Machine translation use outside the language industries: a comparison of five delivery formats for machine translation literacy instruction

Lynne Bowker [1] [0000-0002-0848-1035]

[1] School of Translation and Interpretation, University of Ottawa, Canada
lbowker@uottawa.ca

1   Introduction

Since the launch of the free online tool Google Translate in 2006, which has been followed by the release of a host of similar tools (e.g. Microsoft Bing Translator, DeepL Translator, SYSTRAN Translate, Baidu Translate, Yandex.Translate, Naver Papago), machine translation (MT) has been easily accessible to anyone with an internet connection. Not only are machine translation tools easy to access, they are also easy to use. In many cases, users need only choose their language pair, copy and paste a text, and click "Translate". In other cases, a machine translation widget may be embedded in a web browser or social media platform, meaning that translation is just a click away. It is very easy to see the appeal of a tool that is free, fast, and easy to use! Therefore, it comes as no surprise that these tools are indeed being used widely. While language professionals certainly constitute an important user group, they are by no means the only one. Indeed, various groups outside the language professions use machine translation actively:

- Anazawa et al. [1] describe how practicing nurses in Japan use machine translation to stay on top of the latest developments in the international nursing literature;
- Bowker and Buitrago Ciro [2] explore the use of machine translation by researchers seeking to publish in other languages;
- Nurminen [3] recounts how patent professionals use machine translation to search for international patents;
- O'Brien and Ehrensberger-Dow [4] note that machine translation is sometimes used to support communication in a crisis situation.

In all of these cases, the authors emphasize that some kind of training can help tool users to make better decisions about employing machine translation and to optimize its use. What's more, authors such as Mundt and Groves [5] and Lee [6], among others, have identified university students as a very active group of machine translation users, and what better place to offer machine translation literacy instruction than at a university?

However, even at a university, instruction can take many forms, and over the past couple of years, we have had the opportunity to try out five different formats for delivering machine translation literacy instruction. In this paper, we first introduce the basic

notion of machine translation literacy and share some general content that we believe could usefully be included in a machine translation literacy module for non-translation students. Next, we briefly present the five different machine translation literacy instruction formats that we have pilot tested, as well as some general feedback received from the participants. This is followed by a comparative summary of some strengths and weaknesses of each format, along with some general conclusions.

2   Machine translation literacy

As noted previously, machine translation tools are easy to access and straightforward to use, but this does not mean that people without a translation background instinctively know how to use these tools critically. Machine translation literacy is less about knowing which buttons to press and more about deciding whether, when or why to use this technology [2]. In this way, it has a strong cognitive or conceptual element that focuses more on critical thinking tasks, such as evaluating the suitability of a text for translation by machine, or weighing the benefits and risks of using machine translation against other translation solutions. Owing to space limitations, it is not possible to provide a comprehensive description of the contents of a machine translation literacy module; however, key elements that can be usefully covered as part of such a module are briefly summarized below. It is also important to recognize that machine translation literacy does not take one single form; rather, it is a customizable concept that can (and should) be adapted to meet the needs of the target audience. The summary below focuses on the needs of university students who are not studying to become language professionals, but machine translation literacy instruction for other groups (e.g. primary or secondary school students or teachers, translator trainees, journalists, health care workers, workers in NGOs) may incorporate different elements or explore them to a different depth.

For non-translation undergraduate students, the basic machine translation literacy module that we designed had four main components, which were covered in more or less depth, depending on the format and time available:

1. Understanding data-driven approaches to machine translation
2. Transparency and machine translation use
3. Risk assessment and machine translation
4. Interacting with machine translation

2.1   Understanding data-driven approaches to machine translation

Having a basic understanding of how data-driven approaches to machine translation (including neural machine translation [7]) work will enable students to better understand the strengths and limitations of these tools. For instance, understanding the notion of sensitivity to training data can help users to realize why these tools can be more or less useful for different language pairs, domains or text types (e.g. low vs high resource situations).

