

Reflective practice

Google Translate in Academic Writing Courses?

Sara Kol*, Miriam Schcolnik** and Elana Spector-Cohen***
Tel Aviv University, Israel

*sarakol @ post.tau.ac.il | **smiriam @ post.tau.ac.il | ***espector @ tauex.tau.ac.il

Abstract

The aim of this study was to explore the possible benefits of using *Google Translate* (GT) at various tertiary English for Academic Purposes (EAP) course levels, i.e., to see if the use of GT affects the quantity and quality of student writing. The study comprised preliminary work and a case study. The former included an awareness task to assess student awareness of GT mistakes, and a correction task to assess their ability to correct the mistakes identified. The awareness and correction tasks showed that intermediate students identified 54% of the mistakes, while advanced students identified 73% and corrected 87% of the mistakes identified. The case study included two writing tasks, one with GT and one without. Results showed that when using GT students wrote significantly more words. They wrote longer sentences with longer words and the vocabulary profile of their writing improved. We believe that GT can be a useful tool for tertiary EAP students provided they are able to critically assess and correct the output.

Keywords: English for Academic Purposes, academic writing, *Google Translate*.

1. Introduction

Most tertiary institutions around the world teach writing in English, both at the undergraduate and graduate levels, so students can function in academia and later in their chosen professions. Students of English for Academic Purposes (EAP) need to deal with lexical, morphological and syntactic language difficulties as well as learn the structure and conventions of academic writing.

1.1. Machine Translation, Past and Present

In this day and age, digital tools, such as online dictionaries, spelling and grammar checkers and search engines are ubiquitous and can aid the process of writing. Automatic translation or machine translation (MT) is also a digital tool. "Machine translation is the process of translating text from one language into another using a computer." (*Cambridge English Dictionary* online). Even though it was not originally developed for educational purposes, MT has been adopted by many students writing in foreign languages.

In 1951, following a few early attempts to produce automatic translation, Yehoshua Bar-Hillel from MIT wrote a research report on MT, and in the following year organized the first MT conference. In 1954, IBM demonstrated the first MT system using a sample of 49 Russian sentences translated into English. The system was limited to 250 words and 6 grammar rules. The demonstration stimulated great interest worldwide (Hutchins, 2007).

In the following decades, research on MT continued, hardware was improved, and a variety of approaches were applied. One such approach was the rule-based approach, which bases the translation on a series of linguistic rules including morphology and syntax; another approach was the corpus-based approach, which bases the translation

on texts taken from databanks (Hutchins, 2007); still another was the phrase-based approach, which breaks down sentences into words and phrases to be independently translated (Hsu, 2016).

In November 2016, Google switched from the phrase-based approach to Neural Machine Translation, an approach that uses Artificial Intelligence (AI) to learn from millions of examples. Artificial Intelligence has greatly improved the quality of GT (Schuster, Johnson & Thorat, 2016), which now goes beyond sentence by sentence translation and takes the whole text into account (*Innovations Report*, 2017). This has reduced the number of errors by at least 60% in comparison with the phrase-based approach (Hsu, 2016).

1.2. Machine Translation in Language Classes

The literature on the use of machine translation in language learning and especially in academic writing is very limited. In a survey conducted by Niño (2009), 75% of the students felt that MT was a helpful language tool and 81% expressed that MT had contributed to their language improvement. In her survey of faculty attitudes, Niño found that 23% of the language instructors used MT in their lessons both from L1 into L2, and from L2 into L1, from intermediate level onwards. Thirty percent of the instructors who did not use MT in their lessons said they would be willing to use it.

In a study conducted in Australia, Garcia & Pena (2011) investigated whether MT could help beginner learners of Spanish to communicate better in writing. They found that MT helped students write more (quantity) and better (quality).

In a survey of language instructors in the foreign language department at a regional Swedish university conducted in 2012, 66% of the respondents said they would prefer if their students did not use MT when doing written assignments. However, all teachers agreed that if students did use MT, they would need good language skills to edit the output.

In 2012 a survey focusing on the use and perceptions of MT was conducted among students of Spanish, French, Italian and Portuguese (N = 905) at Duke University. Ninety-one percent of the students said they noticed mistakes made by MT; 43% said they used it for double-checking what they wrote in the foreign language; 85% felt MT helped increase their vocabulary (Clifford, Merschel, Munne, & Reisinger, 2013). In short, students felt that MT helped them learn language.

