A process-based perspective on the effects of machine translation on L2 writing

Margot Fonteyne[a], Joke Daems[b], Maribel Montero Perez[c] and Lieve Macken[d]

[a]Ghent University, Ghent, Belgium, iD, Margot.Fonteyne@UGent.be; [b]Ghent University, iD, Joke.Daems@UGent.be; [c]Ghent University, iD, Maribel.MonteroPerez@UGent.be and [d]Ghent University, iD, Lieve.Macken@UGent.be

Abstract

Most existing empirical work on the effects of Machine Translation (MT) use on second language (L2) writing has concentrated on its impact on writing products, with much less research addressing its effects on L2 learners' behaviours during writing. We therefore investigate whether the L2 writing process varies depending on whether learners are provided access to MT or to an online bilingual dictionary. Twenty-seven L1 Dutch learners of Swedish were assigned four writing tasks, with two tasks completed in each condition (MT or dictionary). While writing, participants' keystrokes were registered. Descriptive measures were used to summarize and compare participants' writing speed, tool engagement, and pausing behaviour across the two conditions. Results indicated that participants wrote more in less time, consulted the tool more frequently but for shorter periods of time, and paused longer between higher textual units when provided access to MT.

Keywords: L2 writing process, machine translation, online dictionaries, keystroke logging.

1. Introduction

The effects of MT use on L2 writing have gained increasing attention in recent years (Jolley & Maimone, 2022; Klimova et al., 2023; Lee, 2023). However, thus far, its effects have mostly been studied from a product perspective, overlooking the possible effects of using these tools on the processes learners engage in while writing a text in their L2. In this study, we adopt a process-based perspective to evaluate the impact of MT use on L2 writing. Our approach involved collecting, analysing, and comparing keystroke logs from 27 L2 writers, writing with access to MT or to an online bilingual dictionary. The L2 writing process was operationalised in terms of online measures of participants' writing speed, tool engagement, and pausing behaviour, which we extracted from the keystroke logs.

1.1. Background

L2 writers' linguistic knowledge is often limited and lacking automatisation (Kormos, 2012). Hence, the lower-order processes of vocabulary retrieval and grammatical encoding are commonly thought to impose greater cognitive demands on L2 writers, leaving them with less time and fewer resources to attend to higher-order processes, such as generating, organising, and evaluating the content of their texts. It has been suggested, however, that MT may increase L2 learners' lower-order processing speed. For example, Ahn and Chung (2020) found that respondents chose MT's time-saving capabilities most often as the main reason for using it during

writing. Similarly, in Clifford et al. (2013), respondents' third most common reason to use MT was that it "saves time" (p. 111). If MT indeed increases L2 learners' efficiency by accelerating their lexical retrieval and grammatical encoding processes, L2 writers should have more time and cognitive resources left to engage in higher-order processing, compared to when relying on more traditional tools such as online bilingual dictionaries.

Through triangulation of keystrokes with other data, studies have found associations between pauses and lower- and higher-order writing processes. Pauses between smaller textual units (within and between words) and larger textual units (between sentences and paragraphs) are typically linked to lower- and higher-order writing processes, respectively (Révész et al., 2019). In terms of duration, large thresholds such as 2000 ms are generally associated with higher-order processes, whereas smaller pause thresholds such as 200 ms are thought to also encompass lower-order processes (Van Waes & Leijten, 2015). Analysing L2 learners' pausing behaviour thus allows us to gather insights into how learners distribute their time and cognitive resources over higher- and lower-order processes during writing.

1.2 Research questions

Considering previous research on L2 learners' perceptions of MT and their writing behaviours, we formulated the following research questions:

1. Do L2 learners' writing speed and tool engagement vary depending on the tool they have access to while writing (i.e. MT or bilingual dictionary), indicating differences in the tools' efficiency?
2. Does L2 learners' pausing behaviour vary depending on the tool they have access to while writing, indicating differences in learners' distribution of time and resources over higher- and lower-order processes?

2. Method

2.1. Participants

A total of 27 learners of Swedish participated in this study. All participants exclusively identified Dutch as their first language. Out of the 27 participants, 21 learners were female, five were male, and one participant preferred not to disclose their gender. Participants were between 18 and 50 years old ($M = 28.74$, $SD = 11.47$, $m = 22$). Their L2 Swedish proficiency levels varied from A2 to B2 on the CEFR scale. At the time of data collection, 17 participants were enrolled in at least one (under)graduate course with Swedish as the main language of instruction at a Flemish university. The remaining ten participants were attending Swedish proficiency courses offered at various adult education centres in Flanders. All participants received monetary compensation (€12.50/hour) and were treated to a homemade Swedish pastry.

