Lexical and Syntactic Diversity: Still Lost in Machine Translation?

Lise Volkart
FTI/TIM, University of Geneva
Switzerland
lise.volkart@unige.ch

Pierrette Bouillon
FTI/TIM, University of Geneva
Switzerland
pierrette.bouillon@unige.ch

Abstract

This paper investigates whether recent developments in machine translation (MT) have affected two well-documented stylistic characteristics of MT: reduced lexical diversity and increased source syntax mirroring. Using a corpus of English into French translations, we compared MT outputs produced by a widely used system over time (2024 vs. 2026) and under different models (NMT vs. LLM-based). Lexical diversity is measured using word translation entropy (HTra), and source syntax mirroring using four syntactic similarity metrics. Results show that lexical and syntactic characteristics of NMT output remain stable over time. The LLM-based MT exhibits a slight increase in lexical diversity, but still presents strong signs of lexical overgeneralisation. Syntactic similarity to the source remains largely unchanged. Overall, the findings suggest that recent MT developments have a very limited impact on these two stylistic characteristics.

1 Introduction

Machine translation (MT) is generally considered a rapidly evolving field. Yet, since 2016, and the so-called neural revolution, progress has been more gradual, and despite the emergence of large language models (LLMs), conventional neural machine translation (NMT) remains a relevant technology (Lazarov, 2026). In terms of output quality, both technologies achieve state-of-the-art performance, particularly for high-resource languages, typically written in Latin scripts (Ataman et al., 2025; Lazarov, 2026). In the past few months, several traditional MT providers announced the implementation of so-called LLM-based MT architectures or the integration of LLM-related features, with claims of improved output quality.[1] Now, beyond general quality, it remains unclear whether such a technology shift affects the stylistic properties of the MT output. Previous work has shown that NMT output differs from human translation (HT), particularly through reduced lexical diversity and increased literalness, with higher syntactic similarity to the source (Vanmassenhove et al., 2019; Toral, 2019; Tezcan et al., 2019; Webster et al., 2020; Vanmassenhove et al., 2021; Volkart and Bouillon, 2024; Volkart, forthcoming). It is timely to determine whether such findings remain valid over time and in light of recent developments of LLM-based MT. In this paper, we address this question through a corpus-based study of English into French MT outputs produced by a widely used online MT system at different points in time, and under different models. Specifically, we compare outputs generated by NMT in 2024 and 2026 with those produced in 2026 under an LLM-based architecture. Using established metrics of lexical diversity and syntactic similarity, we aim to assess both the diachronic evolution of NMT and the potential stylistic impact of LLM-based MT.

Our main objective is to determine whether pre-

[1] See for instance: https://www.supertext.com/en/technology [Accessed: April 2026]
https://www.deepl.com/en/blog/next-gen-language-model [Accessed: April 2026]
https://blog.google/products-and-platforms/products/search/gemini-capabilities-translation-upgrades/ [Accessed: April 2026]

vious findings on the lexical and syntactic characteristics of MT output are still relevant in the context of recent technological developments. More precisely, we pursue two complementary goals: (i) to evaluate the evolution of NMT output over time with respect to lexical diversity and syntactic similarity, and (ii) to assess whether the introduction of LLM-based MT leads to substantial stylistic changes.

We formalise these goals under the following research question:

> *Did lexical and syntactic features of MT output evolve over time and with the integration of LLM-based MT?*

The lexical and syntactic analysis conducted shows that, despite recent technological developments, the stylistic characteristics previously observed in NMT output remain largely unchanged, suggesting that LLM-based MT does not appear to constitute a major shift on the stylistic level.

The remainder of the paper is structured as follows: Section 2 presents related work, Section 3 describes the background of our study as well as the data and metrics. Our results are presented in Section 4 and discussed in Section 5. Finally, Section 6 concludes the paper and Section 7 outlines limitations and directions for future research.

