Analele Universității de Vest din Timișoara. Seria Științe Filologice
2024, Nr. 62/1, pp. 27-41
https://doi.org/10.35923/AUTFil.62-1.03

CAN DEEPL SPEAK EUROPEAN PARLIAMENTESE? A CASE STUDY INTO THE ERROR ANALYSIS OF A REPORT TRANSLATED INTO ROMANIAN

Alina RĂDOI

West University of Timișoara
alina.radoi@e-uvt.ro
https://orcid.org/0009-0006-6176-2035

Can DeepL speak European Parliamentese? A case study into the error analysis of a report translated into Romanian

Machine translation has gained much traction in the last decade, but research has been rather English-or Western-centric. This study looks at the errors produced by a public neural machine translation system (DeepL) when translating a European Parliament report from English into Romanian, an Eastern European language. More precisely, it focuses on a report published by the Committee on Women's Rights. This paper puts forth a fine-grained report on the nature and severity of these errors while assigning a quality score and a pass/ fail threshold to the output. All errors are manually annotated using CATMA, a digital text markup tool. The error typology focuses on the grand categories of accuracy, fluency and fitness for purpose, using the Multidimensional Quality Metrics (MQM) framework as the basis. The qualitative analysis accompanied by examples offers translators and machine translation system developers an insight into the kind of errors they should expect and correct, as well as a template for calculating translation quality scores.

KEYWORDS:

DeepL; error analysis; European Parliament; machine translation; translation quality; Romanian.

Introduction

Machine translation is a key topic in the world of translations studies, one that has been stealing the spotlight in the last decade. However, not all languages get the same star treatment. For example, Romanian could be considered one of the middle tier languages. More research needs to be done on it.

This study aims to fill this research gap, especially since Romanian will become more important due to Moldova's future accession to the EU, which will open up a new market. Information needs to reach citizens fast – in their own language –, and one of the ways in which this process is mitigated is by employing machine translation even within the European institutions themselves (*eTranslation*, n.d.; Txabarriaga, 2023).

Due to the sheer volume of texts that need to be translated, European institutions outsource their translations. What is more, even external translators might resort to using public machine translation systems. The ELIA report (*MQM Council – MQM*

(Multidimensional Quality Metrics), n.d.) shows that such engines are in fact part of the translation workflow.

Still, no system is perfect and errors might appear. Translators need to be aware of such faults, especially when it comes to documents that reach a wider audience, such as European ones. This study aims to bring to the front the error analysis on a DeepL-translated report published by the European Parliament, more precisely by the Committee on Women's Rights.

The methodology, results and conclusions are described below.

Previous work

Romanian is included in European-scale projects in the area of Digital Humanities, such as ParaCrawl (*Releases*, n.d.) which develops parallel corpora for all official European languages. The European Language Resource Coordination (ELRC) (*ELRC-SHARE*, n.d.) also collected and shared data in Romanian which then led to the birth of the eTranslation machine translations system, which is used by the European Commission. This work was continued by the Neural Translation for the EU (NTEU) project, which aimed to build "near-human quality machine translation engines built on industry tested neural networks technologies to/from all official EU languages except English" (*NTEU – Neural Translation for the EU*, n.d.).

Romanian was also included in the EU Presidency Translator project, where the PresidencyMT system provided text, document and website translation in all 24 official European languages (*CEF eTranslation EU Council Presidency Translator Disrupts Language Technology Landscape*, n.d.).

Pinnis *et al.* (2020) analysed the usage data from this Presidency period regarding the use of PresidencyMT system. When it comes to Romanian, the system translated 1.2-1.3 million words per month, in mostly text and document form, the language combinations being English and Romanian, in both directions (Pinnis *et al.*, 2020).

As for academic research, Romanian has been deployed mostly as a candidate target language, though many of these projects focus on the development of NMT systems (Gehring *et al.*, 2017; Garica *et al.*, 2020; Peter *et al.*, 2016; Lakew *et al.*, 2017).

Still, there have been attempts at critically analyzing the errors produced by machine translation systems, be they public or fine-tuned. Rios Gaona, et. al (2023) have studied the quality of machine-translated medical texts from English into Romanian. Previously, the same researchers (Rios *et al.*, 2022) focused on terminology errors in a medical paper abstracts corpus for the same language combination.

