INNOVATION IN LANGUAGE LEARNING International Conference

The Role of Machine Translation in Language Studies and Beyond: Evaluation and Future Directions

Iglika Nikolova-Stoupak

Kyoto University, Japan

Abstract

It is hard to deny the significance of the role that machine translation has taken in the routine of businesses, translation specialists and even young language learners today. Studies show that the use of machine translation by translation students leads to a number of benefits, including enhanced understanding of language, higher speed of translation and even increased creativity. To go further, the computational or mathematical aspect has so much as seemingly been "pushing away" the language-related side of computational linguistics as an academic field. The reason behind such a trend can be traced to a more general association of language and mathematics, accounted for in Tanaka-Ishii's elaborate study, "Statistical Universals of Language" (2021). Is the study of language(s) becoming obsolete? On the contrary, an adequate reaction on the side of humanities studies would be an adaptation of curricula that accounts for the role of technological and research advancements in the field of translation and language studies. In fact, language could fit more than neatly in the historically highly mathematical curriculum of Computer Science. This paper will provide specific guidelines as to novel modalities through which language studies can be incorporated in a middle school curriculum.

Keywords: *translation studies, machine translation, computational linguistics*

1. Introduction

Machine Translation or the automatic translation of text from one language to another by means of computer software is a technological advancement that is difficult to bypass as it has made its way and continues to become ingrained in the experience of individuals and business entities; in personal and professional activities. The technology's role in education is especially visible when one peeks into university-level translation curricula, but even much younger students use it in their foreign language classes. One may so much as say that the definition of translation itself has changed as a result of Machine Translation technology [3]. This paper will firstly examine the role that automatic translation currently accommodates within education. Then, the vastness of this role will be presented as a result of a more generalised tendency of "mathematisation" of language as visible within the growing field of Computational Linguistics. Finally, a natural path for the re-introduction of language learning will be suggested.

2. Machine translation in education

This part of the paper will examine the current role that Machine Translation has in different facets of education. Most studies on the topic discuss university-level studies, specialised in translation, and little place is accorded to general language-learning practices as well as to earlier education, when a number of crucial habits are formed. One should note that in most cases, English is considered as the target language of translation and the most popular automatic translation tool, Google Translate, is often emphasised.

Deng and Yu (2022) offer a contemporary literary review of 26 articles on the use of automatic translation in language learning. Some common activities that students (mostly, at university level) are involved in for the purpose of practice with the technology involve comparison of human and machine-translated text and analysis and editing of errors [1:82]. The authors come at the natural conclusion that automatic translation provides help to students in various ways, such as through improving the quality of their work and boosting their confidence, whilst the technology may be detrimental if relied on excessively and without sufficient critical thinking. A prevailing opinion of students is that additional training with automatic translation would be of help [1:76].

Yanti and Meka (2019) point out that undergraduate translation students demonstrate the correct intuition that Google Translate is useful at word-level and less so sentence-level and paragraph-levels [9:132]. Insightfully, the participants voice an opinion that the acquisition of good grammar skills in their target language helps prevent mistakes that may stem from automatic translation [9:135].

An experimental task with two mixed groups of BA- and MA-level translation students, one of which is allowed to resort to an initial machine-translated text, reveals the following additional benefits of using the technology: less time spent on the task, a higher understanding of complex terms and even higher creativity as proven by the variety of word choices applied [6:174-177]. Also, problems such as too literal translation are shown not to be more frequent within the work of the aided group [6:174].

Kenny and Doherty (2014) propose a holistic approach to translator training, through which human translators are encouraged to participate in all stages of an automated translation process as opposed to simply when it comes to editing or evaluation of output. The authors also point out that it is crucial to provide a common ground between different professionals, thus avoiding a situation where Google Translate's speed and accuracy give a wrong idea of the implied work of either human translators or computer scientists [3:14].

Turning to EFL education, Raza and Nor (2011) point out that the use of automatic translation may be especially useful in cases when teachers do not speak a student's native language and can therefore not use translation as a means of teaching [7]. Examining the dynamics in a highly international EFL class in Malaysia, Jaganathan (2014) reveals the practices of the highest scoring students: they do not highly rely on automatic translation in order to understand the meaning of text and they refer to additional sources, such as dictionaries as well as critically examine all matching translations prior to selecting one [4:6]. To sum up, just like students in specialised translation classes, EFL students can benefit from automatic translation when they use it optimally.

3. Underlying trends in computational linguistics

Machine Translation subscribes within Natural Language Processing, which in turn is associable with the more general scientific field of Computational Linguistics; that is to say, all technology-aided modelling and representation of natural language. As the name implies, the domain requires and produces knowledge of both computing (and the implied mathematics and logic) and linguistics. However, as will be shown in this section, the former seems to be increasingly overpowering the latter.

