Case Study — ProjecTA-U: Where Artificial Intelligence (Science), Machine Translation (Technology) and Translation Studies (Humanities) Meet to Improve Higher Education Students' Access to Global Knowledge

Anna Aguilar-Amat, Pilar Cid-Leal, Marta Fuentes, Olga Torres-Hostench

Introduction

Artificial intelligence techniques and big data are increasingly being used in machine translation. Nevertheless, critical human thinking is still needed to fix machine translation mistakes. In the ProjecTA-U project (Ref. Ref.FFI2016-78612-R) we approach our research from different perspectives (scientific, technological and humanistic) to improve communication in a multilingual and multicultural Higher Education (HE). Some of the research questions of the project are: which language(s) do HE students learn in? How do HE students access academic knowledge? Do they use machine translation as a tool to help with comprehension? To avoid ethnocentric views, would it be possible to use machine translation to encourage the provision of access to academic knowledge in languages other than English? Considering that machine translation has not yet reached the desired efficiency, how could human professional translation skills be transferred to digital users (students and staff) in higher education in a fast and easy way in order to properly fix machine translation errors? How can domain-specific machine translation systems be created and trained?

Objectives

The ProjectTA-U project applied to Higher Education has three aims:

1. To conduct a far-reaching survey on the use of foreign languages and machine translation by students to measure the extent to which they access global university knowledge;

2. To develop training materials to empower university students and staff to be able to fix machine translation mistakes, apply strategies from professional humanistic translation studies and develop critical thinking, and;

3. To present the MTradumàtica web platform developed in the project to potential users and encourage them to use the user-friendly interface to train and customize domain-specific machine translation systems.

Development

An exhaustive questionnaire of 40 items on the use of foreign language and machine translation by HE students was designed, validated and piloted at the Universitat Autònoma de Barcelona (UAB). Preliminary results of the pilot survey conducted on seventy students from the Degree in Primary Education at UAB show that despite 55% of students having a good knowledge of English, only 20% occasionally use and read sources in a foreign language, and that foreign language is always English. Up to 41% sometimes use machine translation at present and up to 61% have used it in the past for academic homework. More specifically, machine translation is more used as a dictionary (54%) and to understand texts (71%) than to write in a foreign language (36%). Despite the widespread use of machine translation, 52% of students do not rely on its output. In the future, the survey will be extended to include other faculties and universities. The results will be accompanied by diagnosis and proposed measures and training materials to improve access to global knowledge and

encourage users to apply critical thinking to machine translation output.

For advanced machine translation users and/or, for instance, research centres interested in machine translating their own domain-specific materials, we developed MTradumàtica, a free, Moses-based web platform for training and using machine translation systems with a user-friendly graphical interface. Its goal is to offer users a free open tool to customise their own machine translation with domain-specific engines. Contrary to generic machine translation systems, which do not allow for customization of content and whose output may not be focused on a specific domain, "these engines are built from domain-specific parallel corpora. This means that users can use their own resources or open resources (such as corpora from the Opus collection: http://opus.lingfil.uu.se) and customise their own engines according to their needs." [1]

MTradumàtica is currently available for testing, and can be installed either on a server accessible from any computer or standalone on a PC. "A compressed package is provided (less than 3 MB) with installation instructions, which will allow any user to install their own version of MTradumàtica." [2]

Users can create a domain-specific engine by uploading sentence-aligned parallel files in two languages on a specific domain in the usual Moses text format or TMX format. Then, the system uses these files to train a translation model and a language model. A translation model is built by automatically comparing both languages and establishing direct statistical relationships between their elements. A language model is built from monolingual texts to provide more natural and correct output in the target language. Finally, the engine is trained with the provided corpora and is then ready to be used.

