right side of the vertical bar were replaced by tokens tagged with the same lemma, we counted that edit as a morphological inflection edit type; if the tokens were tagged with a different lemma, we counted them as a lexical choice edit type.

For this filter we used the lemmatiser module from SpaCy, a leading open-source software library for advanced natural language processing. This lemmatiser was trained on the Constraint Grammar converted version of the Brazilian Portuguese corpus Bosque, which is available in SpaCy in three sizes. For this study, we have used the largest one available.

The example in Figure 13 shows an instance of a morphological inflection edit type.

15788 1993 sb foi [descoberta|descoberto] pelo observatório roque de los muchachos.

Figure 13 Example of morphological inflection edit.

In Figure 13, the adjectival female gender ending "a" in "descoberta" ("discovered") was replaced by the male gender ending "o" in "descoberto", to match the male gender in Portuguese for the noun which hypernames "15788 1993 sb", a transneptunian object or dwarf planet.

An instance of lexical choice edit is shown in Figure 14.

chuck fletcher é o [gerente|diretor] geral do minnesota wild.

Figure 14 Example of lexical choice edit.

In Figure 14, "gerente" ("manager") was replaced by "diretor" ("director"), each word having a different lemma.

4.2 Sample filtering cycle

At each filter run, Editing Brackets are inspected for patterns matching a particular edit type. Figure 15 shows an example of a post-edited text with multiple edit types and their status at each filter run. The pattern matching each edit type is highlighted in bold font. Once a pattern is found, this is filtered out at the following run, as is highlighted in Figure 15 with a strikethrough line.

The example in Figure 15 shows one word order edit type at filter 1.

Filter 1 Word order	**[|o ex-aluno alfred moore scales de a]** universidade de a carolina de o norte em chapel hill **[aluna alfred moore escalas alfred|]** foi um governador democrata de a carolina de o norte. ele nasceu em reidsville, carolina de o norte e [|foi] sucedido por james [,|] w. reid.
Filter 2 Punctuation	~~[|o ex-aluno alfred moore scales de a]~~ universidade de a carolina de o norte em chapel hill ~~[aluna alfred moore escalas alfred|]~~ foi um governador democrata de a carolina de o norte. ele nasceu em reidsville, carolina de o norte e [|foi] sucedido por james **[,|]** w. reid.
Filter 3 Addition	~~[|o ex-aluno alfred moore scales de a]~~ universidade de a carolina de o norte em chapel hill ~~[aluna alfred moore escalas alfred|]~~ foi um governador democrata de a carolina de o norte. ele nasceu em reidsville, carolina de o norte e **[|foi]** sucedido por james ~~[,|]~~ w. reid.
Filter 4 Deletion	~~[|o ex-aluno alfred moore scales de a]~~ universidade de a carolina de o norte em chapel hill ~~[aluna alfred moore escalas alfred|]~~ foi um governador democrata de a carolina de o norte. ele nasceu em reidsville, carolina de o norte e ~~[|foi]~~ sucedido por james ~~[,|]~~ w. reid.
Filter 5 Word Edit	~~[|o ex-aluno alfred moore scales de a]~~ universidade de a carolina de o norte em chapel hill ~~[aluna alfred moore escalas alfred|]~~ foi um governador democrata de a carolina de o norte. ele nasceu em reidsville, carolina de o norte e ~~[|foi]~~ sucedido por james ~~[,|]~~ w. reid.

Figure 15 - Example of inspection of editing brackets on the filtering cycle

For a pattern to qualify as a word order edit type, a minimum of a single token needs to be shared by two Editing Brackets and there has to be a shift in position in the text. In our example, the first two editing brackets share the tokens alfred and moore. As we have mentioned before, we consider the other unique tokens in those brackets side effects of word ordering. Therefore, they are assigned to the word order category

and removed from upcoming filters. At filter 2, an edit type matches a punctuation edit type: a comma was removed by the post-editor. At filter 3, an edit type is detected matching addition: the verb "foi" ("was") was added to the text. At filters 4 and 5, no Editing Brackets are left to be inspected and the filtering cycle is thus concluded.