A comprehensive contemporary work that is worth mentioning as a milestone in Computational Linguistics is Tanaka-Ishii's *Statistical Universals of Language* (2021). The study offers an exhaustive yet understandable presentation of a number of mathematical trends in human language that are statistically-provable through modern corpus technology (such as Zipf's Law and several of its derivatives) as well as discusses them as a measure of the validity of computer-generated language models. Tanaka-Ishii admittedly takes a step back from language as the tool of communication we are familiar with in order to shed light on its inner workings [8:185]. The "meaning" that is intrinsic to language is notably omitted from the vast majority of the discussion.

In order to further analyse the relationship of leading researchers in the field of Computational Linguistics with language and linguistics, the author of this paper conducted a survey among students in a Natural Language Processing laboratory in the University of Kyoto. The very rigorous entrance exam that students are required to take prior to admittance is built on a strong basis of advanced mathematics and computing theory. The seven respondents unambiguously showed a personal interest in language, 85.7 % of them indicating an advanced to proficient level of English as a foreign language. Also, all of the respondents indicated knowledge of at least one additional foreign language. To go even further, when prompted to freely describe the main reasons for their choice of specialisation, two students indicated a shared interest in language and computing, and one underlined their eagerness to help facilitate communication between "people with different language backgrounds".

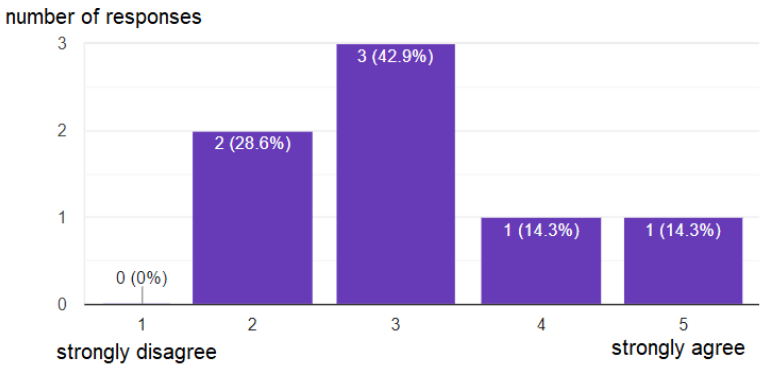

Figure 1. "I think that additional training in language/linguistics would benefit my academic work."

However, the voiced interest in language seems to not have made it to professional activity, all students indicating a prior background in Computer Science or related scientific fields as well as admitting to a larger interest in the “sciences” rather than “humanities”. 71.4% of the students have conducted research involving a human language they are not personally proficient in. An enquiry on students’ interest in deepening their professional knowledge of language and linguistics is met with varied responses (Fig. 1).

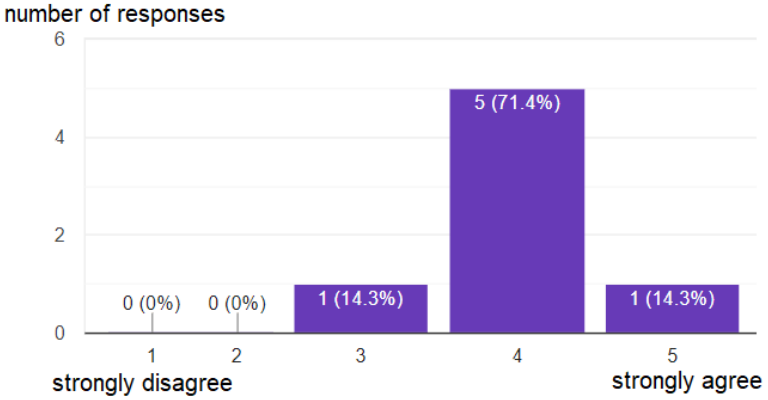

Figure 2. “I think that machine translation can replace human translation.”

The respondents’ opinion concerning the current place and future potential of technological advancements like Machine Translation proves highly optimistic: 85.7% disagree that computers will always demonstrate limitations in the understanding and production of language and, relatedly, the same number of students agree that machines can replace human translation (Fig. 2).

4. An early “Computational Linguistics” curriculum

Whilst the ongoing trend for the automation and mathematisation of language is both undeniable and understandable, this bringing together of the two once seemingly irreconcilably different scientific and educational spheres implies just as strongly that language now has a place within curricula that have traditionally been almost exclusively mathematical, such as computer studies. Of course, there are good reasons for mathematics to play a major role in introductory computer science. As Henderson (2005) explains, a computer scientist needs to possess logical and abstract thinking skills, a sensitivity to patterns and the ability to move from specific to general. Interestingly enough, however, computing is also comparable to mathematics on grounds of its use of symbols and role in communication [2], features that are also readily associable to language. The following is a preliminary curriculum, meant to integrate the notion of computational linguistics within the study of computing as early as in middle school (or the beginning of high school, with minor modifications). Special focus is placed on machine translation due to its already active role in students’ personal and educational lives.