Kriston (2017) compared three public NMT systems (Google Translate, SDL Free Translator and Bing Microsoft Translator), highlighting grammatical and lexical errors in a business article translated from Romanian into English. Google Translate seems to be a favourite among Romanian researchers, as Dumitran (2021) looked at the errors it produced in translating texts about the coronavirus pandemic, while Pungă *et al.* (2023) focused on everyday and newspaper / news releases language, both in the English into Romanian direction.

To the best of my knowledge, no research has been undertaken for a detailed error

analysis on European documents translated into Romanian using DeepL.

Research questions

This study aims to answer two research questions:

- RQ1: What kind of errors appear in the target text?
- RQ2: How severe are these errors?

Methodology

Corpus text and machine translation system

The source text is a report published by the Committee on Women's Rights and Gender Equality in 2023 which is authored by this committee only. I selected this type of text so as to ensure fidelity to the language and terminology that this institutional body would use on its own.

The report is called *Report on sexual harassment in the EU and MeToo Evaluation (2022/2138(INI))* and contains 7,602 words, of which 1,318 words were analysed for the purposes of this research.

The paragraphs that were extracted were carefully selected to ensure that each structural element of the text is properly represented. I was mindful to follow a proportion principle – if, for example, the explanatory statement was rather long, I would select more or longer paragraphs from that particular section.

The target text ended up containing 1,468 words. It was produced using DeepL, which employs convolutional neural networks and can translate up to 1,500 characters at once.

Due to its length, it was not possible to translate the source text as a whole, so it had to be split into sections. Even though this might affect the quality of the translation because the entire context is not provided, this is how these public instruments are working at the moment and how the general user would need to navigate their interface.

Digital error annotation tool

After translating it, the target text was uploaded to CATMA[1], a free annotation tool developed at the University of Darmstadt, in Germany. It allows for the creation of a list of annotation categories and for marking up the text. As can be seen in the screenshot below, which contains a tutorial example, on the right-hand side there exist the tags that were decided upon by the user, whereas the text to be annotated is on the left-hand side.

[1] https://catma.de/.

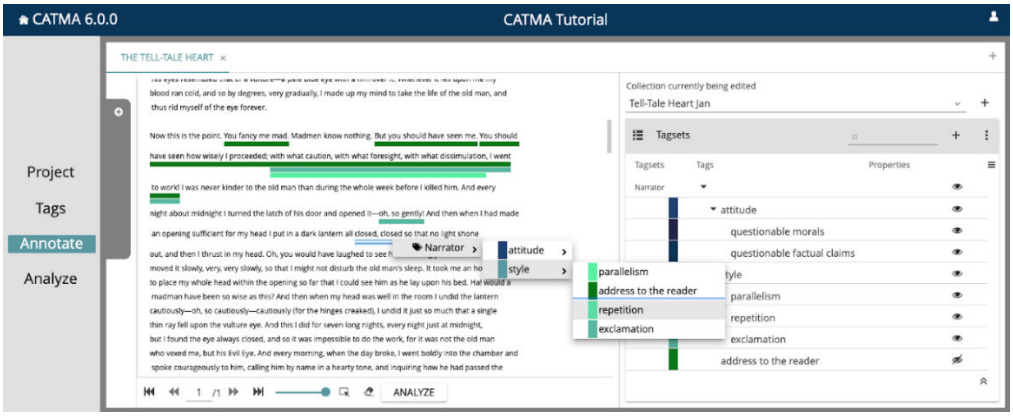

Figure 1. Screenshot from CATMA

This tool also displays an analysis and visualization component. Analyzing the frequency of errors is paramount to this research, as it indicates what category of errors is the most pervasive or which would require the most effort (Daems *et al.*, 2017) to post-edit.

Error typology

For the purposes of this research, the Multidimensional Quality Metrics (MQM) framework was used as the error typology. Its goal is to enable users to identify specific problems associated with words and phrases in the text (*MQM Council – MQM (Multidimensional Quality Metrics)*, n.d.). It follows that it is a rather "Skopos-oriented perspective on translation quality" (Lommel *et al.*, 2014, p. 2).

The error analysis undertaken in this study is comprised of the following three steps:

1. **Fluency analysis** – the target text is read throughout and fluency errors are annotated monolingually (without reading the source text). Fluency errors are those pertaining to the MQM categories of:
    - linguistic conventions
    - style
    - locale conventions
2. **Accuracy analysis** – the source and target text are compared bilingually. Accuracy errors are those pertaining to the MQM categories of:
    - terminology
    - accuracy
3. **Fitness for purpose analysis** – the errors are be weighted and assigned scores.

