ACADEMICS LOST IN MACHINE TRANSLATION: ON POLITICAL SCIENCE TEXTS

Alina RĂDOI
West University of Timișoara, Romania

Abstract: Being under the constant threat of "publish (in English!) or perish", academics use public machine translation systems in order to translate articles intended for publication in international journals. Though free and accessible, MT systems do not always produce high quality output. Target texts can contain errors ranging from minor to very severe. This interdisciplinary paper focuses on the results of a small-scale study. A political science academic text was translated from Romanian into English using three neural MT systems: Bing Microsoft Translator, DeepL and Google Translate. Errors were manually annotated in CATMA, a digital tool, taking into account the fluency, accuracy and fitness for purpose of the target texts. Translators, MT systems developers and academics alike can profit from these findings by understanding what kind of errors to expect – and eventually –, correct and eliminate when automatically translating academic texts from Romanian into English.

Keywords: academic texts, error annotation, error classification, error typology, neural machine translation, translation assessment, translation quality assessment, Romanian

1. Introduction

Since they became publicly available in 2016, neural machine translation systems have been used around the world by people from very different backgrounds to help them understand messages, be they mundane or professional.

Academics are certainly among the 1 billion users reported by Google Translate in 2021. This is especially because they are under the threat of "publish or perish" and due to the fact that English is now the dominant language in science. Non-native speakers are at a disadvantage when it comes to publishing their articles in high-visibility journals. If they do not have the necessary foreign language skills needed for academic writing or the resources to hire a professional translator, they might employ public machine translation engines to translate their own works into English.

In this paper I started out from this real-life scenario and conducted a small-scale study in the following manner: (i) excerpts from a political science academic text were translated from Romanian into English using three public NMT systems; (ii) the errors were annotated and analysed with the purpose of delineating which engine is best and least suited for this translation task.

Section 2 of this study deals with literature review and the background conditions, while section 3 takes a closer look at the methodology. Section 4 is dedicated to an inspection of the results – both at the general and individual level for each particular system – and section 5 deals with the main conclusions and future directions of research.

2. Previous research

Romanian has been a candidate target language in studies focused on the development of NMT systems (Lakew, et al., 2017; Koehn & Knowles, 2017) and was included in the

2016 edition of WMT (Bojar, et al., 2016; Peter, et al., 2016). It was also on the list of four languages in the EU-funded Health in My Language project, which aimed to build MT systems for translating public health information documents from English (https://www.himl.eu/).

More recently, Romanian was used in studies focused on machine translating user-generated content (Saadany, et al., 2021) and was included in ChatGPT multilingual translation research (Jiao, et al., 2023). Romanian speakers have additionally taken part in eye-tracking studies on the absence of morphological transfer in second-language acquisition (Sagarra, 2014).

Researchers have also focused their efforts on including Romanian in translation quality assessment studies, such as those carried out by the HAITrans Research Group at the University of Vienna. As a result, Rios et al. (2022) concentrated on a corpus of machine-translated medical abstracts from English into Romanian, with a focus on terminology errors. The same authors then investigated the quality of machine-translated medical texts from English into Romanian, this time annotating errors using the Multidimensional Quality Metrics (MQM) (Rios Gaona, et al., 2023).

Google Translate seems to be a favourite among Romanian language researchers especially when it comes to medical texts translated from English into Romanian (Dumitran, 2021; Delfani, et al., 2024). This system was also looked into by Pungă et al. (2023), but in relation to everyday and news/ news releases language.

The topic of translation quality assessment of academic texts translated using machine translation has been approached by Scansani (2020), who zoomed in on course catalogues, while Delorme Benites and Benites (2021) identified errors in PhD abstracts from German into English.

On top of this, researchers have tackled machine translation to emphasize the importance of machine translation literacy in academia (Bowker & Ciro Buitrago, 2019; Steigerwald, et al., 2022) or to underscore its significance as a tool for academic writing support (O'Brien, et al., 2018).

3. Research questions

As mentioned earlier, the underlying reasons for carrying out this research are (i) the prevalence of English in academic circles and (ii) the fact that some researchers have limitations in producing texts directly in this language. However, the cornerstone of this study is the – to the best of my knowledge – very finite number of research pieces concerned with the topic at hand and with this language combination and direction.