In 2013, the Spanish Language Program at Duke University had a written policy forbidding the use of any computer software that compromises the students' learning process, including translation programs (Clifford, Merschel, & Munné, 2013). The Duke University research team sent an email survey of faculty attitudes to several universities and received 43 responses. Seventy-seven percent of the respondents disapproved of the use of machine translation by students, and none of them approved. Eighty-four percent of the instructors teaching beginners felt that MT was not a useful tool. However, 54% of the instructors teaching advanced levels felt MT was useful. Forty-two percent felt that using MT in writing assignments was cheating. In sum, both the Duke University and the Swedish university faculty surveys showed that the majority of participants were against the use of MT by students.

In 2014, Spector-Cohen, Schcolnik, & Kol reported on a survey conducted among tertiary level students (N = 203) to find out whether EAP students use GT when writing in English, their motivation for using it, and their attitudes towards GT. The results of the survey showed that 80% of the students used GT *always*, *often* or *sometimes*. Eighty-two percent reported using it to translate single words, whereas only 28% said they used it to translate whole paragraphs.

1.3. Purpose and Structure of the Study

According to Groves & Mundt (2015), MT can have a profound effect on language teaching. MT programs have been available for a while, but there have been few studies of their use by foreign language students (Spector-Cohen et al., 2014, Clifford, Merschel, & Munné, 2013; Garcia, & Pena, 2011, Niño, 2009).

Google Translate (GT) is the most commonly used automatic translation tool, which is why we decided to focus our study on GT rather than other MT tools. Most students in our tertiary EAP classes use it when writing. Even so, among language instructors there seems to be consensus that GT has no place in EAP courses. Many instructors in fact forbid their students to use GT because “*You don’t teach writing by having a machine write for you.*” Or “*Machine translation is so inaccurate, how can it help students write?*” These comments are consistent with the lack of willingness of tertiary institution instructors to use technology in their courses.

In a study of information technology use in tertiary institutions in the U.S. (Brooks & Pomerantz, 2017) in which 43,559 students from 124 institutions in 10 countries and 40 U.S. states participated, the students reported that faculty are banning or discouraging the use of laptops, tablets, and smartphones in the classroom more often than in previous years.

The main issue that motivated our study was whether the use of GT should be allowed in EAP writing programs. Over the years, we have encouraged and taught mindful use of various digital tools in our EAP courses. Since we see GT as a potentially useful tool, we decided to explore the possible benefits of using this tool at various course levels. We assumed GT could be beneficial to EAP writers, provided the GT output was assessed critically and corrected, i.e., students would need a sufficient level of language knowledge to notice the mistakes made by the machine and to correct them. The main purpose of the study was to develop guidelines for use of GT in EAP writing courses, favoring mindful use over simple copy-paste.

Our study comprised preliminary work and a case study. The purpose of the preliminary work was to determine the course level to focus on in the case study. It included an *awareness task* to assess student awareness of the mistakes made by GT, and a *correction task* to assess their ability to correct the mistakes identified. The purpose of the case study was to determine if the use of GT affected quality and quantity of advanced student writing, including language, number of words, readability and vocabulary level. The case study included two writing tasks, one without the use of GT and one where GT was allowed.

2. Preliminary work

2.1. Method

Population: The awareness task was run in B1 and B2 classes (Common European Framework levels, N = 79) at Tel Aviv University and the Interdisciplinary Center Herzlyia (IDC). The correction task was then run in B2 classes only (N = 49) at IDC.

Instruments:

- An *awareness task* to assess student awareness of mistakes in the translation of 10 sentences translated from Hebrew to English by GT.
- A *correction task* to assess students’ ability to correct the mistakes they identified.

Data Analysis: To measure mistake awareness and ability to correct mistakes, we devised formulas for calculating scores as follows:

- *Mistake Awareness Score (MAS)* is the ratio of the number of mistakes students identified, out of the total number of mistakes made by GT (minus a penalty for marking correct items as mistakes).
- *Mistake Correction Score (MCS)* is the ratio of the number of mistakes students corrected, out of the number of mistakes they identified (minus a penalty for incorrectly changing correct items). Penalties were given only when the alternative provided in the change was incorrect. Where the correction was unnecessary but possible, no penalty was given.

Procedure: The awareness task presented students with 10 short sentences in Hebrew followed by their GT translation. Students were asked to identify the mistakes in the translation. In light of the findings, that showed B1 students could only identify about half of the mistakes, we decided to run the correction task in B2 classes only, where students were asked to correct the mistakes found.