Figure 1. Screenshot of MTradumàtica, a domain-specific machine translation system trainer.
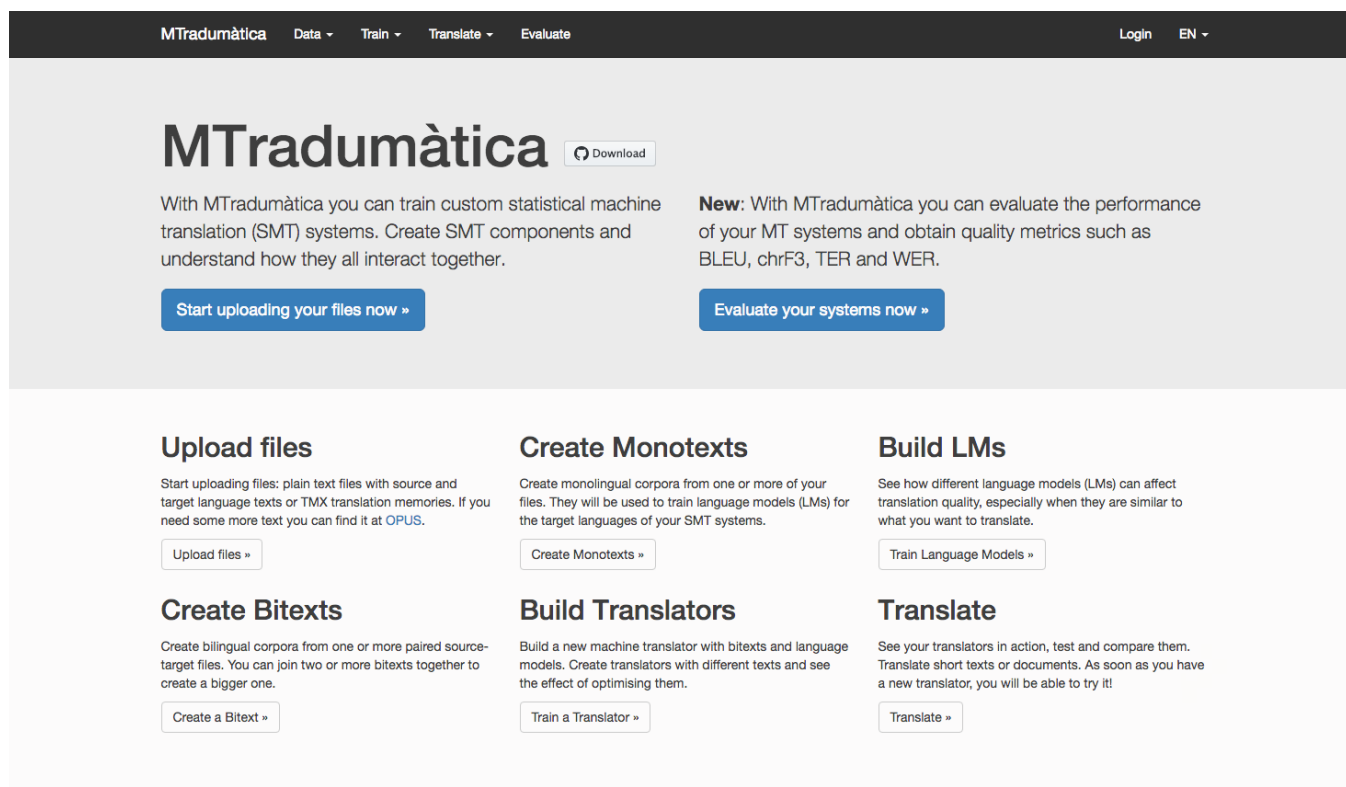
1. Doğru, Gokhan, Sergio Ortiz-Rojas, and Adria Martin-Mor. (2017). "MTradumatica: Free Statistical Machine Translation Customisation for Translators." [Poster]. Annual Conference of the European Association for Machine Translation (EAMT). https://ddd.uab.cat/pub/posters/2017/174910/MTradumatica_EAMT_May_2017.pdf

2. Martin-Mor, Adria. (2017). "MTradumatica: Statistical machine translation customisatio n for translators." Skase journal of translation and interpretation, v. 10, n.1 pp. 25-40. http://www.skase.sk/Volumes/JTI12/pdf_doc/02.pdf

Key facts that show that there is room for improvement

Even though HE students are supposed to have a good knowledge of English, they prefer to read academic sources in their mother tongue. Students need to be motivated to access knowledge worldwide, not only in English, but also in other foreign languages. Machine translation may help to introduce multilingual and multicultural academic knowledge to our campuses.

Although our preliminary results show that students mainly use machine translation as a dictionary, there is a legitimate concern among HE institutions of the risk of plagiarism from the use of machine translation, i. e., students copying academic sources translated with a machine translation system, which would be more difficult for an anti-plagiarism system to detect. Awareness and training programs are needed to balance the pros and cons of these systems in class.

HE users could be trained in using and training customized machine translation systems as well as in the use of corpora. The analysis of corpus may help to obtain information about concepts and the underlying sub-concepts. Moreover, the importance of a term in a corpus can help to decide on the pertinence of the proposed translation. Users in HE generally use generic machine translation systems without domain-specific features, and they normally use them as a dictionary. Domain-specific corpora can be used to develop customized machine translation systems such as MTradumatica.

Artificial intelligence techniques and big data are increasingly being used in machine translation. Critical human thinking, however, is still needed to fix machine translation mistakes. Users should not rely blindly on machine translation output and should adopt an active role in detecting mistakes and fixing them if they learn how.

Conclusions

There is a need to empower higher education students to engage in autonomous learning by accessing knowledge from all over the world using machine translation combined with critical thinking. The wise use of machine translation could be integrated into HE competences.

Acknowledgements

Contact details

We encourage other universities and colleagues from all fields to participate in the survey and the project. Please contact projecta@tradumatica.net
www.projecta.tradumatica.net
www.Tradumatica.net
Tradumàtica magazine
https://revistes.uab.cat/tradumatica