This research expands on the array of studies on neural machine translation where Romanian is involved, and academic texts are analysed. It reports on the errors that appear when translating a sample of a political science academic text from Romanian into English using three public machine translation systems: Bing Microsoft Translator, Google Translate and DeepL.

The three main research questions are:

1. What kind of errors appear?
2. How frequent are these errors?
3. Out of the three, which public neural machine translation system is best suited for this translation task?

It is important to mention the fact that not only academics can profit from knowing the answer to these questions, but also MT developers and translators alike. Developers may wish to understand what problems arise and how to solve them, while translators might benefit from being made aware of what to expect from public NMT engines when translating academic texts.

The research design and subsequent results are described below.

4. Research design

4.1. Source text

The source text was extracted from the EXPRES corpus developed by the Center for Corpus Related Digital Approaches to Humanities (CODHUS) at the West University of Timișoara (for more information visit: https://codhus.projects.uvt.ro/). It was written in Romanian by a political science researcher and its topic is populist attitudes and discourse in Romania – dynamics between 2019 and 2021 and possible explanations for the impact of the AUR (a young political party in Romania considered to be right-wing extremist). The text is comprised of 6,527 words (including footnotes), of which 521 were analyzed. They form 18 sentences, with an average of 29 words per sentence.

These numbers are quite high. They are however typical for Romanian academic writing, since this language displays long and intricate sentences, as opposed to the anglophone system, where sentences are shorter and the ideas more concise. No modifications were made to the original sentences so as to match the target system of writing, as I believe some interesting conclusions could be drawn from how the MT systems handled such complex sentences.

4.2. Instruments

4.2.1. Machine translation systems

Three public neural machine translation systems were selected for this study: Bing Microsoft Translator, DeepL and Google Translate. They were chosen because of their popularity in Romania and the presumption that academics would choose them to translate their own work.

4.2.2. CATMA

Despite the fact that manual annotation has been deemed slow and costly (Callison-Burch, 2009), I believe that it renders more reliable results. It is also suitable for a rather short sample. The errors were identified and annotated in the CATMA environment (Gius, et al., 2023), which is a free tool developed at the Technical University of Darmstadt, Germany. Its main advantage is that it enables users to craft their own criteria lists and it offers insights into the frequency and distribution of annotated errors using statistical analysis.

4.3. Error typology

This study makes use of the error typology developed by Maja Popović (Popović, 2018), illustrated below. It contains three error levels, in hierarchical order. The five main error categories are lexis, morphology, syntax, semantics and orthography, plus the *Too many errors category*. The last one is intended to be used for segments where errors are difficult to classify. Nonetheless, Popović (2018) advises against using it and rather advocates for trying to be as granular as possible.

Level 1	Level 2	Level 3
Lexis	Mistranslation	Terminology
	Addition	
	Omission	
	Untranslated	
	Should not be translated	
Morphology	Inflection	Tense, number, person
		Case, number, gender
	Derivation	Part of speech
		Verb aspect
	Composition	
Syntax	Word order	Range
	Phrase order	Range
Semantic	Multi-word expressions	
	Collocations	
	Disambiguation	
Orthography	Capitalisation	
	Punctuation	
	Spelling	
Too many errors		

Figure 1. Error typology *(Popović, 2018)*

While undertaking the annotation of errors it became clear that this typology cannot be used as-is; it needed to be amended due to some insufficiencies. As a consequence, *Case, number, gender* became *Case, number, gender, definiteness* because there was no way of marking the lack of or the wrong selection of definite or indefinite articles. Mood was also added to the other inflectional subcategory of Tense, number, person.

Besides this, the category of semantics also acquired a new subtag – *Style*. At first, style errors were annotated under the umbrella of *Collocations* but a change was necessary so as to avoid ambiguities.

4.4. Method

The first step was identifying and annotating fluency and accuracy errors. They were annotated separately, not at the same time, to improve the evaluation process, as Lim et al. (2024) argue for.

The second step consisted of a quantitative analysis. The errors were counted and the degree of fitness for purpose of the three NMT systems was determined based on the number of errors they displayed. The engines were then ranked from the lowest to the highest number of errors they produced, meaning that the lower the number, the better the translation. According to Chatzikoumi (2020), this is a directly expressed judgment by ranking.