Step one consists of a document-level analysis, while step two displays a sentence-level analysis. Läubli *et al.* (2018, p. 4794) argue that document-level assessment is paramount, because errors become "hard or impossible to spot in a sentence-level

evaluation" due to great improvement in machine translation systems. The same authors suggest that this type of assessment brings forth mistranslations of an ambiguous word or textual cohesion and coherence issues (Läubli *et al*., 2018).

Error severity

After classifying the errors, they were assigned a *critical* / *major* / *minor* tag, depending on the severity of the error. Their definition can be found below – it is loosely based on the definitions that DGT formulates in their *Info pack for external contractors* (European Commission, 2024).

- A *critical error* is one that might lead to severe consequences (financial losses, political conflicts, reputation damage) or one that "requires extensive efforts: thorough revision or several corrections/interventions throughout the document" (European Commission, 2024, p. 12);
- A *major error* is one that necessitates correction because it otherwise leads to the text being unintelligible, unusable or unreliable or leads to a loss or change in meaning;
- A *minor error* is one that necessitates correction, but it does not render the text unintelligible, unusable or unreliable, and can be corrected without comparing the target text to the source text.

Results

Table 1 showcases the total number of fluency and accuracy errors encountered in this target text, together with a fractioning of the type of mistakes, as well as their severity level.

Though the difference between them is not that great, fluency errors are more prevalent than accuracy errors – 27 to 21 mistakes. This goes against the established literature in the field which states that NMT systems are deceptively fluent but lack accuracy (Way, 2018). However, no fluency error was assigned the *critical* tag, so even though their number is higher, their impact is lower. The text displayed 8 critical accuracy errors.

Some errors will be presented in more detail below.

Step	Parent category	Child category 1	Child category 2	Child category 3	Minor errors	Major errors	Critical errors	Total
FLUENCY ANALYSIS (STEP 1)	LINGUISTIC CONVENTIONS	duplication			2			2
			function words		1			1
			word form	part of speech	1			1
		punctuation			4	1		5
		spelling	capitalization		6			6
		unclear reference				2		2
	STYLE	awkward style			2	1		3
		unidiomatic style			2	5		7
Total fluency errors					**18**	**9**	**-**	**27**
ACCURACY ANALYSIS (STEP 2)	TERMINOLOGY	inconsistent with terminology resource	organization terminology		1	1	6	8
		wrong term			1	2		3
	ACCURACY	do not translate					1	1
		mistranslation	-		2	5	1	8
			overly literal			1		1
Total accuracy errors					**4**	**9**	**8**	**21**
Total number of errors (fluency + accuracy)					**22**	**18**	**8**	**48**

Table 1. Errors in the report translated by DeepL

Fluency errors

The target text contains 27 fluency errors, i.e. linguistic conventions and style errors. Most of them – 18 – are minor errors, while the rest are major errors.

The most prevalent linguistic conventions error is *capitalization*, as can be seen in example (1). However, this is a minor error since it can be corrected without looking at the source text and because it does not render the text unusable or unreliable.

(1)

a. 33. Instructs its President to forward this resolution to the Council and the Commission.

b. 33. 33. încredințează Președintelui sarcina de a transmite prezenta rezoluție Consiliului și Comisiei.

This example also contains a *duplication* error that is minor nonetheless – DeepL doubled the number 33.

Another pervasive but minor error was *punctuation*, displayed in example (2). The engine selected the correct quote mark for Romanian, but failed to place it at the bottom of the page.

(2)

a. having regard to Annex II of its Rules of Procedure entitled 'Code of Appropriate Behaviour for Members of the European Parliament in Exercising their Duties'

b. având în vedere anexa II la Regulamentul său de procedură intitulată "Codul de comportament adecvat pentru deputații în Parlamentul European în exercitarea atribuțiilor lor"

One *punctuation* error that I believe to be major is the following:

(3)

a. having regard to the *European Institute for Gender Equality's Gender Equality Index 2022* report

b. având în vedere raportul *Indexul privind egalitatea de gen 2022 al Institutului European pentru Egalitatea de Gen*

Though it might not be necessary in English to add the quote marks while naming the report, I believe that not doing so in Romanian may render the passage unintelligible. The reader could ask themselves if the report is in fact named like this or if the text should read *raportul privind indexul de egalitate* or even if the name of the institute is included in the title of the report.