2 Related work

Since the advent of NMT systems around 2016, numerous studies, such as Martikainen and Kübler (2016), Toral (2019), Webster et al. (2020), Vanmassenhove et al. (2019), Vanmassenhove et al. (2021) or Volkart and Bouillon (2024), have shown that MT tends to exhibit a lower lexical variety than HT and to mirror source syntax, therefore producing more literal translations. More recent work has begun exploring translation produced by LLMs, mainly in terms of output quality. A number of studies have demonstrated that LLMs can achieve comparable or higher performance than conventional NMT systems, especially on high-resource languages (Hendy et al., 2023; Ataman et al., 2025; Lazarov, 2026). Regarding style, LLMs are often argued to achieve a higher level of fluency and produce less literal translations than NMT (Li et al., 2025). Raunak et al. (2023), for instance, have shown that LLMs produce less literal translation than NMT when translating out of English, particularly for the translation of idiomatic expressions. Other works, such as Li et al. (2025), have shown that LLMs still exhibit significant amounts of *translationese*, i.e. translations that are too literal and therefore sound unnatural (Li et al., 2025). To the best of our knowledge, there are no studies yet focusing specifically on these questions for LLM-based MT output, that is translation produced by MT tools relying on an LLM architecture.

3 Methodology

3.1 Background

The analysis conducted in Volkart and Bouillon (2024) and Volkart (forthcoming) on texts collected in professional contexts have shown that NMT tends to exhibit a lower level of translation solutions variation and a higher level of syntactic similarity than HT. Specifically, these previous studies show that, on the lexical level, NMT tends to overuse the most frequent translation solutions and to underuse the less frequent ones, leading to a lexical overgeneralisation phenomenon, as already demonstrated by other authors. On the syntactic level, these experiments demonstrated that NMT tends to reproduce the source syntactic structure, leading to more literal translations. For the present study, we use the raw MT corpora produced in 2024 for these experiments and compare it to more recent MT outputs (NMT and LLM-based), to estimate whether MT has made any progress on those characteristics over the past two years.

3.2 Corpus

This work is based on two corpora (C1 and C2) compiled and described in Volkart and Bouillon (2024) and Volkart (forthcoming), containing authentic HT and PEMT from English into French collected in professional contexts, alongside a baseline raw MT generated using the online and publicly available NMT system DeepL[2] in February 2024. Original HT and PEMT corpora were collected between 2021 and 2022, C1 contains translations shared with us by the European Investment Bank (EIB), and C2 contains translations shared by a sports organization based in Switzerland. For the present experiment, we include only the baseline NMT and its source corpus. Using DeepL again, we generated two new raw MT corpora in February 2026, using the

[2]https://www.deepl.com

`classic` DeepL model (NMT) and the so-called `next-gen language model`. In DeepL's documentation[3], the classic model is presented as the standard NMT model, while the next-gen model is described as based on an LLM-based MT model. In doing so, we obtain two multiple translation corpora containing a collection of source sentences with multiple MT outputs (NMT from 2024, NMT from 2026 and LLM-based MT from 2026). Tables 1 and 2 summarize the main corpus statistics.

3.3 Metrics

The first metric we rely on is the word translation entropy HTra (Carl et al., 2016), which measures the diversity of translation solutions in a corpus. HTra is computed as the sum over all observed word translation probabilities $p(s \rightarrow t_i)$ of a given source text word s into target text word $t_i...n$ multiplied with their information content $I(p) = -log_2(p)$ (Carl et al., 2016; Volkart and Bouillon, 2024) as in the following equation:

$$H(s) = -\sum_{i=1}^{n} p(s \rightarrow t_i) \times log_2 p(s \rightarrow t_i)$$

The score reflects the amount of different translation solutions present in the target for a given source lemma, as well as the distribution of these solutions (if they are equiprobable, or if some are more represented than others). In that sense, HTra is a good indicator of the lexical diversity level in a translation corpus (for a more extensive discussion on this aspect, see Volkart (forthcoming)). The higher the variety and the more balanced the distribution, the higher HTra (Carl et al., 2016; Bangalore et al., 2016; Volkart, forthcoming).