4.4.1. Fluency

A fluent target text is one that follows the norms of grammaticality and naturalness of the target language (Chatzikoumi, 2020). Fluency errors were identified and annotated at sentence level and monolingually, i.e. without access to the source text. This decision was taken so as not to be influenced by the flow of the source text. The error categories considered at this point were morphology, syntax, semantics and orthography.

4.4.2. Accuracy

An accurate target text is one that preserves the information contained in the source text (Lim, et al., 2024), so there needs to be an equivalence (House, 2015) between the two. Identifying and annotating this type of errors required a bilingual and document-level analysis (Läubli, et al., 2018). The elements selected from Popović's typology for this step were the lexical ones.

4.4.3. Fitness for purpose

According to Secară, fitness for purpose denotes whether texts are "effectively usable by their consumers/ readers in pursuit of their purpose" (Secară, 2005, p. 39). For this study, it indicates whether the selected sample from the academic political science article could be published in a journal and be audience appropriate. To judge this, the three public NMT systems were ranked in accordance with the number of translation errors per each system. The assumption is that if the number of errors is high, the post-editing effort itself is larger, so the text is less fit for purpose while the MT engine is less suitable for this task.

5. Results

This section reports on the overall results of three MT systems and then dwells into a more detailed analysis, for each of them respectively. It begins with ranking them on account of the number of errors they have produced.

5.1. Ranking

Looking at Table 1, it becomes clear that the order of suitability is as follows:

1. DeepL – 17 total errors
2. Bing Microsoft Translator – 29 total errors
3. Google Translate – 30 total errors

The difference between the last two is minor, whereas between DeepL and Google Translate it is almost double. As a result, the target text produced by the first system would be easiest to post-edit. However, due to the nature and style of the source text, all translations would need full post-editing to ensure that they follow academic standards.

Name of errors	DeepL	Bing	Google	Total
Mistranslation	3	4	7	**14**
Omission	1	1	1	**3**
Total number of accuracy errors	**4**	**5**	**8**	**17**
Collocations	3	7	6	**16**
Tense, number, person, mood	4	6	2	**11**
Punctuation	2		6	**8**
Case, number, gender, definiteness	3	7	4	**8**
Style	1	2	2	**5**
Verb aspect			1	**2**
Part of speech		1		**1**
Syntax			1	**1**
Word order		1		**1**
Total number of fluency errors	**13**	**24**	**22**	**59**
TOTAL NUMBER OF ERRORS	**17**	**29**	**30**	**76**

Table 1. Overall errors

The following subsections will comment on the overall errors, then on those specific to each MT system, in order to better understand how these engines function at individual level and what users can expect when employing them.

5.2. Overall errors

In total, 76 translation errors were identified, of which 17 were accuracy errors and 59 fluency errors. Surprisingly, these results are not in line with literature findings which state that NMT systems are deceptively fluent but lack a degree of accuracy (Way, 2018). This result might be due to the fact that the number of *Collocations* errors is quite large – 16 out of 51 fluency errors, i.e. a third of the fluency errors.

When it comes to fluency errors, all engines had difficulties in selecting the most appropriate combination of terms, i.e. C*ollocations*.

Example (1), from Google Translate, showcases that the engine confused the context-dependent preposition for another one. Even though *la* is widely translated as *to*, in this case it was the wrong choice. Also, the Romanian fragment displays an awkward style itself, so it becomes even more difficult to translate properly.

(1)

a. votanții cu cele mai multe abordări populiste sunt la AUR și USR
b. the voters with the most populist approaches are ***to** AUR and USR

To keep the prepositional construction, one possible solution would have been to translate this example as: the voters *of* the AUR and USR display the most populist tendencies. It is clear however that the MT system needed to have made more operations – correct selection of the preposition and a word order rearrangement, besides correcting the mistranslation error.

Bing also encountered problems while selecting the correct preposition, this time in a phrasal verb construction, as can be seen in example (2). In fact, it did not select any, even though the right one was *about*, which forms a linguistic element that denotes the causing of an event.

(2)

a. anul 2021 aduce o modificare importantă
b. 2021 **brings** [**about**] an important change (Bing)

The same situation is displayed in example (3). The DeepL translation lacks a prepositional constituent necessary to form the phrasal verb *to keep something up* which signifies a course of action is preserved.