5. Results

Out of the 8,296 post-edited texts, 5,176 texts were extracted as having been actually modified by post-editing. Altogether, those texts include 11,219 Editing Brackets. Each pair of brackets and tokens within them counts as one Editing Bracket. Table 1 shows how the Editing Brackets are distributed among our categories.

Edit type	Editing Brackets	Proportion
Word Order	1630	14.53%
Punctuation	1417	12.63%
Addition	1528	13.62%
Deletion	998	8.90%
Word Edit		
Morphological inflection	4740	42.25%
Lexical choice	906	8.08%
Total of Editing Brackets	11219	100%

Table 1 Proportion of edit types assigned to the Editing Brackets

As can be seen in Table 1, in terms of number of Editing Brackets, word edit was the category with the highest number of occurrences, accounting for over 50% of all edit types. When considering its subtypes, the majority of edits implicated morphological inflection. Word order was the second highest edit type after morphological inflection, closely followed by addition and then punctuation. Having less than 10% of occurrences, lexical choice and deletion were respectively the two least frequent edit types in our corpus.

Word order edits had a relatively high number of occurrences, but it should be borne in mind that there is not necessarily a one-to-one correspondence between Editing Brackets and edit operations as we have stated in Filter 1 (4.1.1). The example in Figure 9 shows two segments that were moved to the end of the sentence. Three Editing Brackets are counted: one for each segment removed from its original position plus one for the addition of both at the end. A precise number of instances of word order edits requires manual inspection, a shortcoming of our method we will discuss below.

5.1 Edit Type and Quality Score Association

In order to explore edit type and quality score, we opted for a technique used to compute the relatedness between two variables, namely, Pointwise Mutual Information

(PMI) (Turney, 2001). PMI computes association between variables indicating to what extent two variables are more likely to occur together than separately. The higher the value of this measure, the higher the chances of co-occurrence. In our case, we applied this technique to our data to find the extent to which an edit type was associated with particular quality scores, both of machine translation output and post-edited texts. Again, the higher the value of this measure, the higher the chances of an edit type to co-occur with a quality score. To remove data sparsity in the analysis, our original five quality labels were subsumed under three main ones – poor (encompassing very poor and poor), medium and good (encompassing good and very good).

5.1.1 Machine Output

Our results of PMI for the association between edit type and original machine translation output quality scores are shown in Table 2. In this case, we used the quality score assigned to the machine translation output in order to check which edit types were most associated with our three scores.

Machine Output					
poor		**medium**		**good**	
deletion	-0.042	morphological	0.130	word order	0.389
lexical	-0.043	punctuation	0.125	morphological	0.109
addition	-0.053	word order	0.095	deletion	0.067
punctuation	-0.112	addition	0.063	punctuation	0.061
morphological	-0.312	lexical	0.025	lexical	0.061
word order	-0.312	deletion	0.019	addition	0.029

Table 2. PMI between edit types and machine output quality scores

Regarding machine translation quality scores, Table 2 shows that machine translated texts ranked as poor do not tend to co-occur with any particular edit type. Unlike them, machine translated texts ranked as medium tend to co-occur with morphological inflection, punctuation, and word order edit types and, to a very little extent, with addition, lexical choice, and deletion. Figure 16 shows an example of an edit type classed as morphological inflection in a medium-ranked machine translated text.

[o autor|a autora] de a wizard of mars é diane duane e o formato impresso tem um número oclc de 318875313 e um número isbn de 978-0-15-204770-2.

Figure 16 - Example of MT output rated as medium with a morphological inflection edit

In Figure 16, the machine output shows the noun phrase "o autor" ("the author") used as a generic reference to a female proper name ("Diane"). The generic reference in Portuguese "o autor" is made up by an article and a noun, both having male grammatical gender. The morphology of both the male article "o" and the male noun "autor" were

edited so that the reference expression was grammatically female ("a autora") and matched "Diane Duane", assumed to be a female writer.