2.2. Results

Results for the awareness task showed that B1 students were able to identify 54% of the mistakes, while B2 students identified 73%.

Results for the correction task showed B2 students were able to correct 87% of the mistakes they identified.

3. Case study

3.1. Rationale

Since we assumed that awareness of mistakes and the ability to correct them are a prerequisite for productive use of GT for writing, and in light of the results of the preliminary work, we decided to run the case study with B2 students only. We wanted to see if the use of GT affected the quantity and quality of their writing, including the readability level and vocabulary profile of their texts, and their use of academic vocabulary.

3.2. Method

Population: One class (N = 25; 14 male and 11 female) of 1st year students of government and political science (at B2 level) at IDC.

Instruments:

- Two *writing tasks*, one without using GT and one where use of GT was allowed. Both tasks required writing a few paragraphs taking a stand on a topic dealt with in texts read in the course.
- A *short questionnaire* about student use of GT when doing the second writing task: whether they used GT at all, used it to translate words and phrases, full sentences, or whole paragraphs (See Appendix).

Procedure: Early in the semester, students did the two writing tasks, administered a week apart. They were given a choice whether to write by hand or on the computer, to avoid the effect of the medium on the quality of the writing. They were given 40 minutes for each *writing task*.

In the first *writing task* students were not allowed to use GT but were permitted to use a handheld dictionary, either print or electronic. In the second task they were allowed to use GT. After completion, students answered questions about their use of GT.

Data Analysis: The evaluation of the two writing tasks included scores for the quantity and quality of writing, the readability and the level of vocabulary.

- *Writing Score* – To score the writing tasks, we considered language mistakes only, including morphology (e.g., verb form) and syntax (e.g., missing article). Content and paragraph structure were ignored, as were spelling mistakes. If the same type of mistake was repeated, it was counted only once. We calculated the score by using a percentage based on the ratio of the number of mistakes students made, out of the total number of words they wrote.
- *Quantity of Writing* – We compared the number of words in the two writing tasks.
- *Readability Score* – Readability is an index that reflects text complexity and can be expressed as the grade level for which the text is suitable. It is measured by calculating the length of sentences and the number of words, syllables, and characters in the text. For example, when longer words (with more syllables) are used, the readability score increases. We used the *Flesch-Kincaid* scale to measure readability.
- *Vocabulary Profile* – Vocabulary profiling is based on lexical text analysis. It reflects the percentage of low and high frequency vocabulary used in a written text and classifies the words as belonging to different categories (Laufer & Nation, 1995). According to Laufer & Nation, lexical profiling has several advantages over other measures of lexical richness, as it is independent of syntax and focuses on lexis. The first category (K1) includes the 1000 most frequent words in English (e.g., *world, years, peace, and*),

and the second (K2) includes the next 1000 commonly used words (e.g., *ahead, threat, solve*). The third category is the Academic Word List (AWL), which includes 570 words used frequently in academic texts across subjects (e.g., *crossroad, intractable, generation*). We used *Lextutor* (<https://www.lexutor.ca/vp/eng/>) to profile student texts.

- *Questionnaire* - We recorded student use of GT in the second writing task.

3.3. Results

- *Writing Score* - The average grade (based on grammatical accuracy only) for writing task 1 (without GT) was 93.2 and for writing task 2 (with GT) the average grade was 93.4. There is no significant difference between the two grades.
- *Quantity of Writing* - The average number of words written in task 1 was 140 and in task 2, 171. A t-test was conducted to compare the average number of words in the two writing tasks. There was a significant difference between the two tasks, $t = -2.04947$. $p = .023348$. The result is significant at $p < .05$, showing that students wrote significantly more words when using GT.
- *Readability Score* - The readability grade level rose when students used GT. The average grade level for the first writing task (without GT) was 8.6 and that for the second task (with GT) was 10.3. A t-test was conducted to compare the two averages. There was a significant difference between the two grade levels, $t = -2.71851$. $p = .004745$. The result is significant at $p < .05$, showing that students wrote longer sentences with longer words when using GT.
- *Vocabulary Profile* - Words in the texts students wrote were classified into three vocabulary groups: K1 = the first 1000 most commonly used words in English; K2 = the next 1000 most commonly used words; AWL = academic word list.

Table 1. Vocabulary profiling average percentages.

	K1	K2	AWL
Task 1 (without GT)	89.1	2.9	4.6
Task 2 (with GT)	79.0	5.0	5.9

The numbers represent the percentage of words in each group out of the total number of words in the text.