(3)

a. în timp ce votanții PSD tind să își păstreze atitudinile duale
b. while PSD voters tend to **keep** [**up**] their dual attitudes (DeepL)

As mentioned above, for the purposes of this research, a subcategory of *Style* was also introduced. Even though only five style errors were identified, it is worth mentioning them because they are a pain point in the development and usage of machine translation (Niu, et al., 2017; Wang, et al., 2021).

In the example below, *has as its central axis* is not necessarily a mistranslation error because such a construction is grammatical in English and is not terminologically wrong. The idea of an axis is copied from the source text, but the problem is that this wording is usually employed for mathematical texts and is strictly related to mathematical concepts – it does not have the same metaphorical connotation as in Romanian. Because of this, the phrasing becomes awkward and not suited for an academic text.

(4)

a. definiția (…) are ca ax central ideea că politica ar trebui să fie o expresie a voinței poporului
b. the definition (…) **has as its central axis** the idea that politics should be an expression of the will of the people (Google)

Again, example (5), is an instance of unidiomatic style because it directly borrows from the source text, which is itself awkward. Keeping somebody in an area indicates an idea of physicality, which is not present in this case. The reader might become confused as to where exactly the voters are kept. The text induces a sense of ambiguity, although academic writing should be as clear as possible.

(5)

a. chiar dacă {partidul} și-a păstrat și o parte din public în zone mai puțin populiste
b. even though **it** {the party} **kept some of its audience in less populist areas** (DeepL)
c. if **it** {the party} **also kept part of its audience in less populist areas** (Bing)

As for accuracy errors, the most numerous are mistranslation, with 14 total instances.

A very interesting case of a mistranslation is rendered in example (6). The author of the political science article decided to include a real-life scenario where a politician threatened a fellow colleague whom he did not agree with.

(6)

a. Dacă eram legionar știți cum procedam, da? Așa, doar cu cuvintele.
b. If I were a legionnaire, you know how I would do it, right? **Like that, just with words**. (DeepL)
c. If I was a legionnaire, you know how I would act, yes? **Like that, just with words**. (Google)
d. If I was a legionnaire, you know how I did it, right? **So, just with words**. (Bing)

The extra text reality shows us that the political figure is a member of a party considered to be right-wing extremist and the descendant of the fascist party, also known in Romania as the legionary party. In short, he tells his workmate that if he were a legionnaire, he would attack him, but since he is not, he will resort to only using words to fight him.

None of the NMT systems were able to correctly interpret this situation. A translator could have provided a translation along the lines of: *If I were a legionnaire, you know what I would do, right? Right now, I can only fight using my words*. It is very important for translated texts to not contain factual errors because they need to render the same idea as the source text. In this case, all three systems failed this task for this particular segment, producing a major error.

Moreover, this same sentence also contains another mistranslation error – *legionary* translated as *legionnaire*, even though the correct term would have been an Iron Guard member.

The only other type of accuracy errors encountered in the target texts is *Omission*. For the purposes of this study, omission refers to text that was skipped by the MT system or to quotes that do not correspond to the original version. All engines displayed an *Omission* error because they were not able to correctly back translate a quote included in the paper. Its original version was: "a political ideology of governance, which is about legitimate authority not substantive policy programs". It was included in the book *Cultural Backlash: Trump, Brexit, and Authoritarian Populism* by Pippa Norris and Roland Inglehart (Norris & Inglehart, 2019, p. 68).

The author cites this work, but the NMT systems were not able to retrieve the quote, as a translator would. Checking facts, figures and quotes is one of the most understated, yet useful skills that translators possess. Example (7) shows how the engines rendered this quote back into English.

(7)

a. o ideologie politică a guvernanței, care se referă la modalități de legitimare a autorității, și nu la politici publice de substanță
b. a political ideology of governance, ***concerned with ways of legitimising authority rather than substantive public policy** (DeepL)
c. a political ideology of governance, ***which refers to ways of legitimizing authority, and not to substantive public policies** (Google)
d. a political ideology of governance, ***which refers to ways of legitimizing authority, and not to substantive public policies** (Bing)

From this single example it can also be seen that Google and Bing tend to use American English as their target language, while DeepL tends to use British English. Even when prompted to use American English, DeepL chooses to resort back to British English. This may be due to the country where these engines are produced. Google Translate and Bing Microsoft Translator are US products, while DeepL is a German product, more geographically and culturally connected to the UK.

