MultitraiNMT Erasmus+ project: Machine Translation Training for multilingual citizens (multitrainmt.eu)

Mikel L. Forcada
Universitat d'Alacant
mlf@dlsi.ua.es

Dorothy Kenny
Dublin City University
dorothy.kenny@dcu.ie

Juan Antonio Pérez Ortiz
Universitat d'Alacant
japerez@dlsi.ua.es

Gema Ramírez Sánchez
Prompsit
gramirez@prompsit.com

Caroline Rossi
Université Grenoble-Alpes
caroline.rossi@univ-grenoble-al-pes.fr

Pilar Sánchez-Gijón
Universitat Autònoma de Barcelona
pilar.sanchez.gijon@uab.cat

Felipe Sánchez-Martínez
Universitat d'Alacant
fsanchez@dlsi.ua.es

Riccardo Superbo
KantanAI
riccardos@kantanai.io

Olga Torres-Hostench
Universitat Autònoma de Barcelona
Olga.torres.hostench@uab.cat

Abstract

The MultitraiNMT Erasmus+ project has developed an open innovative syllabus in machine translation, focusing on neural machine translation (NMT) and targeting both language learners and translators. The training materials include an open access coursebook with more than 250 activities and a pedagogical NMT interface called MutNMT that allows users to learn how neural machine translation works. These materials will allow students to develop the technical and ethical skills and competences required to become informed, critical users of machine translation in their own language learning and translation practice. The project started in July 2019 and it will end in July 2022.

1 The project

MultitraiNMT consortium is formed by Universitat Autònoma de Barcelona, Universitat d'Alacant, Université Grenoble-Alpes and Dublin City University, together with Prompsit Language Engineering and Xcelerator Machine Translations.

The project is currently supported by numerous associate partners, all of whom are interested in learning about the use of neural machine translation (NMT), and willing to adjust their teaching practices.

MultitraiNMT invites higher education institutions and teachers of translation and second languages to join the project as associate partners to (i) use the project coursebook and associated activities in their classes; (ii) test the MutNMT educational platform and activities for managing NMT engines for didactic purposes and (iii) participate in any other training and/or research activity which fosters the development of machine translation skills in general.

2 The coursebook

The open access coursebook addresses both the technical foundations of machine translation, and the ethical, societal and professional implications of this approach. It will soon be available from Language Science Press. The coursebook is organized in 9 chapters: (1) Multilingualism. (2) Introduction to machine translation. (3) How to choose a suitable MT system. Evaluation of machine translation quality. (4) How to prepare and select texts for machine translation. (5) How to deal with machine translation mistakes. Post-

editing and error fixing. (6) Ethical aspects of machine translation related to the data workflow, sustainability, diversity and decision-making. (7) How neural machine translation works. (8) Custom neural machine translation and (9) Machine translation and language learning.

3 The course activities

The project includes learning activities related to the coursebook and MutNMT that allow language learners and translators to learn about machine translation in general and especially about NMT. There are two types of activities. On the one hand, there are self-learning activities aimed at students working at their own pace; these are short-answer questions with immediate automatic feedback. On the other hand, there are open-answer teacher-guided activities which can be customized and adapted to different contexts. After exploring different formats and repositories of learning objects, we opted for the open-source H5P platform (h5p.org), as it allows each of our questions and activities to be self-contained and easily embeddable by instructors in learning management systems such as Moodle (including grading) or more general environments such as WordPress. Currently all activities are written only in English, using examples in different languages. However, they can be easily exported and translated as needed. Activities are designed taking into account different progress levels to approach different student profiles. The database of activities is here: https://ddd.uab.cat/record/257869.

4 MutNMT

MutNMT is a web application to train neural machine translation engines for didactic purposes (see logo in Figure 1). Currently teachers and students interested in using the tool may access MutNMT at https://ntradumatica.uab.cat (UAB) and/or http://multitrainmt.univ-grenoble-alpes.fr:5000 (UGA). Access is given with any Gmail account. Besides, the code of the web application is available on GitHub at https://github.com/Prompsit/mutnmt. MutNMT lets the user train, inspect, evaluate and translate using NMT engines. The project has contributed to other free/open-source projects, such as JoeyNMT, a command-line tool to train NMT engines.

In what follows, we provide a brief description of the main features of MutNMT:

Data. MutNMT needs corpora in the form of parallel data to learn from. Previewing, downloading and grabbing corpora is possible as part of the basic options for corpora.

Engines. As well as for data, there is a library of engines in MutNMT, that is, already available machine translation systems that have been trained and shared. Of special interest also are the actions allowed: viewing the full training log of an engine, downloading the model, downloading the corpora used to train the engine, and grabbing or removing the corpora. While beginners can only view, experts and administrators will be able to resume the training of an engine.
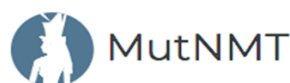

Figure 1: MutNMT logo

Train. This is an advanced feature for experts and administrators, so they can train NMT engines using MutNMT. Users will need to set up engine details, configuration parameters and select corpora for training a particular system.

Translate. All users will be able to copy and paste a series of sentences and translate them using the engines available in the 'Your engines' section. An already trained engine can be used if the user first goes to Engine and selects the *Grab* option from the menu associated with a particular engine. They will get the resulting translation in the text box and will be able to export a standard TMX file with the whole translation.

Inspect. There are several options in this section, all aimed at checking the inside of the translation engines at work. The first one allows users to input a sentence and see its evolution at different steps of processing by a particular engine: preprocessed input, hypothesis generation (n-best), preprocessed output and final output.

Evaluate. As a final step, users will be able to evaluate the machine translated output by comparing it to other machine translated texts or to one or more professional human translations. MutNMT provides popular automatic document- and sentence-level evaluation metrics, as well as an overview of sentences along with their individual scores. Evaluation results can also be downloaded in a spreadsheet.