Students who understand how data-driven machine translation systems work will also recognize that different tools (e.g. Google Translate, DeepL Translator, etc.) are

likely to produce different results. Many of the non-translation students that we worked with had previously believed that while the interfaces of these tools may differ slightly, all machine translation tools were driven by the same engine and would produce the same results. It had not occurred to many of these students that they would get different results by trying different systems (which had been trained using different corpora), and many had never looked beyond Google Translate. Similarly, they had not realized that the systems were constantly "learning" and so the results may improve from one trial to the next, and they should not write off a tool as being unhelpful based on one experience.

Finally, learning about sensitivity to training data also makes students aware of the potential for algorithmic bias, including problems such as inappropriate selection of pronouns in languages that are marked for gender [8].

2.2 Transparency and machine translation use

The concept of transparency is relevant in several ways for student users of machine translation tools. Firstly, it may be important to point out that the use of machine translation for course work may be more or less appropriate depending on the learning objectives of the course and the preferences of the instructor. It is also important to emphasize academic integrity with regard to the need to properly cite and reference material that has been translated from another language; the wording may change, but the original author should still be cited as the source of the ideas. Another reason that transparency is important is that it allows the readers of the text to take the fact that it has been machine translated into account as part of their own decision-making when deciding how much to trust the content. For all these reasons, students are encouraged to be transparent about their use of machine translation.

2.3 Risk assessment and machine translation

The notion of transparency has links to the idea of risk assessment. For students without a background in translation, the idea that translations can have different purposes and take place in different contexts may not be immediately apparent. Students need to learn to evaluate different types of translation tasks and recognize them as being low-stakes or high-stakes tasks where the use of a machine-translated text may carry a lower or a higher risk. This could include educating students about the differences between using machine translation for information assimilation (e.g. understanding a friend's social media post) versus text dissemination (e.g. submitting an essay to a professor), or the difference between using machine translation to compose an email to a friend versus composing an email to a prospective employer.

Another type of risk assessment that is relevant to students is determining whether the material that they want to translate is sensitive or confidential. Most of the students that we worked with had not given much thought to what happens to the text that is entered into a free online system, and many were surprised to learn that this text does not simply disappear once they exit the tool. Making students aware that they should

not enter sensitive information (e.g. banking details, health information, proprietary research) into a free online machine translation tool is an important component of machine translation literacy for this group.

2.4 Interacting with machine translation

Finally, the students that we worked with were eager to learn about how they can interact with a machine translation tool in order to improve the quality of the results. While the vast majority of students are aware of the likely need to make some adjustments to the output, few have given any consideration to the idea that changing the input to reduce ambiguity can result in higher quality output. Therefore, if the goal is to use machine translation as a writing aid to help produce a text in a second language, it may be easier for students to make adjustments to their input text, which is likely written in their dominant language. While this idea of "garbage in, garbage out" seems very obvious to translators, it is not necessarily something that occurs to people outside the language professions.

Depending on the language combinations of the participants, it may be easier or more difficult to work together on practical exercises on pre- or post-editing. The tips that are relevant for one language, or the errors that are made by a given system, may not be the same as those that are typical for another. Nonetheless, students consistently expressed an interest in gaining hands-on experience with pre- and post-editing, so if it is not possible to include this in the instruction session, it could be worth preparing a resource sheet and some exercises that participants can work on later.

3 Machine translation literacy instruction formats

In the 20-month period between October 2019 and May 2021, we had the opportunity to test five different formats for delivering machine translation literacy instruction to undergraduate students who are *not* studying to become language professionals. These five formats include: 1) a library workshop; 2) an English-as-a-second-language course; 3) a translation for non-translators course; 4) an information literacy course; and 5) a digital humanities summer institute course. Below, we will present a summary of these experiences, including some highlights from student feedback.