Another major error pertains to the subcategory of *unclear reference*. In example (4), it is unclear what *sale* refers to – is it the institution or the mechanism? As a result, the paragraph becomes confusing to the reader.

(4)

a. Is convinced that the European institutions should behave as exemplary employers, establishing zero-tolerance standards towards any type of harassment, working actively on harassment prevention, adequate victim protection and holistic supporting mechanism, countering all forms of discrimination, implementing *its* rules strictly

b. Este convins că instituțiile europene ar trebui să se comporte ca angajatori exemplari, stabilind standarde de toleranță zero față de orice tip de hărțuire, lucrând în mod activ la prevenirea hărțuirii, la o protecție adecvată a victimelor și la un mecanism holistic de sprijin, combătând toate formele de discriminare, aplicând cu strictețe normele *sale*

The same I believe to hold true for example (5), where the *unclear reference* error becomes major because it is not clearly stated what the object of the improvement is – improvement towards combatting gender-based violence or improvement towards

gender equality in the EU?

(5)

a. Recalls that gender equality is a core value of the EU and must be mainstreamed in all EU policies, activities and programmes; regrets the slow progress towards gender equality in the EU and recalls that gender-based violence is both a cause and a consequence of gender inequality; stresses the urgent need for *improvement* and insists that the EU and its Member States fully commit to speeding up progress including by implementing gender mainstreaming and gender budgeting in all EU policies, activities and programmes;

b. Reamintește că egalitatea de gen este o valoare fundamentală a UE și trebuie să fie integrată în toate politicile, activitățile și programele UE; regretă progresele lente înregistrate în ceea ce privește egalitatea de gen în UE și reamintește că violența bazată pe gen este atât o cauză, cât și o consecință a inegalității de gen; subliniază nevoia urgentă de *îmbunătățire* și insistă ca UE și statele sale membre să se angajeze pe deplin în accelerarea progreselor, inclusiv prin punerea în aplicare a integrării dimensiunii de gen și a includerii dimensiunii de gen în buget în toate politicile, activitățile și programele UE;

Of course, this error appears as a result of using a convoluted style of writing in the source text, despite the fact that EU institutions also advocate for plain language[2]. However, I believe that it is the job of the translator to make texts as accessible and clear as possible, so this mistake can be imputed to the NMT system, but the system is however not able to correct it.

Moving on to style errors, the most frequent ones were part of the *unidiomatic style* subgroup – 7 total errors, 5 of which were major. Two instances can be seen in the examples below.

(6)

a. highlights the importance of *addressing* cyber violence in the workplace

b. subliniază importanța *abordării* violenței cibernetice la locul de muncă

(7)

a. calls on Member States to pro-actively design and implement legislation and policies that *tackle* sexual violence and harassment in our society

b. invită statele membre să elaboreze și să pună în aplicare în mod proactiv o legislație și politici care *să abordeze* violența și hărțuirea sexuală în societatea noastră

Both *to address* and *to tackle* have been translated using *a aborda* in Romanian. Still, neither of them is correct. The usual collocations[3] used for *a aborda* is *a aborda o problemă* or *a aborda o persoană*. In recent times however, such translations are more

[2] https://op.europa.eu/en/web/accessibility/transcript-plain-language.
[3] https://dexonline.ro/definitie/aborda.

prevalent because Romanian speakers borrow this term from English. Due to the fact that at this point in time it is considered to be a grammatical mistake that needs correction, and one which might render the text unsuable or unreliable, these errors will be counted as major.

As for *awkward style*, which denotes excessive wordiness, I have identified one major error, highlighted by example (8), and two minor ones.

(8)

a. including the introduction of mandatory harassment prevention training for all Members *as soon as they take office at the beginning of their mandate*

b. inclusiv introducerea unei formări obligatorii de prevenire a hărțuirii pentru toți deputații *de îndată ce intră în funcție la începutul mandatului lor*

This passage indicates sloppy writing – it is only logical that Members begin their mandate when they take office. The error is carried through into the target text. Since an NMT system cannot delete information from its output (unless it omits it by mistake), this pleonasm is visible in the Romanian version, too. A translator would at least need to correct it as *la începutul mandatului, de îndată ce intră în funcție*, to highlight the urgency of this matter. This process would however require seeing the text as a coherent structure and having real-world experience, which is something that a machine does not possess.