Figure 17 shows an example of medium-rated machine output with a punctuation edit.

o código de área para austin, texas [|,] é 512.

Figure 17 - Example of MT output rated as medium with a punctuation edit

In Figure 17, a comma was inserted to separate the name of a location ("Texas") from the rest of the sentence, following punctuation conventions in lists of geographical entities.

Machine translated texts ranked as good tend to co-occur with all edit types, word order being the most frequent. The second edit type is morphological inflection followed by, though to a very little extent, all other edit types. An example of a word order type of edit is shown in Figure 18.

[|o iraque é] a terra natal de ahmad kadhim assad [é o iraque|] .

Figure 18 - Example of MT output rated as good with a word order edit

In the sentence in Figure 18, there is a shift whereby subject and complement are reversed in order so that "o Iraque" ("Irak") becomes the subject and "a terra natal de ahmad kadhim assad" ("the birthplace of ahmad kadhim assad") becomes complement. As already noted for our example in Figure 8 and in the literature (Popović et al, 2014), some word order edits, as is the case of the one in Figure 18, are harder to account for as they do not implicate improvement of inadequate or unacceptable machine output. Rather, they point to acceptable translations, which nonetheless are considered worthy of further improvement. In this particular case, edits may be explained on the basis of preferred word order patterns in a language.

5.1.2 Post-edited Text

Our results for PMI exploring the association between edit type and post-edited text quality are displayed in Table 3.

Regarding post-edited text quality scores, data in Table 3 show that post-edited texts ranked as poor tend to co-occur with punctuation and, to a very little extent, with morphological inflection edit types. Figure 19 shows an example of a post-edited text ranked as poor in co-occurrence with a punctuation edit. In this case, the post-editor removed a comma that was between the subject “um feiticeiro de marte” (“A Wizard of Mars”) and the verb “foi” (“was”).

Post-Editing					
poor		**medium**		**good**	
punctuation	0.035	word order	0.165	morphological	0.093
morphological	0.002	deletion	0.067	deletion	0.046
addition	-0.003	addition	0.038	lexical	0.018
lexical	-0.009	lexical	0.031	addition	-0.055
deletion	-0.021	morphological	-0.057	word order	-0.056
word order	-0.034	punctuation	-0.103	punctuation	-0.139

Table 3. PMI between edit types and the post-edited quality scores

However, he/she failed to remove an extra dot after the word “dura”, which most probably prevented this text from obtaining a score above poor. Edits deemed essential but which fail to be implemented by post-editors have been pointed out and in the literature on post-editing (cf., for instance, De Almeida, 2013).

o livro de capa dura. um feiticeiro de marte [,|] foi escrito por diane duane e tem o número isbn 978-0-15-204770-2

Figure 19 - Example of post edited text rated as poor with punctuation edit

Post-edited texts assessed as medium tend to co-occur with word order edit types. To a lesser extent, they co-occur with deletion, addition, and lexical choice edit types. Figure 20 shows an example in which the post-editor performed a word order edit to produce an adequate noun phrase in terms of structure in Portuguese: "o nome completo do personagem cômico" ("the full name of the humorous character"). Yet, the lexical choice "personagem cômico" as a translation for "comic book character" is not accurate in Brazilian Portuguese, which most likely prevented the post-edited text from attaining a score above medium.

[o personagem cómico,|] o nome completo de [|o personagem cômico] auron é lambien.

Figure 20 - Example of post edited text rated as medium with word order edit

Post-edited texts assessed as good tend to basically co-occur with morphological inflection, deletion and lexical choice edit types. Figure 21 shows an example in which the post-editor replaced a plural number noun phrase “os americanos” (“the Americans”) by a singular number noun phrase “o americano” (“the American”) to refer to a single person (Paris Cullins).

ernie colón e [os americanos|o americano] paris cullins estavam entre os criadores de o personagem de quadrinhos bolt, também conhecido como larry bolatinsky.