3.1 Library workshop

The first format that we tested for delivering machine translation literacy instruction to university students took the form of an optional one-hour workshop offered in the autumn semester of 2019 through the university library at two different institutions in Canada: the University of Ottawa and Concordia University. At each institution, this workshop was promoted through the library, the international students office, and the student success centre. The short length of the workshop meant that it was necessarily high-level and was mostly a lecture-style format with some time for questions and answers. Students were given a resource sheet and some ideas for practical exercises (e.g.

tips on pre- and post-editing, comparing different machine translation systems) that they could take away and try at home. Between 25 and 30 students attended at each institution. The vast majority were undergraduate students, although a handful of graduate students participated. The students came from a wide range of disciplines in both the humanities and sciences, and they spoke a diverse range of languages (although none were Anglophone).

At the end of the workshop, participants were asked to provide a short evaluation of the workshop. Most claimed to find it valuable, noting that they had learned new things. Suggestions for improvement included using fewer specialized terms and giving more language-specific tips for pre- and post-editing. More time for practical exercises was also identified as something to strive for in future iterations. More details about this experience can be found in Bowker et al. [9].

3.2   English-as-a-second language course

The second format that we tested consisted of integrating a module on machine translation literacy into a course on English-as-a-second language (ESL). Once again, we tried this experiment at both the University of Ottawa and at Concordia University. At the University of Ottawa, all 22 students in the class were native speakers of Chinese, but they were studying a wide range of subjects. In contrast, at Concordia University, the 23 students spoke a range of languages, but all were studying business and the course therefore focused on business English.

In addition, at Concordia University, we were also invited to offer the workshop to a group of 24 ESL instructors prior to delivering it to the students. This was organized as a type of professional development opportunity for the instructors, and it was very informative because it allowed us to better understand the concerns of the instructors and to adapt the workshop content accordingly.

At both institutions, we essentially participated as a guest speaker to deliver the one-hour workshop in a single class, and again, the short timeframe for delivering the workshop limited the amount and depth of material that could be shared or the number of practical exercises that could be undertaken. Once again, students claimed to find the content interesting and relevant overall, but they expressed a desire for more time to be dedicated to practical exercises.

Additional details about the experience of integrating machine translation into an ESL course at the University of Ottawa, and in a "train the trainers" format and ESL course at Concordia University can be found in Bowker [10] and Bowker [11] respectively.

3.3   Translation for non-translators course

Next, we attempted to move away from the standalone workshop format or guest lecture delivery and to better integrate machine translation literacy into a broader course. In the Fall 2020 semester, we designed a full-semester (12-week) course on translation for non-translators at the University of Ottawa. The course was offered at an introductory (first-year) level and was open to students across the whole campus, regardless of their

major area of study. The course proved to be popular and attracted 50 students, so it was offered again in the Winter 2021 and Summer 2021 semesters, with a similar result. Over the three iterations of the course, students came from more than 20 disciplines and more than 20 different native languages were represented (including English).

Because the course was an introduction to translation more broadly and not only to machine translation, it meant that students had an opportunity to learn some key translation concepts first, and therefore to have a slightly firmer footing in some of the basics before undertaking the machine translation literacy module in the fifth week of the course. In addition, because the course unfolded over a longer period, we were able to increase the amount of time spent on machine translation literacy to three in-class hours, and to increase the level of practical activity, both in class, but also in the form of homework and assignments to be conducted before and after class.

Another benefit of offering machine translation literacy in this format is that we reached many students who have English as a native language – a group that had not been represented at all in our prior efforts to deliver this training through the library or in an ESL class. The Anglophone students confirmed that they too are active users of machine translation, though more for their leisure activities than for their studies. Nonetheless, according to the results of the feedback survey, they claimed to find the machine translation literacy training to be informative, and overall, the majority of students in the translation for non-translators course recommended that this type of training be offered regularly and on a wide-scale across campus.

Additional information about the course on translation for non-translators and how it incorporated machine translation literacy will be shared in an upcoming publication [12].