When using GT, the vocabulary profile of student writing improved, i.e., K1 words decreased, while K2 and AWL words increased. T-tests were conducted to compare averages. There was a significant difference between the first and second writing tasks for all three word groups. For K1, $t = 8.60713$. $p < .00001$. The result is significant at $p < .05$, showing that students used fewer basic words when writing with GT. For K2, $t = -4.78664$. $p < .00001$. The result is significant at $p < .05$, showing that students used more words from the second 1000 word group when writing with GT. For AWL, $t = -2.18186$. $p = .017133$. The result is significant at $p < .05$, showing that students used more academic words when writing with GT.

To sum up, there was a significant difference in the lexical items used by students when using GT, the number of K2 and AWL words grew significantly and the K1 words decreased.

- *Questionnaire*

Eighty-three percent of the students reported they used GT while writing. Seventy-five percent of those who used it did it to look up words and phrases. Very few students said they used it to translate full sentences or paragraphs.

Students who said they didn't use GT explained why: *I didn't need it to write the paper; I don't want to get used to relying on software, I prefer practicing my English; I want to improve my English with the vocabulary I already have in my possession, my own tools.*

4. Discussion and conclusion

Preliminary Work – The results of the awareness and correction tasks showed that B2 students are well aware of most of the mistakes made by GT and are capable of correcting them. This indicates a level of English that is sufficient for critical assessment of the GT output and revision. The finding also validates our decision to run the case study with advanced students.

Case Study – No significant difference was found in the writing task scores. We believe this can be explained by the fact that the only types of mistakes counted were grammatical mistakes. Since students used GT mainly as a dictionary, to look up words and phrases, GT could not affect their grammatical accuracy. Moreover, the writing task with GT was done one week after the task without. We can assume that one week is not enough time for instruction to affect the scores and evidence an improvement in grammatical accuracy.

Students wrote significantly more words when using GT. One explanation could be that they had more time to write, not needing to think about how to express their thoughts in English, as they could just look up words and expressions quickly. However, we cannot discount the possibility that having done one task, students gained more self-confidence to write more extensively in the second task.

The readability level of the students' writing in the second task was significantly higher. This can be explained by the fact that students used a higher level of vocabulary in the second task, employing longer words, as evidenced in the vocabulary profiling done.

In academic writing, the ability to choose formal content words is needed (Swales & Feak, 2009). The vocabulary profiling revealed that students used a higher level of vocabulary when using GT. This is not evidence that the words have already become part of their productive written vocabulary, but it does show that they were exposed to them. The course in which the case study was conducted deals with topics in the area of government, where the specialized, academic vocabulary suggested by GT will probably reappear. Since exposure is the first stage in vocabulary learning, it is to be expected that the students will eventually be able to incorporate those words into their active vocabulary.

According to Nation (2001), the quality of academic writing depends strongly on the use of academic vocabulary. He claims that both knowledge and motivation affect the use of vocabulary. Our EAP students are motivated to use academic vocabulary but in many cases do not know suitable words in English, although they may well know them in their L1. We think that using GT can help them start to use those words and with enough re-entry, incorporation of the words into the students' productive vocabulary should occur.

The findings of the case study, though interesting, cannot be generalized due to the small number of participants. That is a limitation of the case study. In light of the constant improvement of artificial intelligence and thus of GT, it would be interesting to conduct similar studies with larger samples of advanced students of English in the future.

5. Pedagogical suggestions

Even though it may be difficult for EAP instructors to introduce machine translation into their courses, as it could be perceived to undermine teaching and learning a foreign language (Groves & Mundt 2015), we believe that GT can be a useful tool for language students at all levels, particularly because it is continuously improving its translations. Since most students are already using the tool, instructors should show their students how to use it effectively rather than ignoring it or forbidding its use. As pointed out previously, machine translation in its current state is only able to produce texts of limited quality. Thus, students will need to be able to check it for accuracy, cohesion and quality of translation. This process could be exploited in the classroom to enhance teaching and learning (Groves & Mundt 2015).

Based on the findings of the study and our experience teaching the use of digital tools in tertiary EAP courses, we propose the following guidelines for instructors.

Lower levels (C and B1 levels in the CEF)

- Even though lower level students might be tempted to enter whole paragraphs in their native language, they need to be made aware of the fact that the output will require editing, which they are probably not equipped to do. Therefore, we recommend that lower level students use GT only for words (as a dictionary) and short phrases.
- Instructors should show their students that when they input a word in their native language, they need not use the first word that comes up in the translation box. Instead, students need to choose the best translation for their specific context. GT facilitates the choice by providing a list of alternative translations with their meanings in the language entered. Figure 1 is an example of the list provided by GT when entering the Hebrew word להפיק (/lehafeek/). The main translation given in the box is *produce*, which is the most common use of the word.