Our second axis of research concerns the source sentence structure mirroring, which we estimate using the following metrics from the *ASTrED*[4] library (Vanroy et al., 2021):

- Syntactically Aware Cross (SACr): quantifies the degree of reordering of word sequences that occurred between source and target (Vanroy et al., 2021)
- Aligned syntactic tree edit distance (ASTrED): computes the tree edit distance between source and target dependency trees while taking word alignments into account (Vanroy et al., 2021).
- Label change for part-of-speech (POS) (Label POS): corresponds, for a given source-target sentence pair, to the number of source-target word-aligned pairs that have different POS tags, normalised by the total number of alignments for that sentence (Vanroy et al., 2021)
- Label change for dependency labels (Label DEP): corresponds, for a given source-target sentence pair, to the number of source-target word-aligned pairs that have different dependency labels, normalised by the total number of alignments for that sentence (Vanroy et al., 2021)

Each of these four metrics captures a specific aspect of the syntactic similarity between a source sentence and its translation (Volkart, forthcoming). Altogether, these metrics provide a comprehensive estimate of the degree of literality of a translation on the syntactic level.

HTra is computed for all content lemma appearing at least three times in the source corpus, based on the word-alignment generated using `awesome-align` (Dou and Neubig, 2021). Syntactic similarity metrics are computed using the ASTrED python library (Vanroy et al., 2021). Tagging and lemmatization are performed using SpaCy's transformer models for English, French and German.[5]

4 Results

4.1 HTra

HTra scores for the different MT outputs for all POS and for each POS category are presented in Table 3. We obtain very similar results for C1 and C2. First of all, HTra scores do not differ between NMT24 and NMT26, indicating that the level of lexical diversity has not improved in the translation produced by DeepL's NMT model over the past two years. Regarding the translation produced by the `next-gen` model, it demonstrates a slightly higher word translation entropy, meaning that the transition to an LLM architecture seems to have improved the lexical diversity of the MT output, and this across almost all the POS categories. The differences between NMT24 and LLM-MT26 are statistically significant according to the Wilcoxon

[3]https://support.deepl.com/hc/en-us/articles/14241705319580-About-DeepL-language-models
[4]https://github.com/BramVanroy/astred
[5]https://spacy.io/models

Subcorpus	# segments	# tokens	mean segment length
NMT24	1 852	47 294	25,54
NMT26	1 852	47 338	25,60
LLM-MT26	1 852	47 351	25,61

Table 1: Descriptives statistics for corpus C1.

Subcorpus	# segments	# tokens	mean segment length
NMT24	2 280	48 760	21,39
NMT26	2 280	52 390	22,98
LLM-MT26	2 280	52 732	23,13

Table 2: Descriptives statistics for corpus C2.

signed-rank test for all POS together, for nouns and verbs, though effect sizes[6] are generally low (ranging from -0.09 to -0.17), indicating limited practical significance. While NMT has remained stable over the past two years in terms of translation solutions diversity, the introduction of an LLM-based architecture is associated with a modest improvement that, although limited, deserves further investigation.

On the technical level, a higher HTra score can be explained by two different factors: either a higher number of unique translation solutions or a better distribution of them, and, eventually, a combination of both. To better understand the improved HTra observed for the nouns and the verbs in the LLM-based MT, we attempt to isolate the main factor contributing to this score increase. To do so, we extract all the nouns and verbs presenting a higher score for the LLM-based MT with all their translation solutions, and manually annotate the main explaining factor for the score increase (either quantity or distribution). The cases for which LLM-MT26 has an equal or lower number of translation solutions compared to NMT24, but a more balanced distribution are annotated as "due to distribution" ("distribution" hereafter). The cases for which LLM-MT26 has more unique translation solutions compared to NMT24 are annotated as "due to variety" ("variety" hereafter) and split into three subgroups according to the number of additional solutions (1, 2 and 3 or more). Results of this classification are presented in Tables 4 and 5.

Again, C1 and C2 exhibit very similar tendencies. For the nouns, in about 80% of the cases, the higher HTra observed in the LLM-based translation is due to a greater translation solutions variety rather than to a more balanced distribution. Amongst those, around 50% of cases exhibit only one additional translation solution. This is very similar for the verbs, with almost 90% of the increase cases due to a larger variety and a bit more than 50% of them having only one additional solution. This clearly indicates that the LLM-based MT, while having a slightly higher HTra score, still follows the same tendency as the NMT: it exhibits the same centralisation and generalisation tendencies on the lexical level, but with occasional use of less frequent translation solutions.