Figure 21 - Example of post edited text rated as good with morphological inflection edit

6. Discussion

Our study found a sharp decrease of low quality scores when machine translated texts rated as "very poor" and "medium" were post-edited by humans (almost 50% and 32% decrease respectively). Moreover, it showed an increase of 54% in "very good" quality scores after human post-editing. Such results clearly confirm human post-editing of machine translation as a necessary step in the workflow for obtaining texts of higher quality as previously reported by Toral (2018) and Laubli (2018). Additionally, they support the view that post-editing is a very productive task to tap for insights into which recurrent edits performed by humans can feed a translation model so as to obtain a better machine translation output.

Regarding most frequent edit types, morphological inflection was the most frequent one. This does not confirm the findings reported in Caseli and Inácio (2020), who pointed out lexical choice as the most frequent error annotated when comparing machine translated and human translated texts. The fact that they did not target post-editing in particular and that directionality in their work was from Brazilian Portuguese into English might be an explanation for this result.

The chances of co-occurrence between an edit type and a quality score was measured using a popular statistical metric in Natural Language Processing known as Pointwise Mutual Information (PMI) (cf. Turney, 2001). This metric is often used to detect words that occur together rather than separately in a given corpus. Thus, we sought to investigate whether or not an edit type is associated with a quality mark. Our results showed higher co-occurrence between word order edits and machine translation output ranked as good and post-edited texts ranked as medium. Moreover, word order edits showed no tendency to occur either with poor quality scores of machine translated output or post-edited texts.

It is worth noting that machine translations ranked with good scores were very likely to be post-edited because of word order, while poor post-edited translations were very unlikely to have a word order edit. This means that when the machine output is poor and the text gets a poor score after being post-edited, word order was not an edit type impacting the quality of the post edited translation. Yet, a manual analysis of some of our examples for word order edits, like those in Figures 8 and 18, point to word ordering as a type of post-editing with leveraging potential to gather insights as to what motivates these operations carried out by human post-editors even when the quality of the machine translation output is good and very good. A manual analysis of instances of word order edits in cases where these are not deemed necessary can point to patterns of preferences in word order that may be ascribed to the grammatical systems of information and theme-rheme from a systemic-functional perspective (cf. Halliday; Matthiessen, 2014). This type of analysis is currently underway and results will be reported in follow-up papers.

7. Conclusion

In order to automatically detect edit types with a view to exploring their impact on the quality of the post-edited texts, we introduced a method we called "Editing Brackets" to

contrast each automatic translation and its corresponding post-edited text. Based on the results and drawing partially on Popović (2011), we automatically classified each difference between both texts into five edit types: word order, punctuation, addition, deletion, and word edits (morphological inflection and lexical choice). The Editing Brackets method proved efficient to screen edit types automatically and allow for an overall classification. Our method was not envisioned to compete with existing forms of approaching post-editing, which largely draw on editing distance measures; rather, it is a method to detect edits and class them into types to be readily analysed as edit categories. Detecting and sorting out word edits was successfully carried out thanks to the good performance of the lemmatisation model used, which allowed for separating morphological inflection from lexical choice edit types. Quite expectedly, punctuation edits were easy to detect automatically, as they comprise a closed set of tokens (punctuation marks) to be queried inside the brackets. Additions and removals were also readily detected after the word order filter had been run.

Our study set out to explore a method to automatically detect and classify human post-edits by automatically pre-assigning edit types to Editing Brackets, leaving for the human analyst the task of pursuing a more fine-grained analysis in studies aimed at more insightful assessment of machine translation and human post-editing. Our method proved useful in sorting edit types and finding out which types contribute the most to different quality scores.

Edit types most easily detected, such as punctuation, can be readily accounted for and implemented in a post-editing API. Other edit types, such as word order, are natural candidates for a more fine-grained analysis, both in terms of computing the number of their occurrences as well as accounting for their occurrence, particularly when the produced texts are as adequate and acceptable as the machine translated sources themselves.

Considering that word order editing requires different ranges of editing operations in connection to ranks other than, and higher than, the word and that this edit operation is associated in the literature with cognitive post-editing effort (cf. Popović et al., 2014), their examination is worth pursuing, not only to improve their automatic detection, but also to account for their occurrence, especially when the machine translation output has reached very good scores.