Let us compare this to the example below which contains a minor *awkward style* error.

(9)

a. very few of these concrete measures have been fully implemented *and more needs to be done*

b. foarte puține dintre aceste măsuri concrete au fost puse în aplicare pe deplin *și mai sunt necesare mai multe eforturi*

Though not necessarily wrong – grammatically speaking – repeating *mai* creates an irritating rhythm, one that could have been avoided has the first one been deleted, especially because it was not compulsory within this structure.

Accuracy errors

There are 21 accuracy (terminology and accuracy) errors in the target text. Most of these – 9 – are major errors.

Focusing on terminology first, we can see that there are 5 critical *organization terminology* errors. These are usually institution names or document titles that have been wrongly translated, as in the examples below. I believe that they are critical errors because they might cause reputation damage to the European Parliament by misinforming the citizens and also because the entire document would need extensive revision to make sure that all such terms are correct. Option c. showcases the correct

translation.

(10)

a. having regard to the *European Institute for Gender Equality's Gender Equality Index 2022 report*

b. având în vedere raportul **Indexul privind egalitatea de gen 2022 al *Institutului European pentru Egalitatea de Gen*

c. având în vedere raportul „Indicele egalității de gen 2022[4]" al Insitutului European pentru Egalitatea de Șanse între Femei și Bărbați[5]

(11)

a. having regard to Annex II of its Rules of Procedure entitled '*Code of Appropriate Behaviour for Members of the European Parliament in Exercising their Duties*'

b. având în vedere anexa II la Regulamentul său de procedură intitulată "**Codul de comportament adecvat pentru deputații în Parlamentul European în exercitarea atribuțiilor lor*"

c. având în vedere anexa II la Regulamentul său de procedură institulată „Codul comportamentului adecvat al deputaților în Parlamentul European în exercitarea funcțiilor lor[6]"

Example (12) contains another *organization terminology* error, but this one is only major due to the fact that there is no fixed translation for this term in the IATE inventory.

(12)

a. having regard to the Bureau Decision of 2 July 2018 on the functioning of the advisory *committee dealing with harassment complaints concerning Members of the European Parliament* and its procedures for dealing with complaints

b. având în vedere Decizia Biroului din 2 iulie 2018 privind funcționarea *comitetului consultativ care se ocupă de plângerile de hărțuire referitoare la deputații în Parlamentul European* și procedurile sale de soluționare a plângerilor

There is one entry[7] in IATE for an obsolete term, *Advisory Committee dealing with Harassment Complaints between Accredited Parliamentary Assistants and Members of the European Parliament and its Prevention at the Workplace*, which is translated as *Comitetul consultativ privind hărțuirea la locul de muncă și prevenirea acesteia, în ceea ce privește plângerile depuse de asistenți parlamentari acreditați împotriva unor deputați în Parlamentul European*.

[4] https://eige.europa.eu/sites/default/files/the_index_press_release_final_ro.pdf

[5] https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/european-institute-gender-equality-eige_ro

[6] https://www.europarl.europa.eu/at-your-service/ro/transparency/appropriate-behaviour

[7] https://iate.europa.eu/entry/result/3569641/en-en-ro

Following this example, the correct term could have been *comitetul consultativ privind plângerile de hărțuire împotriva unor deputați în Parlamentul European*. A human translator (and terminologist) was absolutely needed in this case to solve the issue and to ensure terminological accuracy.

Funnily, DeepL did not even select the correct translation for the name of the Parliament Committee the source text is written by. That is why example (13) was categorized as a critical error. The correct translation can be seen in c.

(13)

a. Committee on Women's Rights and Gender Equality

b. Comisia pentru drepturile *femeii și egalitatea de gen

c. Comisia pentru drepturile femeilor și egalitatea de gen[8]

Moving on to *wrong term* errors, there is one error in the minor severity category and two in the major one. Example (14) puts forth the minor one. I tagged it like this because *legislator* is usually translated as *legiuitor* in Romanian and the subject is even called as such in Romanian law. Even though *legislator* used to be a rarely used term in Romanian, it is now given as a synonym in DEX.[9]

(14)

a. both as the directly elected representatives of EU citizens and as *legislators*

b. atât în calitate de reprezentanți aleși în mod direct de către cetățenii UE, cât și în calitate de *legislatori*

In the example below there appears a major *wrong term* error, as *bystander* was wrongly translated as *spectatori*, when it should have been *martori.*

(15)

a. whereas research shows harassment training must be complemented by *bystander* training

b. întrucât cercetările arată că formarea în domeniul hărțuirii trebuie să fie completată de o formare a *spectatorilor*

When it comes to errors pertaining to the MQM accuracy category they are split into two: *do not translate* and *mistranslation*.