5.3. Results for DeepL

Out of the 17 total errors (13 for fluency and 4 for accuracy), the largest number of errors produced by DeepL are those regarding *Tense, number, person*.

Example (8) shows that verbs are in the present tense in the source text and that the NMT system copies this structure in the target text. However, this fragment is part of an enumeration where the past simple tense had previously been used.

(8)

a. După alegerile din 2020, votanții AUR devin dominanți la capitolul atitudini populiste, în timp ce votanții PSD tind să își păstreze atitudinile duale
b. After the 2020 election, AUR voters ***become** dominant in populist attitudes, while PSD voters ***tend** to keep their dual attitudes

In order to preserve the concordance of tenses that the English language is keen on, all the verbs should have been in the past simple tense. This is especially necessary since the author was not telling a story, where the lack of concordance could have been considered acceptable, but was listing events which had also taken place before the publication of his paper.

Another subcategory of errors which DeepL displays is *Case, number, gender, definiteness.* In fact, *definiteness* was the only problematic one, as can be seen in example (9).

(9)

a. Este prea devreme să afirmăm că AUR (…)
b. It is too early to say that [**the**] AUR (…)

Though it might sound awkward to a Romanian reader who knows the party and its acronym, the English language needs to add a definite article before acronyms, as it does in the case of other political institutions such as the EU and the UN. The engine unfortunately did not select this article and followed the grammatical rules of Romanian.

Despite these errors, DeepL is still the best suited NMT system in this study, as it produced the lowest number of errors – 17. As a result, it would be easiest to post-edit this sample target text.

5.4. Results for Bing Microsoft Translator

Bing Microsoft Translator produced 29 total errors – 24 for fluency and 5 for accuracy.

Besides collocation errors which have been discussed above, the most striking ones are those regarding *Case, number, gender, definiteness* and *Tense, number, person*. Of the three NMT systems included in this study, it registers the highest number of errors in these subcategories. As was the case with DeepL, *definiteness* is actually the tag that was attached to errors in this subcategory. No errors related to *case, number, gender* were identified.

Example (10) is a perfect depiction of this type of error, especially since it contains two different instances.

(10)

a. Este prea devreme să afirmăm că AUR și-ar putea revendica o anume filiație şi din naționalismul interbelic, nu doar din cel al epocii Ceaușescu
b. It is too early to say that [**the**] AUR could claim a certain lineage from ***the** interwar nationalism, not only from that of the Ceausescu era

In the first instance, the acronym of the name of the political party must be preceded by the definite article, as was discussed above. In the second one, interwar nationalism should have the zero article attached to it instead of the definite one. Most probably Bing borrowed the definite article from the Romanian text, where the morpheme *-ul*, glued to the noun *nationalism*, follows the grammatical norms of the language. English however does not follow the same rules.

When it comes to issues of *Tense, number, person*, Bing follows DeepL's pattern, displaying a lack of sequence of tenses as seen in example below.

(11)

a. Tot după aceste alegeri, votanții USR încep să devină mai populiști, adoptând tot mai multe abordări anti-elite, chiar dacă propriul partid se afla la putere.
b. Also after these elections, USR voters ***begin** to become more populist, adopting more and more anti-elite approaches, even if their own party was in power.

This NMT system also displayed two kind of errors that were not encountered in the other target texts, namely *Part of speech* and *Word order*. Interestingly, they both had to be annotated to the same grammatical structure, which is exemplified by number (12).

(12)

a. Definiția clasică a populismului se construiește în jurul a două repere – „poporul" și „elita" – aflate în opoziție și are ca ax central ideea că politica ar trebui să fie o expresie a voinței poporului.
b. The classical definition of populism is built around two landmarks – "the people" and "the elite" ***in opposition** and has as its central axis the idea that politics should be an expression of the will of the people.

In order for the target text to have complete fluency, the prepositional phrase *in opposition* should be replaced by an adjective (*opposing*), so by another part of speech. It should also be placed before the *landmarks* noun. If we wanted to keep this structure as-is, we would need to add a pronoun and a verb – *landmarks (…) that are in opposition* (…).

