

MACHINE TRANSLATION AS AN AID FOR WRITING BY COMPUTER SCIENCE UNIVERSITY STUDENTS

J.M. Arriola

University of the Basque Country UPV/EHU (SPAIN)

Abstract

This paper describes a practice in which students have integrated the use of machine translation tools (MT) into academic writing tasks, in line with recent works [1], [2], [3]. More specifically, the use of Elia [4], the automatic neuronal translator of Elhuyar. Fourth-year students of the Bachelor's degree in Computer Engineering used MT tools for the following tasks: writing articles for Wikipedia and performing localization projects from English to Basque. We complemented the use of MT in the context of Basque for specific purposes with the exploitation of existing lexical and terminological resources. The students will apply different strategies according to the writing task. We demand them to use all these tools critically and encourage practices such as post-editing text outputs to obtain better results. Most students were not aware of the limitations of technology in the translation of specialized texts. We provide strategies to improve translations and train them for a better knowledge of the language that will be beneficial to get the most from MT tools. We are committed to integrating MT into teaching, although some predictions believe that MT will replace the need for language learning and teaching [5].

Keywords: MT and language teaching and learning, MT-supported activities.

1 INTRODUCTION

Increasingly, education trends advocate learning based on the integration of digital and technological resources. Many scholars have highlighted that the 21st century has witnessed much integration of MT into education, especially in language learning [2]. In the same direction, concerning foreign language teaching and learning, nowadays nobody working in the field of teaching and learning Foreign Language (FL) s can imagine life without machine translation tools [1]. Although the use is unavoidable, the didactic and methodological implications require further study. For instance, in the university context of language teaching, the use of machine translation tools online as a methodological resource in the classroom is not yet fully established [6].

In this research, we explore the integration process in the use of MT tools in the teaching of the following subjects taught by the Basque Language and Communication Department: Norm and Use of Basque Language and Applied Linguistics. Due to the current sociolinguistic particularities of Basque standardization and normalization, issues are still so important for Basque. These subjects cover the four macro skills (speaking, listening, reading and writing), but especially emphasize that the student manages to handle the language in scientific and technical texts in their area of study, focusing the contents according to their interests. The important thing is to help the student's learning experience.

In this context, we reaffirm ourselves in one of the competences of our educational program: Management of consultation tools (especially available on the Internet) to adequately respond to communication needs that may arise in diverse situations.

Taking into account the big impact of Elia, the automatic neuronal translator of Elhuyar, we would argue that Basque is certainly affected by the new language technologies that have been developed for Basque. In a more general scenario, there are also predictions that MT will replace the need for language learning and teaching [5]. The issue has many implications (linguistic, sociolinguistic, pedagogical,...), from the point of view of language teaching, in our case language for specific purposes, it is essential to undertake the integration of translation tools both in teaching and in learning practices. The aim of this work is to contribute to the integration of MT systems in academic writing classes.

The use of MT in the context of Basque for specific purposes was complemented by the exploitation of existing lexical and terminological resources. The students should take into account the different strategies depending on the writing task. We demand them to use all these tools critically and encourage practices such as post-editing text outputs to obtain better results.

Most students were aware of the limitation of technology and reached out to several strategies to test and improve translations quickly, although with a significant lack of adequacy and attention to the needs of specialized texts.

This paper is structured as follows: in Section 2, we present the methodological approach carried out with a group of computer science engineering degree students, in Section 3 we describe some issues observed in the application of the proposed methodology, in Section 4 we conclude and outline the future work.

2 METHODOLOGY

Through an initial questionnaire about which resources and tools use the students in tasks related to technical language in the field of computer science, we observed that most of them use automatic translator tools whenever they have to do a job. Elia is the most used and as complementary source the Elhuyar dictionary to solve linguistic doubts. Therefore, we decided to dedicate part of our classes to teaching them how to use these tools and the necessary strategies to correct and produce texts that are useful in the process of communication. Computer students are familiar with the use of technology, but it is not just cutting and pasting, translating and re-cut and pasting. The use of the MT tools also requires the necessary strategies to correct automatic errors and edit a product that, touched by the human mind, is useful in the process of communication. There are extensive works about how and when to use MT for language learners and about the understanding of typical MT errors and how to spot them [7].

Our methodological approach is relatively rudimentary but is in line with the goal of integrating MT tools into the academic writing process. We gave them an introductory lesson on the main phases of the text elaboration process: contextualization, structuring and textualization. In turn, they have emphasized the importance of the text genre in the structuring phase of the text.

The methodology was divided into three main phases: 1) Discussion forum, 2) Planning, and 3) Review.

2.1 Discussion forum

As an initial step, for us is fundamental to focus on the communicative context of each task, so that we organize a forum about the communication situation to be discussed. Once, every student has already expressed in the forum what they think to work it out, they should follow a procedure. Regarding Wikipedia article writing, to address adequately the wording of this article and the possible result, it is necessary to develop the planning phase.