Weeks 1-2: Machine Translation: Fails

The topic can be entered gradually and cheerfully as students are shown a variety of thematic memes and google-translated songs. The YouTube channel “Twisted Translations” has especially hilarious, children-friendly content, and the process of generating lyrics can be recreated as an activity. A discussion may issue about the types of mistakes artificial technology makes, the reasons behind them and the limitations of a computer’s “thinking”. A notion is to be established that automatic translators rely on a large corpus i.e. a large number of examples that they have “seen” previously.

Weeks 3-4: Machine Translation: Efficient and Ethical Use

A discussion may be initiated about students’ current habits of using tools like Google Translate. By means of demonstration, students are to be let know that automatic translation can provide wrong output when text is extensive or ambiguous. An overreliance on the tool is to be emphasised as detrimental to one’s language acquisition as well as unethical. A link is to be drawn with plagiarism and it is to be noted that machine-translated writing is in fact recognisable by a variety of modern tools.

Weeks 5-6: Machine Translation: Activities

In accordance with class-appropriate foreign language studies, students are to be given in-class translation tasks that consist in 1) translating a short text while not having access to automatic translation tools 2) automatically translating and then post-editing a short text. The experience is to be critically evaluated and compared.

Weeks 7-8: Natural Language Processing: Discovery of Modern Tools

Students are to be let discover and experiment with a variety of technological tools, such as state-of-the-art chatbots, a converter to Shakespearean English, generators of images from text, etc. Specific tools may be selected and presented by students, and an exhibition of generated creative output may be hosted.

Weeks 9-10: Grammar in Computational Linguistics

A parallel is to be drawn with students' first language grammar classes. For instance, the parts of a sentence can be automatically annotated and illustrated with easy-to-perceive schemes, and words can be brought to their most basic forms (tokenisation). The concept of "context-free-grammar" is to be defined and illustrated by means of simple examples; students are to create their own simple "language" with a very limited set of symbols and rules.

Weeks 11-12: Language as Code

Students may be introduced to the principles of cryptography and asked to complete simple tasks, such as the application of a provided key to reveal a message. Then, examples of programming languages from binary code to natural language are to be shown in a historical perspective.

Week 13: Conclusion and Personal Thoughts

Students are to present their thoughts on a relevant topic of their choice in the form of an essay or presentation. Example topics: Will human translators still be needed in the future? Will computers ever be able to think? What is the difference between natural and programming language?

5. Conclusion

This paper has looked at current facts related to the role of machine translation in different types and levels of education as well as discussed the deeper cause behind the escalating association between computing and language. To meet this trend, a provisional middle school curriculum of computational linguistics has been offered. Of course, this curriculum is to be adapted in accordance with advancements in the fast developing field. It can also be taken as a baseline and adapted to different levels of study and even to teacher training.

References

[1] Deng, X. and Yu, Z. (2022). A Systematic Review of Machine-Translation-Assisted Language Learning for Sustainable Education, *Sustainability*, 13:75-98.

[2] Henderson, P. (2005). The Role of Mathematics in Computer Science and Software Engineering Education, *Advances in Computers*, 65:349-395.

[3] Kenny, D. and Doherty, S. (2014). Statistical Machine Translation in the Translation Curriculum: Overcoming Obstacles and Empowering Translators, *The Interpreter and Translator Trainer,* 8: 276-294.

[4]Jaganathan, P. (2014). An Analysis of Google Translate Use in Decoding Contextual Semanticity Among EFL Learners, *Asian Journal of Research in Social Sciences and Humanities*, 4:1-13.

[5] Lee, S. M. (2021). The Effectiveness of Machine Translation in Foreign Language Education: A Systematic Review and Meta-Analysis, *Computer-Assisted Language Learning.*

[6] Schmidhofer, A. and Mair, N. (2018). Machine Translation in Translator Education, *CLINA*, 4: 163:180.

[7] Raza, M. and Nor, F. (2018). Google Translate in an EFL Classroom, *Technology in EFL Classroom*, 30:7-21.

[8] Tanaka-Ishii, K. (2021), *Statistical Universals of Language: Mathematical Chance vs Human Choice*, Springer, Cham.

[9] Yanti, M. and Meka, L. M. C. (2019). The Students' Perception in Using Google Translate as a Media in Translation Class, *Proceedings of the 3rd INACELT (International Conference on English Language Teaching)*, 128-146.