The first one is quite straightforward, exemplified by (16). DeepL translated the title of a report (as part of the footnotes) that only has an English version. It can lead to severe consequences should the reader want to look up the document as they cannot find it under the Romanian name. The mistake was marked as critical.

(16)

a. International Labour Organization (ILO), Lloyd's Register Foundation,

[8] https://www.europarl.europa.eu/committees/ro/femm/home/highlights
[9] https://dexonline.ro/definitie/legislator/definitii

Experience of violence and harassment at work: the first global survey, 2022.

b. Organizația Internațională a Muncii (OIM), Fundația Lloyd's Register, **Experiența violenței și hărțuirii la locul de muncă: primul sondaj global*, 2022.

Mistranslation errors are also split in two – the child 1 category (*mistranslation* per se) and child 2 category (*overly literal*). Looking at the example below where the error was marked as *overly literal*, one can see that the phrase could have a negative connotation, so exactly the opposite intention. *A ieși în față cu ceva* means in Romanian to boast about something. Because it produced a change in the meaning of the original text, this mistake is a major one.

(17)

a. applauds the millions of people who *came forward with* their stories of sexual violence to break the silence and seek justice

b. salută milioanele de persoane care **au ieșit în față cu* poveștile lor de violență sexuală pentru a rupe tăcerea și a căuta dreptate

There are 8 *mistranslation* errors, 5 of which are major. Interestingly, all of these relate to the verb *a aborda*. The reasoning behind assigning this tag has been explained in detail in the Fluency subsection above.

The only critical error, which triples as a *mistranslation*, *organization terminology* and *capitalization* error, can be seen below. In this case, it is a critical *mistranslation* error because it might lead to misinformation, so to reputation damage to the European Parliament. Option c. indicates the correct translation.

(18)

a. having regard to its resolution of 14 February 2023 on the Council of Europe Convention on preventing and combating violence against women and domestic violence: *EU accession*

b. având în vedere Rezoluția sa din 14 februarie 2023 referitoare la Convenția Consiliului Europei privind prevenirea și combaterea violenței împotriva femeilor și a violenței domestice: **Aderarea la UE*

c. având în vedere Rezoluția sa din 14 februarie 2023 referitoare la Convenția Consiliului Europei privind prevenirea și combaterea violenței împotriva femeilor și a violenței domestice: aderarea UE

Aderarea UE (EU accession) and *aderarea la UE* (accession to the EU) are completely different matters.

Error weights and penalties

Table 2 shows the general error weight for each parent category identified to be present in the target text translated by DeepL. A raw analysis was made, along with a pass/fail score calculation.

The pass score was set at 85%, as I believe that any score below this would imply

that the post-editing effort (Chatzikoumi, 2020) would be too great and it would be easier to just translate the text from scratch. However, it would be unrealistic to believe that such a machine-translated text could in any way be used as-is, without post-editing it, especially when considering its importance – it is a publicly-available document issued by a freely-elected European institution whose purpose is to inform around 20 million Romanian speakers.

The calculation was done according to MQM Council instructions (*MQM Council – MQM (Multidimensional Quality Metrics)*, n.d.). The first step was calculating the Error Type Penalty Total (ETPT) score.

		Minor	**Major**	**Critical**	
Severity Penalty Multiplier		1	5	25	**Error Type Penalty Total (ETPT)**
Error Type	Linguistic conventions	14	3	-	**29**
	Style	4	6	-	**34**
	Terminology	2	3	6	**167**
	Accuracy	2	6	2	**82**
Absolute Penalty Total		**22**	**90**	**200**	**312**

Table 2. Error weights

The ETPT was calculated as follows: ((Minor Error Type count × Minor Severity Multiplier) + (Major Error Type count × Major Severity Multiplier) + (Critical Error Type Count × Critical Severity Multiplier)) × Error Type Weight. The Error Type Weight is just 1, according to the MQM Council's recommendations.