2.2 Planning

We asked the students to complete a form containing the steps to be taken. Such planning will consist of the following main steps:

- a) Article starting point: translation, mixed or completely original. If an existing article on Wikipedia is selected as the starting point, the pdf of the article will also be uploaded to the virtual classroom.
- b) Language of origin of the translation.
- c) Indicate the languages to be consulted for cases of linguistic ambiguity:
- d) A sample of the text to translate. The students are going to pick up a significant part of what they need to do. Therefore, they should list in a table the corresponding version for each fragment translated with the MT tool and their own version. They must compare both versions and explain what improvements they have made and the reason for them.
- e) Now, students can do their own first version based on the initial output of the MT system.

2.3 Review

The review comprises two tasks:

First, task: comparing the scheme and the initial work (first writing). Assess whether it has acted correctly, especially as regards the distribution and cohesion of ideas, as well as the consistency of arguments. Is there any contradiction in your work? Is there unnecessary information or content ("noise")? Is the succession of ideas well organized? If necessary, reorder or rewrite the item. Is there any information that is unclear or that the reader does not understand well? If so, try to define and adapt it.

Second task: complete the second scripture: review it and correct it. In this case, it shall pay particular attention to two issues: 1. Analyze and adapt the text (syntax, connectors...) 2. Correction of grammatical errors and finally, use the Spelling Checker/Corrector for Basque XUXEN¹.

The aim of the proposed methodology is to take the profit of applying machine translation systems to support translation tasks with the purpose of giving the students support to get them closer to the target text that they will have to polish.

3 SOME ISSUES

We draw our data from a sample of 32 Basque students of Applied Linguistics as an optional subject while writing articles for Wikipedia and performing localization projects from English to Basque. In this section we briefly describe some issues observed in the application of the proposed methodology and the complementary tools they we have used in our two tasks; we also present some conclusions expressed by students resulting from the practices.

3.1 Some practices observed in these tasks

The students have adopted the methodology in so different ways.

3.1.1 *Elia as dictionary*

Regarding, the localization task many students used the automatic translator Elia to perform vocabulary queries. In these cases, they have used localization platforms such as Crowdin, Transifex, Weblate... in order to obtain results that are more reliable. However, they have mentioned two important problems: 1) Lack of information about the translation context and 2) Absence of translations for the Basque language.

3.1.2 *Elia as the only tool*

Some students were surprised by Elia's breakthrough and consider that they do not need more resources. These students used it exclusively for translation. They have used it, above all, to translate complex phrases into Spanish, because it offers very good results. In these cases, the differences between the machine translation and the translation proposed by the student are negligible. They are therefore required to pick up both versions and indicate their contributions. Otherwise, the qualification will be for the tool and not for the student.

3.1.3 *Elia as central tool, but complemented with other resources*

In this group, the students noted that the translations provided by the MT tool should be reviewed later as it is not perfect in no way. All translations made in this work have followed a methodology to achieve the highest quality in translations. When doing the translation, they analyze the context to make sure they understand it perfectly. Having understood the context and the translation itself, they compared their translation with Elia's translation.

The translation/revision task was performed using other linguistic resources:

- The spelling checker Xuxen (<http://xuxen.eus/eu/home>);
- Terminological databanks (Euskalterm, <http://www.euskadi.eus/euskalterm/>);
- Specialized corpora (Garaterm (http://garaterm.ehu.es/garaterm_ataria/en/kontsultak/),
- And some specialized dictionaries (Zientzia eta Teknologia Hiztegia, <https://zthiztegia.elhuyar.eus/>), etc.

In short, they have applied the recommended methodology.

3.2 Some linguistic issues observed in this practice

In this section, we will shortly describe some common linguistic problems encountered in both tasks.

3.2.1 *The translation of abbreviations*

The translation of acronyms requires revision and use of specific sources. For instance:

e.g. *AdAway quiere configurar una conexión VPN para controlar el tráfico de red (original).*
AdAwayk VPN konexio bat konfiguratu nahi du sareko trafikoa kontrolatzeko. (Elia´s translation)
AdAwayk SPBeko konexio bat konfiguratu nahi du sareko trafikoa kontrolatzeko. (Revised translation)

The acronyms should be written according to the rules of Basque language for science and technology.

3.2.2 The translation of parameters or variables

Parameters or variables must be maintained and not translated. Therefore, we examined these cases in the revision phase.