4.2 Syntactic similarity

Table 6 presents the sentence similarity scores for the different MT outputs. For all metrics, a lower score indicates a higher degree of source sentence similarity. NMT24 and NMT26 present identical scores for the four metrics. The Wilcoxon signed-rank test, though, indicates some statistically significant differences between them for some of the scores (Label DEP and AStrED for C1 and SACr Label DEP and ASTrED for C2). As this may seem counter-intuitive, it can be easily explained: the Wilcoxon signed-rank test reveals a systematic discrepancy between the paired differences, which

[6] computed as $r = \frac{z}{\sqrt{n}}$

	C1			C2		
	NMT24	**NMT26**	**LLM-MT26**	**NMT24**	**NMT26**	**LLM-MT26**
Adjectives	0,97	0,97	**1,00**	1,01	1,01	**1,04**
Adverbs	**1,10**	1,09	1,04	1,25	1,26	**1,27**
Nouns	0,71	0,72	**0,75***	0,86	0,86	**0,94**$^{\diamond}$
Verbs	1,27	1,27	**1,39**$^{\diamond}$	1,45	1,45	**1,56**$^{\diamond}$
All POS	0,92	0,92	**0,97**$^{\diamond}$	1,07	1,07	**1,14**$^{\diamond}$

Table 3: HTra scores for the different MT outputs (NMT24, NMT26 and LLM-MT26) for corpus C1 and C2. $^{\diamond}$ indicates a significant difference with NMT24 at $p < 0,001$, and * at $p < 0,05$. Significance was tested using Wilcoxon signed-rank test.

Main HTra increase factor	**Percentage of nouns**	**Percentage of verbs**
↗ distribution	20,67%	12,30%
↗ variety	**79,33%**	**87,70%**
of which $+1$	**56,98%**	**55,74%**
of which $+2$	18,44%	22,95%
of which $> +3$	3,91%	9,02%

Table 4: Percentage of nouns and verbs according to the main HTra increase factor for LLM-MT26 for corpus C1

Main HTra increase factor	**Percentage of nouns**	**Percentage of verbs**
↗ distribution	17,62%	11,11%
↗ variety	**82,38%**	**88,19%**
of which +1	**50,08%**	**52,08%**
of which +2	15,33%	20,83%
of which $> +3$	11,11%	15,28%

Table 5: Percentage of nouns and verbs according to the main HTra increase factor for LLM-MT26 for corpus C2

means that a large number of sentences have scores varying in the same direction. Considering the large sample size, it is sufficient to produce a statistically significant result. Nevertheless, the effect sizes remain very limited (ranging from 0.05 to 0.06). We interpret this as the NMT26 not producing the exact same translation as the NMT24, but still with a very similar average level of syntactic similarity with the source.

The LLM-based MT presents marginal differences with NMT24, with a slightly lower SACr score and a moderately higher ASTrED for C1, indicating that LLM-based MT presents a higher similarity in terms of word order, but a lower similarity at the level of deep syntactic structure.

For C2, some differences (Label POS, Label DEP and ASTrED) emerge as statistically significant according to the Wilcoxon signed-rank test, despite identical average scores. However, the associated effect sizes remain very small (ranging from 0.08 to 0.09), again suggesting limited practical significance. Taken together, these results indicate that, although the LLM-based MT actually differs from the NMT24 at the sentence level, the overall degree of syntactic similarity remains largely unchanged. Given the marginal differences between the subcorpora and the lack of consistency across C1 and C2, these findings are difficult to interpret. More importantly, the observed differences are substantially smaller than those reported between human translation (HT) and raw MT in other studies (Tezcan et al., 2019; Shaitarova et al., 2023; Volkart and Bouillon, 2024). This suggests that the different MT outputs are consider-

ably more similar to one another than to HT. In this respect, the well-documented tendency of MT systems to adhere more closely to source syntax than HT appears to persist.

5 Discussion

This experiment shows that the stylistic characteristics of MT output tend to be stable over time, even when switching from NMT to LLM-based MT. This may indicate that modern MT systems have reached a plateau in terms of the level of naturalness they can achieve, and the integration of LLMs does not fundamentally change the picture. This has several implications: first of all, it means that post-editing (PE) by professionals still holds significant value to ensure translations remain stylistically rich and fluent in the target language. On the other hand, it also means that the MT interference that could be measured in post-edited MT (Volkart, forthcoming) is also likely to remain unchanged. Finally, those results show that other strategies may be needed to recover lexical and syntactic diversity in translation produced by automated systems. These strategies might include directly prompting LLMs for translation or leveraging them to perform editing targeted on stylistic features. To this end, LLMs may bring significant value.