One error particular to Bing can be seen in example (13). The verb *to do* is in the indicative mood. It should instead be translated as *I would do it*, i.e. conditional mood. As a result, the subcategory of *mood* was added to Popović's typology.

(13)

a. Dacă eram legionar știți cum procedam, da?
b. If I was a legionnaire, you know how I ***did** it, right?

5.5. Results for Google Translate

Google Translate was the least suited engine for this task – the target text amounted to 8 accuracy errors and 22 fluency errors, so 30 errors in total. The number of errors was the highest of the three systems.

Though it is not usual for NMT systems to generate punctuation errors, this one produced six, which is quite a significant number, especially since the total number of words in the source text was not very high.

(14)

a. motiv pentru care eticheta se poate aplica unor partide
b. why the label can apply***.** to parties

(15)

a. soluții de sine stătătoare pentru problemele aflate în dezbaterea publică
b. stand-alone solutions to other problems***. *.** in the public debate

It is interesting to see however that these punctuation errors appear in the middle of a sentence, completely disrupting its flow, and not at the end of it. This might be the result of a training problem – maybe the training corpora contained such mistakes.

Other errors specific to Google Translate, at least for this study, are V*erb aspect* and *Syntax*.

The first one, *Verb aspect*, can be seen below.

(16)

a. În aceeași perioadă, electoratul PSD avea atitudini mai puțin populiste, propriul partid fiind la guvernare.
b. In the same period, the PSD electorate had less populist attitudes, their own party ***being** in government.

In Romanian it is customary to use the gerund in structures such as this one, but not in English. In MT system's rendering, the verb *to be* is in the progressive aspect, but it should be in the simple aspect. The correct translation would be: *as / because / since their own party was in government*. There is also another possibility, yet this one implies a phrase shift: *Their own party being in government, the PSD electorate had less populist attitudes in the same period*.

The next example zooms in on a *Syntax* error.

(17)

a. liderii săi nu au același comportament cu zgomotoșii lor predecesori și, spre deosebire de aceștia, nu par înclinați spre spectacol cu orice preț
b. its leaders do not have the same behavior as their noisy predecesessors and, unlike them, ***not they seem** bent on spectacle at any cost

Even though it might be advisable to use granular tags to annotate translation errors, in this case none other could fit the situation. It is clearly a pure syntax error – the auxiliary verb *do* is missing completely, the subject-verb order is not followed, the negation is not in the right position.

6. Conclusions

The results of this small-scale study point to DeepL being the most appropriate public NMT system for translating political science academic texts from Romanian into English. This is due to the fact that it produced the smallest number of errors.

For these results to be more reliable, a more detailed error typology would be needed, as well as a more complex or more varied corpus of source texts.

However, this study may offer foundational insights into the type of errors that the user can expect. As a result, academics who wish to rely on machine translation engines to translate their own articles will make an informed decision as to the exact MT system they want to use. Moreover, translators can benefit from understanding what kind of cognitive effort to expect when post-editing such texts. Developers of such systems can leverage the information presented here to modify how their products function by filtering and cleaning the corpora used for training them.

References

1. Bojar, O. et al., 2016. *Findings of the 2016 conference on machine translation.* Berlin, Association for Computational Linguistics.
2. Bowker, L. & Ciro Buitrago, J., 2019. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community.* s.l.:Emerald Publishing.
3. Callison-Burch, C., 2009. *Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk.* Singapore, Association for Computational Linguistics, pp. 286-295.
4. Chatzikoumi, E., 2020. How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering,* 26(2), pp. 1-25.
5. Delfani, J. et al., 2024. Google Translate Error Analysis for Mental Healthcare Information: Evaluating Accuracy, Comprehensibility, and Implications for Multilingual Healthcare Communication. *arXiv preprint arXiv:2402.04023.*
6. Delorme Benites, A. & Benites, F., 2021. *Neural machine translation in academic contexts.* Online, ZHAW Zürcher Hochschule für Angewandte Wissenschaften.
7. Dumitran, A., 2021. *Translation error in Google Translate from English into Romanian in texts related to coronavirus.* s.l., s.n., pp. 37-43.
8. Gius, E. et al., 2023. *CATMA 7 (Version 7.0),* s.l.: Zenodo.
9. House, J., 2015. *Translation Quality Assessment: Past and Present.* 1 ed. s.l.:Routledge.
10. https://www.himl.eu/, n.d. [Online] Available at: https://www.himl.eu/
11. Jiao, W. et al., 2023. Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine. *arXiv preprint arXiv:2301.08745.*
12. Koehn, P. & Knowles, R., 2017. *Six Challenges for Neural Machine Translation.* Vancouver, Association for Computational Linguistics, pp. 28-39.
13. Lakew, S. M. et al., 2017. *Improving Zero-Shot Translation of Low-Resource Languages.* Tokyo, Japan, s.n.
14. Läubli, S., Sennrich, R. & Volk, M., 2018. *Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation.* Brussels, Association for Computational Linguistics, p. 4791–4796.
15. Lim, Z. W., Vylomova, E., Cohn, T. & Kemp, C., 2024. Simpson's Paradox and the Accuracy-Fluency Tradeoff in Translation. *arXiv:2402.12690.*