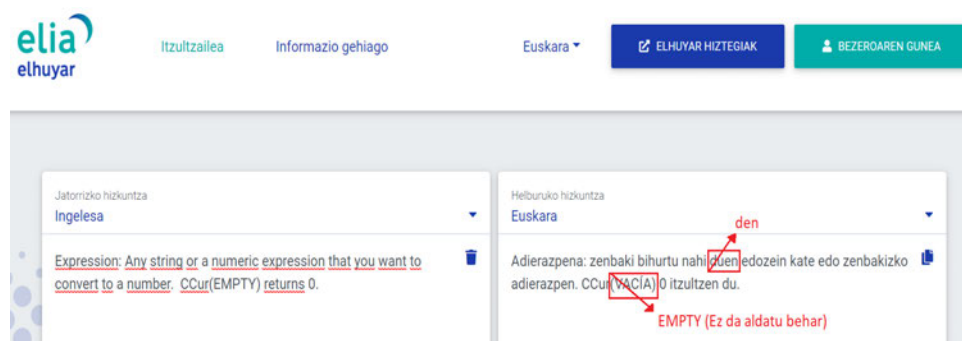


Figure1. Elia´s interface example: translation of English to Basque.

Elia translated the parameter (EMPTY) into (VACIA), but the parameters are not the object of translation. The students highlighted some of the translation errors (see Figure 1).

3.2.3 The translation of specific terms

On the one hand, translation tools cannot cope with all specialized texts. Therefore, we require additional resources to deal with terminology, because terms are nuclear elements of specialized texts. The use of terminological databases, specialized dictionaries and corpus is essential. As we can see in the below example of the term (“rooteado”), we do not always find the solution in the available resources. This is frequent in Basque because specialized Basque language is still in progress.

e.g. *Se requiere Android rooteado. [Original]*
Androiderako haria behar da. [Elia´s version, lit. “It takes thread for Android”]
Android rooteatu bat beharrezkoa da. [Student´s version]

On the other hand, although in general, the fluidity of the translator is very good and it is complicated to see failures in the content, in this case, the error is very clear. The translation of that short phrase makes no sense.

We have described some aspects that we consider relevant to the conference, leaving some linguistic aspects to other more appropriate forums. Besides, we are conscious that loss of information or failure to reflect original content because of machine translation requires further in-depth study.

4 CONCLUSIONS

One of the objectives was to provide students with tools to facilitate their autonomy when facing writing tasks. In this sense, I believe that students have become aware of the importance of following a methodology and the role of reviewing the text. In that sense, we observed that students have improved texts through the described method.

However, we seek to continue studying how machine translation tools fit students' needs. Success lies in reaching students to help them reflect on their learning and advance their process. Various studies show that more studies are needed on the needs of students to integrate machine translation into teaching; and suggest new directions for future research [8], [9]. There is an interesting debate about the challenges and risks of language technology tools for Basque [10]. However, the question is how to integrate machine translation into academic writing. As for the future, we plan to explore new ways to incorporate these tools into writing.

ACKNOWLEDGEMENTS

This article was supported by IXA group (Basque Government (excellence research group IT1343-19) as well as by COST Action CA19102 project 'Language in the Human-Machine Era'.

REFERENCES

- [1] B. Klimova, M. Pikhart A. D. Benites, Lehr C., C. Sánchez-Stockhammer, "Neural machine translation in foreign language teaching and learning: a systematic review". *Education and Information Technologies*, 2022.
- [2] X. Deng, & Z. Yu, "A Systematic Review of Machine-Translation-Assisted Language Learning for Sustainable Education. Sustainability", 14 (13), 7598. MDPI AG. 2022.
- [3] S. O'Brien, M. Simard, M. J. Goulet, "Machine Translation and Self-post-editing for Academic Writing Support: Quality Explorations" In: *Moorkens, J., Castilho, S., Gaspari, F., Doherty, S. (eds) Translation Quality Assessment. Machine Translation: Technologies and Applications*, vol 1. Springer, 2018.
- [4] Elia, Elhuyar's machine translator, Accessed 20 September 2022. Retrieved from <https://elia.eus/start>.
- [5] S. Crossley, "Technological disruption in foreign language teaching: The rise of simultaneous machine translation. Language Teaching", 51(4), 541-552. 2018.
- [6] M. L. Conde Noguero. "Los traductores automáticos en línea como recurso metodológico en el aula de Español como Lengua Extranjera", *Transforming education for a changing world / Camilo López García (ed. lit.), Jesús Manso (ed. lit.)*, pp. 304-312. 2018.
- [7] A. Carré, D. Kenny, C. Rossi, P. Sánchez-Gijón & O. Torres-Hostench. Machine translation for language learners. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 187–207. Berlin: Language Science Press. 2022.
- [8] A. G. Dorst, S. Valdez, H. Bouman, "How, when and why do humanities students at a Dutch university use machine translation?" *Translation and Translanguaging in Multilingual Contexts* Vol. 8:1 (2022), pp. 49–66. 2022.
- [9] J.J. Jolley, L. Maimone, "Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature", *L2 Journal*, Volume 14 Issue 1, pp. 26-44. 2022.
- [10] I. Leturia, Hizkuntza- eta hizketa-teknologiak eta Euskara: arriskuak, aukerak, ekidinezinak, beharrezkoak, Accessed 20 September 2022. Retrieved from <https://www.sarean.eus/hizkuntza-eta-hizketa-teknologiak-arriskuak-aukerak-ekidinezinak-beharrezkoak/>.