Acknowledgments

This research was partially funded by the agencies CNPq, CAPES, and FAPEMIG. In particular, the researchers were supported by CNPQ grant No. 310630/2017-7, CAPES Post doctoral grant No. 88887.508597/2020-00, and FAPEMIG grant APQ-01.461-14. This work was also supported by projects MASWeb, EUBra-BIGSEA, INCT-CYBER, and ATMOSPHERE. The authors also wish to express their gratitude to Deepl for kindly granting a license to translate our corpus, and to the students at UFMG who took part in the post-editing experiment. We would like to thank the anonymous reviewers who carefully read our manuscript and provide many insightful comments and suggestions.

References

Aziz, Wilker; Lucia Specia (2011). Fully automatic compilation of Portuguese-English and Portuguese-Spanish parallel corpora. In: *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology: Cuiabá, MT, Brazil, October 24-26,* pp. 234-238. <https://aclanthology.org/W11-4533.pdf>. [Accessed: 20211207].

Aziz, Wilker; Castilho, Sheila; Specia, Lucia. (2012). PET: a Tool for Post-editing and Assessing Machine Translation. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC12).* European Language Resources Association (ELRA), pp. 3982-3987. <http://www.lrec-conf.org/proceedings/lrec2012/pdf/985_Paper.pdf>. [Accessed: 20211207].

Caseli, Helena; Marcio, Inácio (2020). NMT and PBSMT Error Analyses in English to Brazilian Portuguese Automatic Translations. In: *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020): Marseille, 11-16 May*. European Language Resources Association (ELRA), pp. 3623-3629. <https://aclanthology.org/2020.lrec-1.446.pdf>. [Accessed: 20211207].

Chatterjee, Rajen; Federmann, Christian; Negri, Matteo; Turchi, Marco (2019). Findings of the WMT 2019 Shared Task on Automatic Post-Editing. In: *Proceedings of the Fourth Conference on Machine Translation (Volume 3: Shared Task Papers, Day 2): Florence, Italy, August.* Association for Computational Linguistics, pp. 11–28. <https://dx.doi.org/10.18653/v1/W19-5402>. <https://aclanthology.org/W19-5402.pdf>. [Accessed: 20211207].

Costa, Felipe; Ferreira, Thiago; Pagano, Adriana; Meira, Wagner (2020). Building The First English-Brazilian Portuguese Corpus for Automatic Post-Editing. In: *Proceedings of the 28th International Conference on Computational Linguistics: Barcelona, Spain (Online), December 8-13.* International Committee on Computational Linguistics, pp. 6063–6069. <https://dx.doi.org/10.18653/v1/2020.coling-main.533>, <https://aclanthology.org/2020.coling-main.533.pdf>. [Accessed: 20211207].

Costa, Felipe; Ferreira, Thiago; Pagano, Adriana; Meira, Wagner. (2022, in press). Exploring Semantic Annotations to Measure Post-Editing Quality. In: Ji, Meng; Oakes, Michael P. (ed.). *Corpus Exploration of Lexis and Discourse in Translation*. London: Routledge.

De Almeida, Giselle. (2013). T*ranslating the post-editor: an investigation of post-editing changes and correlations with professional experience across two Romance languages* [PhD thesis]. School of Applied Language and Intercultural Studies, Dublin City University. <https://doras.dcu.ie/17732/>. [Accessed: 20211207].

Gardent, Claire; Shimorina, Anastasia; Narayan, Shashi; Perez-Beltrachini, Laura (2017). Creating Training Corpora for NLG Micro-Planning. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: Vancouver, Canada, July 30-August 4 (Volume 1: Longs Papers).* Association for Computational Linguistics, pp. 179-188. <https://dx.doi.org/10.18653/v1/P17-1017>, <https://aclanthology.org/P17-1017.pdf>. [Accessed: 20211207].