3.4 Information literacy course

Up to this point, our efforts had focused mainly on targeting venues where participation was elective, so as a next step, we investigated the possibility of embedding machine translation literacy instruction into a compulsory course that focused not on language or translation but rather on digital and information literacy. One reason for doing this is that we wanted to see how machine translation literacy instruction would be received by those who were not expressly looking for it. At the University of Ottawa, all first-year students in the Faculty of Arts are required to take a minimum of four compulsory courses that focus on developing critical thinking and academic writing skills. These courses must be selected from a pool of courses offered by the Department of Philosophy, the Department of English, and the Interdisciplinary Studies program. Each year, the Faculty of Arts invites professors from different departments within the Faculty to team up and pitch a theme for an interdisciplinary course that meets the following requirements:

> **AHL 1100 Introduction to Interdisciplinary Study in the Arts (3 units)**
> Exploration of at least two disciplines in the Faculty of Arts whose conjunction illuminates contemporary situations and debates. Development of critical reading and academic writing.

This course has variable topics. Students may take this course twice with different topics.

For the Winter 2021 semester, we successfully pitched the topic "New Literacies for the Digital Age" to be co-taught by a professor from the School of Information Studies and a professor from the School of Translation and Interpretation (i.e., the present author). A total of 80 students registered for the course, and they came from 13 different programs in the Faculty of Arts. While the majority were native English speakers, there were also speakers of eight other languages, including 14 Chinese speakers, 8 French speakers, and a smaller number of speakers of Arabic, Hindi, Persian, Spanish, Ukrainian and Vietnamese.

Along with modules on more traditional aspects of information literacy (e.g. effective searching in library catalogues and on the internet, referencing and citation), the course also contained instruction on media literacy (e.g. fake news) and scholarly communication literacy (e.g. predatory publishing), as well as a module on machine translation literacy. Similar to the case of the module that had been integrated into the course on translation for non-translators, it consisted of three in-class hours, along with some homework and assignments to be done outside of class. Once again, as reported on the feedback survey, the vast majority of the students claimed to find the machine translation literacy module to be useful and they recommended that it should continue to be offered as part of a compulsory course on information literacy.

To learn more about the experience of integrating machine translation literacy instruction into a broader course on digital and information literacy, consult Bowker [13].

3.5 Summer institute course

The final format in which we piloted the delivery of machine translation literacy instruction was as part of the 2021 Digital Humanities Summer Institute: Technologies East (DHSITE) [14], which is open to students from all disciplines and levels who are interested in exploring aspects of the Digital Humanities (DH). The summer institute took place during the last two weeks of May 2021, and it consisted of an offering of six 18-hour mini-courses on different subjects in DH (e.g. Python programming, text analysis, linked open data), from which students could choose up to two. We offered a course on machine translation in which there were eight participants, including both undergraduate and graduate students. Of these, three had a background in translation, while the other five came from disciplines that included computer science, business, music, psychology, and public administration. Four different native languages (English, French, Chinese and Polish) were represented. The diversity of backgrounds, languages and levels brought richness to the discussions but also posed challenges with regard to pitching the material appropriately.

As this format had 18 hours of in-class time, as well as additional time for homework outside of class, it was possible to explore the subject of machine translation much more deeply than in the previous formats. This meant that, in addition to the key elements of machine translation literacy content described previously, there was also time to con-

sider the history of machine translation, methods of evaluating machine translation systems as well as their output, and a broader range of ethical issues surrounding tool use. At the time of writing, the formal course evaluations have not yet been received, but anecdotally, we can report that the participants were active and engaged throughout the course, and even students with a translation background appeared to be learning new things.

4 Comparison of different formats for delivering machine translation literacy instruction

Having experimented with five different formats for delivering some kind of machine translation literacy instruction, we can observe that the various formats have different strengths and weaknesses. We have summarized some of the main pros and cons in Table 1.