2.2. Materials

We developed four picture-based email writing tasks, all eliciting both descriptive and argumentative writing. Depending on the task, the participants were presented with images depicting three different events, holidays, travel destinations, or workshops. All images were connected to Sweden or its culture. For instance, in the 'event' task the images depicted an ice hockey cup, Sweden's national selection competition for Eurovision (*Melodifestivalen)*, and the Nobel Prize award ceremony in Stockholm. Participants were first asked to describe the three options presented to them using the images as references. Subsequently, they were instructed to provide arguments in favour of their chosen option and against the two other options. The emails had to be addressed to a (group of) friend(s) and written in Swedish. Participants were instructed to complete the four different writing tasks in Microsoft Word, with a maximum allotted time of 30 minutes for each task. Participants were told to aim for texts of 300 words, with descriptive and argumentative parts of equal lengths.

The study followed a within-subject design: participants were allowed to consult the MT tool *DeepL* for two of the tasks and were provided access to the online bilingual dictionary *Van Dale* for the other two tasks.

Participants were only allowed to translate between the language pairs Dutch-Swedish and Swedish-Dutch when using the tools. Usage of any other online or paper resources, including Word's built-in spelling and grammar checker, was prohibited.

2.3. Procedure

The procedure received ethical approval from the faculty. Each participant took part in two individual sessions, which lasted between 2.5 and 3 hours. Participants first provided informed consent and then engaged in practice tasks aimed at familiarizing them with the keyboard, software, and stimulated recall procedure. We assessed participants' L2 proficiency level and typing speed in Swedish using a proficiency test (Folkuniversitetet, n.d.) and a copy task (Van Waes et al., 2019), respectively. Additionally, participants completed a background questionnaire. Lastly, they were assigned two writing tasks. In the second session, participants first completed the remaining two writing tasks. Afterwards, participants took part in a stimulated recall session about their last writing task and were interviewed about their writing tool preferences and perceptions of the experiment.

To control for potential order, learning, and carry-over effects, we counterbalanced the order of the four writing tasks (events, holidays, destinations, or workshops) using a balanced Latin square design. In addition, we counterbalanced the order of the tools accessible to participants during the tasks (MT or dictionary). During each writing task, we registered participants' online writing behaviours using screen capture, eye tracking, and keystroke logging. For the keystroke logs we made use of Inputlog (Leijten & Van Waes, 2013).

2.4. Analysis

The 108 keystroke logs (four for each of the 27 participants) were analysed in terms of writing speed, tool engagement, and pausing behaviour. We obtained the writing speed measures by running a summary analysis on each log in Inputlog and extracting measures on participants' total process time (the time the participant needed to complete the task in minutes) and production rate (number of characters in the product, divided by the total process time). For the tool engagement statistics, we generated source analyses with Inputlog. Tool engagement was expressed with two measures: tool consultation frequency (number of times the participant opened the tool, divided by their total process time) and duration (mean duration of the times the participant had the tool open in seconds). Additionally, we used pause analyses generated by Inputlog to calculate pausing measures. Differences in pausing behaviour were studied in terms of total pause frequency (number of times the participant paused in Word, divided by the total time they spent in Word in minutes) and duration (median duration of their pauses in Word), and in terms of frequency and duration at different pause locations. To categorize the pauses by location, we employed Inputlog's classification scheme, distinguishing whether the pauses occurred within words, between words, or between sentences. We computed all frequency and duration measures twice, once with a 200 ms and once with a 2000 ms pause threshold. After extracting the writing speed, tool engagement, and pausing measures from each of the 108 keystroke logs, we grouped them by condition (MT or dictionary) and then calculated the mean values and standard deviations for each condition.

3. Results

For writing fluency, participants using the dictionary completed the tasks in an average of 28.63 minutes, at an average rate of 56.05 characters per minute (Table 1). The average participant in the MT condition exhibited a slightly shorter process time of 28.00 minutes and a slightly higher production rate of 58.59 characters per minute. Notably, the high standard deviations indicate considerable variability in writing speed across participants.

Regarding tool engagement statistics, participants consulted the dictionary approximately 0.99 times per minute, while for MT the frequency was slightly higher (1.06). The mean consultation length for the dictionary was 13.36 seconds, whereas this was notably shorter for MT (10.44s).

Table 1. Mean values and standard deviations (in parentheses) of the writing speed and tool engagement measures per condition.

Condition	Bilingual dictionary	MT tool
Total process time (in min)	28.63 (3.27)	28.00 (3.52)
Number of characters (per min)	56.05 (17.83)	58.59 (16.84)
Consultation frequency (per min)	0.99 (0.38)	1.06 (0.48)
Mean consultation duration (in s)	13.36 (3.96)	10.44 (3.70)

The results of the pause analyses for the two conditions are presented in Table 2. Mean values are given for all pauses and for pauses at three different pause locations, with two different pause thresholds. For pausing frequency, differences between conditions are small. In terms of pause duration, we do find that participants pause noticeably longer between sentences in the MT condition, especially when the larger pause threshold of 2000 ms is applied to the data. With MT, their median inter-sentence pause duration is 3.96 seconds on average, whereas they only pause between sentences for an average of 3.37 seconds with the dictionary.