Gardent, Claire; Shimorina, Anastasia; Narayan, Shashi; Perez-Beltrachini, Laura (2017). The WebNLG Challenge: Generating Text from RDF Data. In: *Proceedings of the 10th International Conference on Natural Language Generation: Santiago de Compostela, Spain, September 4-7.* Association for Computational Linguistics, pp. 124-133. <https://dx.doi.org/10.18653/v1/W17-3518>, <https://aclanthology.org/W17-3518.pdf>. [Accessed: 20211207].

Gusfield, Dan (1997). Preface (Abridged) of Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology. *Acm Sigact News*, v. 28, n. 4, pp. 41-60.

Halliday, M.A.K. (aut.); Matthiessen, Christian M.I.M. (revised) (2014). *Halliday's Introduction to Functional Grammar*. 4th ed. Milton Park [etc.]: Routledge.

Läubli, Samuel; Sennrich, Rico; Volk, Martin (2018). *Has Machine Translation Achieved Human Parity? A case for Document-level Evaluation* [Preprint]. <https://arxiv.org/abs/1808.07048v1>. [Accessed: 20211207].

Levenshtein, Vladimir I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady,* v. 10, n. 8 (February), pp. 707-710. <https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf>. [Accessed: 20211207].

Popović, Maja; Ney, Hermann (2011). Towards Automatic Error Analysis of Machine Translation Output. *Computational Linguistics*, v. 37, n. 4 (December), pp. 657–688. <https://dx.doi.org/10.1162/COLI_a_00072>. [Accessed: 20211207].

Popović, Maja; Lommel, Arle; Burchardt, Aljoscha; Avramidis, Eleftherios; Uszkoreit, Hans. (2014). Relations between different types of post-editing operations, cognitive effort and temporal effort. In: P*roceedings of the 17th Annual conference of the European Association for Machine Translation: Dubrovnik,Croatia, June 16-18.* European Association for Machine Translation, pp. 191-198. <https://aclanthology.org/2014.eamt-1.41>, <https://aclanthology.org/2014.eamt-1.41.pdf>. [Accessed: 20211207].

Popović, Maja. (2018). Error Classification and Analysis for Machine Translation Quality Assessment. In: Moorkens, J.; *et al.* (eds.). *Translation Quality Assessment.* Cham: Springer International. (Machine Translation: Technologies and Applications; 1), pp. 129-158. <https://doi.org/10.1007/978-3-319-91241-7_7>. [Accessed: 20211207].

Popović, Maja. (2011). Hjerson: An Open Source Tool for Automatic Error Classification of Machine Translation Output. *The Prague Bulletin of Mathematical Linguistics,* n. 96 (October), pp. 59–67. <https://doi.org/10.2478/v10108-011-0011-4>, <https://www.readcube.com/articles/10.2478%2Fv10108-011-0011-4>. [Accessed: 20211207].

Snover, Matthew; Dorr, Bonnie; Schwartz, Richard; Micciulla, Linnea; Makhoul, John (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers: Cambridge, August 8-12.* The Association for Machine

Translation in the Americas, pp. 223-231. <https://aclanthology.org/2006.amta-papers.25.pdf>. [Accessed: 20211207].

Snover, Matthew; Madnani, Nitin; Dorr, Bonnie J.; Schwartz, Richard (2009). Fluency, Adequacy or HTER? Exploring Different Human Judgments with a Tunable MT Metric. In: *Proceedings of the Fourth Workshop on Statistical Machine Translation: Athens, Greece, 30-31 March*. Association for Computational Linguistics, pp. 259-268. <https://aclanthology.org/W09-0441.pdf>. [Accessed: 20211207].

Toral, Antonio; Castilho, Sheila; Hu, Ke; Way, Andy (2018). *Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation* [Preprint]. <arXiv:1808.10432v1>. [Accessed: 20211207].

Turney, Peter D. (2001) Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. In: De Raedt, L.; Flach, P. (eds.). *Machine Learning: ECML 2001*. Berlin [etc.]: Springer. (Lecture Notes in Computer Science; 2167), pp. 491-502. <https://doi.org/10.1007/3-540-44795-4_42>. [Accessed: 20211207].