	Strengths	Weaknesses
Library workshop	• Low level of commitment required (for both participants and instructors) • Open to anyone on campus • Potential for immediate feedback • Potential to gauge interest in a more advanced follow-up workshop	• Challenging (and time-consuming) to promote • People don't recognize that they need it and so may not register • Participants have no background in translation so it's a steep curve • Very short, resulting in superficial treatment and limited practice • May only be offered once or twice per year (will take a long time to reach a critical mass of people) • No opportunity for longitudinal observation (e.g. to see if the information is put into practice or if behaviour changes over time) • "Train the trainer" required beforehand if delivered by a non-translator
Integrated into a compulsory English-as-a-second language course	• No need for marketing • Can reach a wide range and large group of international students • Can work with authentic texts in the context of course requirements (e.g. texts students need to produce for assignments)	• May meet resistance from language teachers who fear that MT use may be contrary to language learning objectives • May be misinterpreted by students who could perceive MT use as a shortcut to alleviate the need to learn a language

	• MT has potential to act as an aid for language learning and reinforcement (presents MT in a positive light, rather than as a taboo or shameful practice)	• Participants' knowledge of translation is limited and often restricted to a language learning context • "Train the trainer" required beforehand if delivered by a non-translator • Will not reach Anglophones or those with a high level of English
Integrated into an optional translation course for non-translators	• Participants are interested and motivated to learn about translation • Participants learn some basic translation concepts first and can build on these in the MT literacy module of the course • Can spend more time on it, and incorporate more practical work (e.g. homework, exercises) • Opportunity for more longitudinal observation (e.g. to see if knowledge is put into practice or leads to a change in behaviour) • Can reach both English speakers and speakers of English as an additional language • Course taught by a translation professor already up to speed (or able to get up to speed quickly) on MT	• High level of commitment required by participants (must take a whole course on translation, not just a module on MT) • As an optional course, it will only reach those who are actively seeking this knowledge
Integrated into a compulsory information literacy course	• Reaches a wider range of students (including those who may not realize they need it, and English speakers, who may not think MT is as relevant to them) • Can spend more time on it (3-6 hours), and incorporate more practical work (e.g. homework, exercises)	• Participants have no background in translation • "Train the trainer" required beforehand if it is to be delivered by a non-translator
Digital Humanities Summer Institute course on MT	• Participants are highly motivated to learn about MT • Can explore concepts thoroughly and incorporate more practical work (e.g. homework, exercises) • Course taught by a translation professor already up to speed on MT	• Reaches relatively few students • High level of commitment required by participants • Challenging to manage different backgrounds and levels of prior knowledge

Table 1. Comparative summary of some strengths and weaknesses of different formats for delivering machine translation literacy instruction.

5   Concluding remarks

Free online machine translation systems are very attractive because they are easily accessible and easy to use. However, this does not mean that users – especially those without a background in translation – instinctively know how to use them in a critical way. Therefore, there is an emerging need for machine translation literacy, and correspondingly, a need for machine translation literacy instruction. Having said that, there is no single right way to help users develop machine translation literacy. Rather, as noted previously, this is a highly customizable concept, and the content can (and should) be adapted to meet the specific needs of the target audience. In this article, we have focused on sharing the results of our efforts to teach machine translation literacy to (primarily) undergraduate students who are *not* training to become translators or other language professionals. We have tested five different delivery formats, each of which have strengths and weaknesses that may make them more or less suitable for different contexts. While no approach is perfect, we believe that all have some value and indeed the feedback received in each case was largely positive (though there was always room for improvement).

A general take-away from this experience is that it confirms that students of many backgrounds are eager to learn how to make better use of machine translation, and it is very important to recognize that things which are obvious to language professionals are *not* obvious to those without a translation background (e.g. GIGO, different machine translation systems generate different results). In other words, there is no reason why students would instinctively know how to be informed and critical users of machine translation tools, so there is scope for and benefit to offering some form of machine translation literacy instruction to this group.

Beyond undergraduate students, machine translation literacy instruction is relevant for other groups too, including those in the language professions, but also graduate students and more established scholars, as well as secondary or even primary school students. Indeed, the next delivery format that we will be piloting during Canada's Science Literacy Week in September 2021 is a machine translation literacy workshop for teens that will be delivered in collaboration with the University of Ottawa's Faculty of Engineering Outreach team.