6 Conclusion

In this experiment, we compared the levels of lexical diversity and syntactic similarity of different MT outputs produced using DeepL. Our main goal was to estimate whether NMT had evolved substantially on these characteristics between 2024 and 2026, and whether the introduction of LLM-based MT constitutes a significant game-changer.

Our results show that the NMT generated by DeepL in 2026 exhibits the same level of lexical diversity as in 2024. The NMT tendency toward lexical overgeneralisation and the reduction of the translation solutions variation compared to HT remains constant and this for the two analysed corpora. The LLM-based MT exhibits a slightly higher level of word translation entropy, mainly due to occasional addition of translation variants, but without a significant reduction of the lexical overgeneralisation phenomenon.

On the syntactic level, our experiment demonstrates that the level of source sentence similarity has not changed substantially over the past two years, neither with the transition to an LLM-based architecture. While the more recent systems do not generate a translation identical to older outputs, the overall level of source sentence similarity remains very similar.

Those results show that the lexical and syntactic features of raw MT output generated by one of the most popular online MT engines have not evolved substantially over time, neither with the integration of LLM-based MT. This confirms that the MT tendency to overgeneralise on the lexical level and to follow the source syntax structure is stable over time and that results from previous studies remain highly relevant. Our work therefore shows that, on the stylistic levels, MT systems still struggle to produce human-like, diverse and natural sounding translations.

7 Limitations and Future Work

While this study highlights relevant concerns about some stylistic features of MT output, it remains within the scope of the case study. The analysed corpus, while presenting the advantage of being collected from authentic translation workflows, remains very specific and the analysed MT outputs originate from a single general-purpose online MT provider, without extensive and detailed information on model characteristics. In that sense, our results are not strictly generalisable and may reflect system-specific behaviour. Nonetheless, they provide a snapshot of the capabilities of a system widely used by professional translators and the general public. Furthermore, they invite us to consider MT progress with caution and to keep in mind that some limitations and drawbacks of MT systems might silently persist even if the technology seems to improve. Of course, it is now necessary to explore these aspects further by conducting diachronic studies on MT style on a larger scale and with a larger variety of systems. It is also essential to explore more widely translations done by LLMs and to measure the influence of different prompting strategies on the lexical and syntactic levels.

8 Acknowledgement

We would like to thank the language services who kindly accepted to share their translation memories for this research project.

	C1			C2		
	NMT24	**NMT26**	**LLM-MT26**	**NMT24**	**NMT26**	**LLM-MT26**
SACr	**0,19**	**0,19**	0,17$^{\diamond}$	**0,17**	**0,17***	0,16
Label POS	0,13	0,13	0,13	0,13	0,13	0,13$^{\diamond}$
Label DEP	0,19	0,19*	0,19	0,20	0,20*	0,20$^{\diamond}$
ASTrED	0,55	0,55*	**0,57**$^{\diamond}$	0,57	0,57$^{\diamond}$	0,57$^{\diamond}$

Table 6: Syntactic similarity scores for the different MT outputs (NMT24, NMT26 and LLM-MT26) for corpus C1 and C2. $^{\diamond}$ indicates a significant difference with NMT24 at $p < 0,001$, and * at $p < 0,05$. Significance was tested using Wilcoxon signed-rank test.

References

Ataman, Duygu, Alexandra Birch, Nizar Habash, Marcello Federico, Philipp Koehn, and Kyunghyun Cho. 2025. Machine Translation in the Era of Large Language Models:A Survey of Historical and Emerging Problems. *Information*. https://doi.org/10.3390/info16090723.

Bangalore, Srinivas, Bergljot Behrens, Michael Carl, Maheshwar Ghankot, Arndt Heilmann, Jean Nitzke, Moritz Schaeffer, and Annegret Sturm. 2016. Syntactic variance and priming effects in translation. In *New directions in empirical translation process research*, pages 211–238. Springer.