Table 2. Mean values of the pausing measures per condition (BD: bilingual dictionary. MT: MT tool) by pause threshold (PT) in total and by location.

		Total		By pause location					
	PT			Within words		Between words		Between sentences	
Condition		BD	MT	BD	MT	BD	MT	BD	MT
Pause frequency (per min)	200	56.47	55.36	21.87	20.97	14.73	14.57	1.15	1.13
	2000	3.94	3.90	0.28	0.26	1.74	1.62	0.26	0.25
Median pause duration (in s)	200	0.41	0.41	0.29	0.28	0.68	0.65	0.98	1.13
	2000	3.30	3.31	2.63	2.66	3.21	3.27	3.37	3.96

4. Discussion and conclusions

In this paper, we presented a comparative analysis of the impact of two tools, an MT tool and an online bilingual dictionary, on the L2 writing process. We focused on differences in writing speed, tool engagement, and pausing. This section discusses the study's findings in relation to our research questions. However, it should be noted that these findings are only preliminary, as they are solely based on descriptive statistics.

Regarding our first research question, which addressed whether L2 learners' tool engagement and writing speed vary based on the tool they use, we observed that participants consulted the MT tool slightly more often. They also spent less time in it during a consultation. Furthermore, participants produced their texts at a slightly faster pace. These findings suggest that using MT may lead to a more efficient writing process. However, we also found high variability in participants' writing fluency. Considering previous research that has identified L2

proficiency as a strong predictor of fluency (Révész et al., 2022), the wide range of participants' L2 Swedish proficiency levels may explain this observation.

For our second research question, which investigated whether L2 learners' pausing behaviour differs depending on the available tool, we discovered that, when applying a pause threshold of 2000 ms, participants paused noticeably longer at higher textual units when using MT. Given that longer pauses between larger units are associated with higher-order processes, this finding suggests that participants focused more on higher-order concerns while writing with MT. However, this result needs to be validated through triangulation with other data sources.

Our objective is to address the limitations of this preliminary analysis in future research. First, we plan to construct linear mixed effects models for each of the measures presented. These models will allow us to make more robust generalizations about the impact of tool type on the L2 writing process and provide insights into potential moderating factors, including L2 proficiency level. Second, we will analyse the eye-tracking and stimulated recall data to explore differences in learners' reading behaviour during writing and the cognitive processes underlying learners' online writing behaviours. These analyses will yield additional insights into how learners allocate their time and cognitive resources to the various higher- and lower-order writing subprocesses. Finally, by annotating the eye-gaze, revision, and tool engagement behaviour in the screen recordings, we aim to uncover more fine-grained differences in L2 writers' patterns and the strategies they employ when using the two different tools.

Acknowledgements

References

Ahn, S., & Chung, E. S. (2020). Students' perceptions of the use of online machine translation in L2 writing. *Multimedia-Assisted Language Learning*, *23*(2), 10–35. https://doi.org/10.15702/mall.2020.23.2.10

Clifford, J., Merschel, L., & Munné, J. (2013). Surveying the landscape: What is the role of machine translation in language learning? *@tic. Revista d'innovació Educativa*, *10*, 108–121. https://doi.org/10.7203/attic.10.2228

Folkuniversitetet. (n.d.). *SVENSKA – Inplaceringstest B*. https://uiss.org/wp content/uploads/2017/08/INPLACERINGSTESTB.pdf

Jolley, J. R., & Maimone, L. (2022). Thirty years of machine translation in language teaching and learning: A review of the literature. *L2 Journal*, *14*(1), 26–44. https://doi.org/10.5070/L214151760

Klimova, B., Pikhart, M., Benites, A. D., Lehr, C., & Sanchez-Stockhammer, C. (2023). Neural machine translation in foreign language teaching and learning: A systematic review. *Education and Information Technologies*, *28*(1), 663–682. https://doi.org/10.1007/s10639-022-11194-2

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390–403. https://doi.org/10.1016/j.jslw.2012.09.003

Lee, S.-M. (2023). The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *36*(1–2), 103–125. https://doi.org/10.1080/09588221.2021.1901745

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition*, *41*(3), 605–631. https://doi.org/10.1017/S027226311900024X

Révész, A., Michel, M., Lu, X., Kourtali, N., Lee, M., & Borges, L. (2022). The relationship of proficiency to speed fluency, pausing, and eye-gaze behaviours in L2 writing. *Journal of Second Language Writing*, *58*, 100927. https://doi.org/10.1016/j.jslw.2022.100927

Van Waes, L., & Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to L1 and L2. *Computers and Composition*, *38*(Part A), 79–95. https://doi.org/10.1016/j.compcom.2015.09.012

Van Waes, L., Leijten, M., Pauwaert, T., & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with Inputlog. *Journal of Open Research Software*, *7*(1), Article 1. https://doi.org/10.5334/jors.234