The Severity Multiplier is displayed at the top of the table – 1 for minor errors, 5 for major errors and 25 for critical errors. It indicates the severity of the error through exponentially increasing numbers.

After this came the Absolute Penalty Total (APT), which is the sum of all ETPTs. All in all, the APT for the *Report on sexual harassment in the EU and MeToo evaluation (2022/2138(INI))* translated by DeepL is 312 – 249 penalty points for accuracy and 63 penalty points for fluency. What this result shows is that accuracy errors were much more severe than fluency errors.

Next, the Per-Word Penalty Total (PWPT) was calculated by dividing the APT to the Evaluation Word Count (the number of target-text words). In this case, 312 divided by 1,468, so the result is 0.21. To calculate the Raw Quality Score (RQS), the PWPT must be subtracted from 1. The result was 0,79.

In percentages, the result would be 79%, so the target text reached a 79% quality score, meaning that it did not receive a pass score, the minimum of which was set at 85%. It requires significant post-editing operations.

Conclusions

The MQM framework proved to be a reliable measure of machine translation quality, in terms of both accuracy and fluency. Due to its flexible structure, the user can pick and choose those categories that best fit their particular scenario and calculate objective quality scores.

This paper has shown that DeepL is moderately appropriate for translating a European Parliament report from English into Romanian, at least on the topic of sexual harassment. However, the output requires full post-editing and minute attention to details on the part of the post-editor since DeepL has made eight critical errors in total. The number might even be regarded as too large by some evaluators, and they could advise simply to translate the report from scratch by a professional.

No matter the decision, this article has tried to showcase what translators could expect when using public machine translation systems on such a genre of texts. It might even prove useful to developers wishing to improve their product.

Further investigations will be made on this topic in order to acquire a fuller picture of the state of these engines and the quality of their output.

References:

CEF eTranslation EU Council Presidency Translator Disrupts Language Technology Landscape. (n.d.). Retrieved September 13, 2024, from https://ec.europa.eu/digital-building-blocks/sites/digital-building-blocks/sites/pages/viewpage.action?pageId=533365211.

Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, *26*(2), 137–161. https://doi.org/10.1017/S1351324919000469.

Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017). Identifying the Machine Translation Error Types with the Greatest Impact on Post-editing Effort. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01282.

Dumitran, A. (2021). Translation Error in Google Translate From English Into Romanian in Texts Related to Coronavirus. *Conference Proceedings of eLearning and Software for Education (eLSE)*, *17*(02), 37–43.

ELRC-SHARE. (n.d.). Retrieved September 13, 2024, from https://www.elrc-share.eu/.

eTranslation. (n.d.). Retrieved September 18, 2024, from https://commission.europa.eu/resources-partners/etranslation_en.

European Commission. (2024). *Translation quality evaluation. Info pack for external contractors*. European Commission.

Kriston, A. (2017). *Machine translation in translating business texts: Myth versus reality*. https://dspace.upt.ro/xmlui/handle/123456789/5263.

Läubli, S., Sennrich, R., & Volk, M. (2018). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4791–4796). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1512.

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica*, *12*, 0455–0463.

MQM Council – MQM (Multidimensional Quality Metrics). (n.d.). Retrieved May 14, 2024, from https://themqm.org/mqm-council/.

NTEU – Neural Translation for the EU. (n.d.). Retrieved September 13, 2024, from https://nteu.eu/.

Pinnis, M., Bergmanis, T., Metuzāle, K., Šics, V., Vasiļevskis, A., & Vasiļjevs, A. (2020). A Tale of Eight Countries or the EU Council Presidency Translator in Retrospect. In J. Campbell, D. Genzel, B. Huyck, & P. O'Neill-Brown (Eds.), *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 2: User Track)* (pp. 525–546). Association for Machine Translation in the Americas. https://aclanthology.org/2020.amta-user.25.

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*, *43*, Article 43. https://doi.org/10.5755/j01.sal.1.43.35154.

Releases. (n.d.). Retrieved September 13, 2024, from https://paracrawl.eu/.

Txabarriaga, R. (2023, June 16). *EU Starts Machine Translating Press Content After 'Little Housekeeping Announcement.'* Slator. https://slator.com/eu-starts-machine-translating-press-content/.

Way, A. (2018). Quality Expectations of Machine Translation. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 159–178). Springer International Publishing. https://doi.org/10.1007/978-3-319-91241-7_8.