Carl, Michael, Moritz Schaeffer, and Srinivas Bangalore. 2016. The CRITT translation process research database. In *New directions in empirical translation process research*, pages 13–54. Springer. https://doi.org/10.1007/978-3-319-20358-4_2.

Dou, Zi-Yi and Graham Neubig. 2021. Word alignment by fine-tuning embeddings on parallel corpora. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2112–2128. ACL. https://doi.org/10.18653/v1/2021.eacl-main.181.

Hendy, Amr, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. 2023. How good are GPT models at machine translation? a comprehensive evaluation. *Preprint : arXiv:2302.09210*. https://doi.org/10.48550/arXiv.2302.09210.

Lazarov, Todor. 2026. LLMs and AI in the Language Industry: An Overview. In *2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, pages 1–6. https://doi.org/10.1109/ACDSA67686.2026.11468231.

Li, Yafu, Ronghao Zhang, Zhilin Wang, Huajian Zhang, Leyang Cui, Yongjing Yin, Tong Xiao, and Yue Zhang. 2025. Lost in Literalism: How Supervised Training Shapes Translationese in LLMs. In Che, Wanxiang, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12875–12894, Vienna, Austria, July. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.630.

Martikainen, Hanna and Natalie Kübler. 2016. Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions. *ILCEA*, (27), November. https://doi.org/10.4000/ilcea.3863.

Raunak, Vikas, Arul Menezes, Matt Post, and Hany Hassan Awadalla. 2023. Do GPTs Produce Less Literal Translations? pages 1041–1050, Toronto, Canada, July. https://doi.org/10.18653/v1/2023.acl-short.90.

Shaitarova, Anastassia, Anne Göhring, and Martin Volk. 2023. Machine vs. Human: Exploring Syntax and Lexicon in German Translations, with a Spotlight on Anglicisms. In *The 24th Nordic Conference on Computational Linguistics*. https://aclanthology.org/2023.nodalida-1.22/.

Tezcan, Arda, Joke Daems, and Lieve Macken. 2019. When a 'sport' is a person and other issues for NMT of novels. In *Machine Translation Summit XVII*, pages 40–49. European Association for Machine Translation. https://aclanthology.org/W19-7306/.

Toral, Antonio. 2019. Post-editese: an Exacerbated Translationese. In *Proceedings of MT Summit XVII*, volume 1, pages 273 – 281, Dublin, Ireland. https://aclanthology.org/W19-6627/.

Vanmassenhove, Eva, Dimitar Shterionov, and Andy Way. 2019. Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation. In *Proceedings of MT Summit XVII*, volume 1, pages 222 – 232, Dublin, Ireland. https://aclanthology.org/W19-6622/.

Vanmassenhove, Eva, Dimitar Shterionov, and Matthew Gwilliam. 2021. Machine Translationese: Effects of Algorithmic Bias on Linguistic Complexity in Machine Translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2203–2213, Online. ACL. https://aclanthology.org/2021.eacl-main.188.

Vanroy, Bram, Orphée De Clercq, Arda Tezcan, Joke Daems, and Lieve Macken. 2021. Metrics of syntactic equivalence to assess translation difficulty. In *Explorations in empirical translation process research*, pages 259–294. Springer. https://doi.org/10.1007/978-3-030-69777-8_10.

Volkart, Lise and Pierrette Bouillon. 2024. Post-editors as Gatekeepers of Lexical and Syntactic Diversity: Comparative Analysis of Human Translation and Post-editing in Professional Settings. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*, Sheffield. https://aclanthology.org/2024.eamt-1.33/.

Volkart, Lise. forthcoming. *Traduction automatique neuronale et post-édition : vers une uniformisation des textes traduits ? Étude sur la langue post-éditée en contexte professionnel*. Ph.D. thesis, University of Geneva.

Webster, Rebecca, Margot Fonteyne, Arda Tezcan, Lieve Macken, and Joke Daems. 2020. Gutenberg Goes Neural: Comparing Features of Dutch Human Translations with Raw Neural Machine Translation Outputs in a Corpus of English Literary Classics. *Informatics*, 7(3). https://doi.org/10.3390/informatics7030032.