Translations of להפיק

verb	
■ produce	להפיק, ליצר, להציג, להצמיח, לשאת פרי, לעשות
■ bring out	להוציא לאור, להבליט, להפיק
■ derive	להפיק, לנבוע, להגזר, לשאוב
■ issue	להנפיק, להוציא, לפרסם, ליצא, להפיק, לנבוע
■ elicit	להפיק, להוציא, לדובב, להתרים
■ extract	להוציא, להפיק, למצות, לסלז, להעתיק, לעקור שן
■ output	להפיק, ליצר
■ send forth	להפיק
■ profit	להרוויח, להפיק, להשתכר

Figure 1. Example of a list of alternative translations in *Google Translate*.

Given the richness of the translations provided, even lower level students should be able to choose an appropriate translation. This is true when entering single lexical items. When entering an idiomatic expression, GT may provide both the literal and the idiomatic translations, the most common one appearing in the box, and the other one in a popup that appears when clicking the translation.

Advanced levels (B2 and A levels in the CEF)

- Advanced students should be able to use GT for short sentences as well as words and phrases. Instructors should show their students that GT translates many sentences correctly, but that they need to be aware of potential mistakes. Instructors can show them that in some cases alternative translations of the sentence may be provided in a popup, and the alternative may be more suitable than the main translation.
- Instructors may choose to conduct an awareness and correction task in class, in which students, working either individually or in groups on the same translation output, go over and edit it. Then, student editing can be shown and discussed. This activity can constitute preparation for the use of GT in writing. It can raise awareness of the kinds of mistakes they are likely to encounter and how to correct them.

References

- Brooks, D.C. & Pomerantz, J. (2017, October). *ECAR Study of Undergraduate Students and Information Technology, 2017*. Research report. Louisville, CO: ECAR.
- Cambridge English Dictionary* (2018). "Machine Translation". Cambridge University Press. Retrieved from <https://dictionary.cambridge.org/dictionary/english/machine-translation>.

Clifford, J., Merschel, J., Munne, D. & Reisinger, D. (2013, April). The Elephant in the Room: Machine Translation in Language Learning at Duke University. Retrieved from <http://cit.duke.edu/wp-content/uploads/2013/04/Elephant-in-the-Room1.pdf>

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.

Garcia, I & Pena, M.I. (2011). Machine translation-assisted language learning: writing for beginners. *Computer Assisted Language Learning*, 24(5), 471-487.

Groves, M. & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, 37, 112-121.

Hsu, J. (3 Oct 2016). Google Translate Gets a Deep-Learning Upgrade. *IEEE Spectrum: Technology, Engineering, and Science News*. Retrieved from <https://spectrum.ieee.org/tech-talk/computing/software/google-translate-gets-a-deep-learning-upgrade>

Hutchins, J. (2007). Machine Translation: a concise history. Retrieved from <http://www.hutchinsweb.me.uk/CUHK-2006.pdf>

Laufer, B. & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307-322. Oxford University Press.

Innovations Report (2017). "Machine translation: going beyond sentence by sentence". Retrieved from <http://www.innovations-report.com/html/reports/information-technology/machine-translation-going-beyond-sentence-by-sentence.html>

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *CALL*, 21(2), 241-258.

Schuster, M., Johnson, M. & Thorat, N. (November 22, 2016). Zero-Shot Translation with Google's Multilingual Neural Machine Translation System. Retrieved from <https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html>

Spector-Cohen, E., Schcolnik, M. & Kol, S. *Google Translate in Foreign Language Programs?*, presented at the 13 th Language and Society Conference: Languages and Discourse in the Public Space, Levinsky Education College, Tel Aviv, 2014.

Swales, J.M. & Feak, C.B. (2009). *Academic Writing for Graduate Students*. Michigan: University Michigan Press.

Wikipedia (2018). "History of machine translation". Retrieved from https://en.wikipedia.org/wiki/History_of_machine_translation

Appendix

Questionnaire after second writing task in case study:

1. Did you use GT? YES - NO
2. If your answer is YES, what did you use it for? *Check ALL correct answers.*
 - To translate single words
 - To translate groups of words
 - To translate whole sentences
 - To translate whole paragraphs
3. If you didn't use GT, why not?