Moving forward, a key question to consider when planning machine translation literacy instruction is *who* will do the training. Will it always be done by a machine translation expert or even a language professional? If machine translation literacy is to become embedded in other contexts (e.g. information literacy instruction, digital literacy instruction, ESL teaching, high school or primary school), then it will be necessary for people from other backgrounds (e.g. librarians, teachers) to become involved in delivering machine translation literacy instruction. In cases where the training will be delivered by a non-language professional, some type of "train the trainer" preparation will likely be necessary, as was done for the ESL instructors at Concordia University [10]. With a view to helping to "train the trainers", we are in the process of developing a range of resources for machine translation literacy, which can be found on the Machine Translation Literacy Project website [15]. Some additional resources are also available

through the European Union Erasmus+ project "MultiTraiNMT — Machine Translation training for multilingual citizens" [16, 17].

Acknowledgements

References

1. Anazawa, R., Ishikawa, H., & Kiuchi, T.: Use of online machine translation for nursing literature: A questionnaire- based survey. *Open Nursing Journal, 7*(1), 22–28 (2013).
2. Bowker, L., & Buitrago Ciro, J.: *Machine translation and global research*. Bingley: Emerald (2019).
3. Nurminen, M.: Raw Machine Translation Use by Patent Professionals. A case of distributed cognition, *Translation, Cognition & Behavior, 3*(1), pp. 100–121 (2020).
4. O'Brien, S., & Ehrensberger-Dow, M.: MT literacy: A cognitive view, *Translation, Cognition & Behavior 3*(2), pp. 145–164 (2020).
5. Mundt, K., & Groves, M.: A double-edged sword: the merits and the policy implications of Google Translate in higher education, *European Journal of Higher Education 6*(4), pp. 387–401 (2016).
6. Lee, S.: The impact of using machine translation on EFL students' writing, *Computer Assisted Language Learning, 33*(3), pp. 157–175 (2020).
7. Koehn, P.: *Neural machine translation*. Cambridge: Cambridge University Press (2020).
8. Monti, J.: Gender issues in machine translation: An unsolved problem? In: L. von Flotow & H. Kamal (Eds.), *Routledge Handbook of Translation, Feminism and Gender*, 457–468. London: Routledge (2020).
9. Bowker, L., Kalsatos, M., Ruskin, A., & Buitrago Ciro, J.: Artificial Intelligence, Machine Translation, and Academic Libraries: Improving Machine Translation Literacy on Campus, *The Rise of AI: Implications and Applications of Artificial Intelligence in Academic Libraries*. (Eds. S. Hervieux and A. Wheatley). Chicago: *Association of College & Research Libraries (ACRL)* (2021).
10. Bowker, L.: Chinese speakers' use of machine translation as an aid for scholarly writing in English: A review of the literature and a report on a pilot workshop on machine translation literacy, *Asia Pacific Translation and Intercultural Studies*, 7(3): pp. 288–298 (2020a).
11. Bowker, L.: Machine translation literacy instruction for international business students and Business English instructors, *Journal of Business and Finance Librarianship*, 25(1-2): pp. 25–43 (2020b).
12. Bowker, L.: *Introducing translation to non-translators*. London: Routledge (forthcoming).
13. Bowker, L.: Promoting linguistic diversity and inclusion: Incorporating machine translation literacy into information literacy instruction for undergraduate students, *The International*

Journal of Information, Diversity and Inclusion 5(3) (2021). https://jps.library.utoronto.ca/index.php/ijidi/issue/archive

14. Digital Humanities Summer Institute: Technology East (DHSITE): https://dhsite.org/dhsite-2021/

15. Machine Translation Literacy Project: https://sites.google.com/view/machinetranslationliteracy/

16. MultiTraiNM—Machine Translation training for multilingual citizens: https://www.multitrainmt.eu/index.php/en/

17. Ramírez-Sánchez, G., Pérez-Ortiz, J.-A., Sánchez-Martínez, F., Rossi, C., Kenny, D., Superbo, R., Sánchez-Gijón, P., & Torres-Hostench, O.: MultiTraiNMT: Training materials to approach neural machine translation from scratch. *TRITON 2021: Proceedings of the Conference* (2021).