16. Niu, X., Martindale, M. & Carpuat, M., 2017. *A Study of Style in Machine Translation: Controlling for Formality of Machine Translation Output.* Copenhagen, Association for Computational Linguistics, pp. 2814-2819.
17. Norris, P. & Inglehart, R., 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism.* s.l.:Cambridge University Press.
18. O'Brien, S., Simard, M. & Goulet, M.-J., 2018. Machine Translation and Self-post-editing for Academic Writing Support: Quality Explorations. In: J. Moorkens, S. Castilho, F. Gaspari & S. Doherty , eds. *Translation Quality Assessment. Machine Translation: Technologies and Applications.* s.l.:Springer Cham, pp. 237-261.
19. Peter, J.-T., Alkhouli, T., Guta, A. & Ney, H., 2016. *The RWTH Aachen University English-Romanian.* Berlin, Germany, Association for Computational Linguistics, pp. 356-361.
20. Popović, M., 2018. Error Classification and Analysis for Machine Translation Quality Assessment. In: J. Moorkens, S. Castilho, F. Gaspari & S. Doherty, eds. *Translation Quality Assessment. Machine Translation: Technologies and Application.* s.l.:Springer.
21. Pungă, L., Manda, I. & Chitez, M., 2023. How much Romanian does Google Translate know? A corpus-informed genre-specific error analysis of English-into-Romanian translations. *Studies about Languages,* Issue 43, pp. 77-92.
22. Rios Gaona, M. A., Chereji, R.-M., Secară, A. & Ciobanu, D., 2023. *Quality Analysis of Multilingual Neural Machine Translation Systems and Reference Test Translations for the English-Romanian language pair in the Medical Domain.* Tampere, Finland, European Association for Machine Translation, pp. 355-364.
23. Rios, M., Chereji, R.-M., Secară, A. & Ciobanu, D., 2022. *Impact of Domain-Adapted Multilingual Neural Machine Translation in the Medical Domain.* Luxembourg, arXiv.
24. Saadany, H. et al., 2021. Challenges in translation of emotions in multilingual user-generated content: Twitter as a case study. *arXiv preprint arXiv:2106.10719.*
25. Sagarra, N., 2014. Absence of morphological transfers in beginners: evidence from eye tracking. In: Z. Han & R. Rast, eds. *First Exposure to a Second Language. Learners' Initial Input Processing.* s.l.:Cambridge University Press, pp. 139-170.
26. Scansani, R., 2020. *Machine translation for institutional academic texts: Output quality, terminology translation and post-editor trust. Phd Thesis.* Bologna, Italy: University of Bologna.
27. Secară, A., 2005. *Translation Evaluation: A State of the Art Survey.* Leeds, Centre for Translation Studies, University of Leeds, pp. 39-44.
28. Steigerwald, E. et al., 2022. Overcoming Language Barriers in Academia: Machine Translation Tools and a Vision for a Multilingual Future. *BioScience,* October, 72(10), pp. 988-998.
29. Wang, Y., Cuong, H. & Federico, M., 2021. *Towards Modeling the Style of Translators in Neural Machine Translation.* Online, Association for Computational Linguistics, pp. 1193-1199.
30. Way, A., 2018. Quality Expectations of Machine Translation. In: J. Moorkens, S. Castilho, F. Gaspari & S. Doherty, eds. *Translation Quality Assessment. Machine Translation: Technologies and Application.* Dublin: Springer, pp. 159-178.