ATA CHRONICLE

Balancing Act:
Caring for Someone with Dementia While Keeping a Practice Afloat

ata American Translators Association
The Voice of Interpreters and Translators

MultiTraiNMT: Neural Machine Translation for Everyone

Neural machine translation (NMT) is on everyone's mind. Its quality has become stunning, if not frightening, and it continues to improve even as we speak. Tech giants are investing tremendous amounts of capital in NMT applications. Language services providers, big and small, are deploying it in production. And translators are increasingly using it in their workflow. Indeed, post-editing machine translation (MT) has become a default *modus operandi* for many.

Due to its end-to-end computational architecture designed for deep learning, NMT is easier to understand than its predecessor, statistical MT. But hardcore MT research remains a privilege of relatively few. For most users, MT remains a black box with often unpredictable behavior, but thanks to recent efforts aimed at increasing MT literacy[1], the situation has started to change. The MultiTraiNMT project is a major step in this direction.

What Is MultiTraiNMT?

Funded by the Erasmus+ program of the European Union, MultiTraiNMT is a project specifically intended "to develop, evaluate, and disseminate open-access materials and open-source applications that will lead to the enhancement of teaching and learning about MT among language learners, language teachers, trainee translators, translation teachers, and professional translators across Europe."[2] And not only Europe.

Developed within the past three years by a team of experts from the Universitat Autònoma de Barcelona, Universitat d'Alacant, Université Grenoble-Alpes, and Dublin City University, along with Prompsit Language Engineering and KantanMT, MultiTraiNMT invites all interested parties to join it as partners to:

1. "Use the project coursebook and associated activities in their classes."

2. "Test the MutNMT educational platform and activities for managing NMT engines for didactic purposes."

3. "Participate in any other training and/or research activity which fosters the development of MT skills in general."[3]

Three interrelated components of the project are briefly described below.

The Book

Released in July 2022, the open-access coursebook, *Machine Translation for Everyone* (see links in the sidebar on page 39), covers much ground—from the technical foundations to the ethical and broadly societal implications of MT. While explicitly intended for classroom use, the book's nine chapters, written by experts in the relevant fields, are remarkably clear and accessible to everyone. Each chapter can be read on its own and is complete with ample references to more specialized literature.

The Activities

There are two types of activities developed for each chapter of the coursebook:

1. Self-learning questions ranging from multiple-choice to crossword puzzles and fill-in-the-blank exercises (see Figure 1 on page 37), with immediate automatic feedback for those learning at their own pace.

2. Open-ended, customizable teacher-guided mini-projects that invite readers to reflect on many interesting and challenging issues surrounding MT and write short essays. (See Figure 2 on page 37.)

There are currently over 200 excellent and thoroughly prepared activities, and the authors deserve much praise for putting them together with such care to detail, using the open-source H5P platform, which allows users to integrate them into learning management systems such as Drupal or Moodle and publishing environments such as WordPress. Translation instructors may further adapt the activities to their needs. And it's a great self-test, too: if you can answer most of the questions correctly, you probably know a lot about MT!

To appreciate this point, browse through the questions. You'll be quizzed on a broad range of topics—from the basics of neural networks to the famous semantic alchemy of word embeddings, to MT evaluation metrics such as BLEU and TER, to the opportunities and challenges of adapting a particular MT engine to a given task, and of using MT in second-language learning. If you discover significant gaps in your background, read the book! It has all the answers and is a very rewarding read even if you're already familiar with this material. Among other things, it tries to offer a unified perspective on a field that has become very mosaic.

MutNMT

Deriving its name from Mut, the mother goddess of ancient Egypt, MutNMT is a web application that allows you to get under the

For most users, MT remains a black box with often unpredictable behavior, but thanks to recent efforts aimed at increasing MT literacy, the situation has started to change.

Figure 1: A fill-in-the-blank exercise for Chapter 5: "How to Deal with Errors in Machine Translation: Post-Editing"

Figure 2: A short essay assignment for Chapter 6: "Ethics and Machine Translation"

hood of MT without any coding! Anyone with a Google account can access five out of the seven features of the application: Data, Engines, Translate, Inspect, and Evaluate. (See Figure 3 on page 38.) Let's take a look at each of these features.

Data: A rapidly expanding collection of parallel corpora already uploaded to the system by expert users. Some corpora have millions of sentence pairs. These are used to train NMT engines. Any user can "grab" an available corpus and add it to their individual collection ("Your Corpora"). The corpora can also be previewed and downloaded as a zipped archive of two parallel text files.

Engines: Offers a growing list of NMT models trained by expert users on the available corpora. Again,

users can "grab" any engine and add it to their individual collection ("Your Engines") for translation and inspection. You can also view the training log of a given engine and learn a ton of useful information. The corpora and engines that are no longer needed can be removed from the individual collections.

Translate: This is where you can choose an engine from your individual collection to translate a sentence or a small file. This may take some time. Important: don't expect DeepL quality! Rather, come to appreciate the amazing fact that a neural model trained entirely from scratch on a relatively small

corpus[4] with a simple toolkit[5], for just one hour of graphics processing unit (GPU) time, for primarily didactic purposes can often produce a reasonable translation—and in such a transparent way!

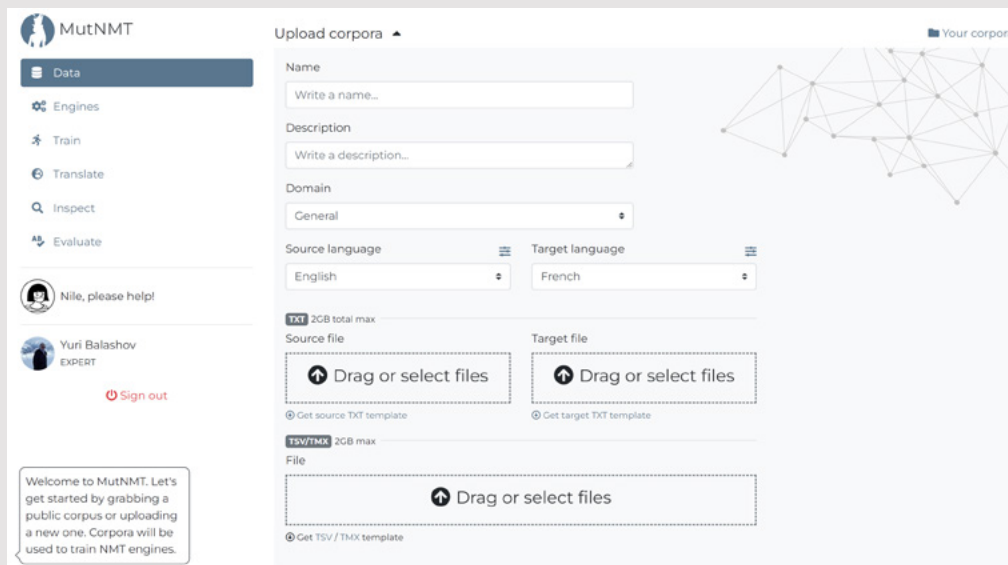**Inspect:** Allows you to gain more insight into what happens when the
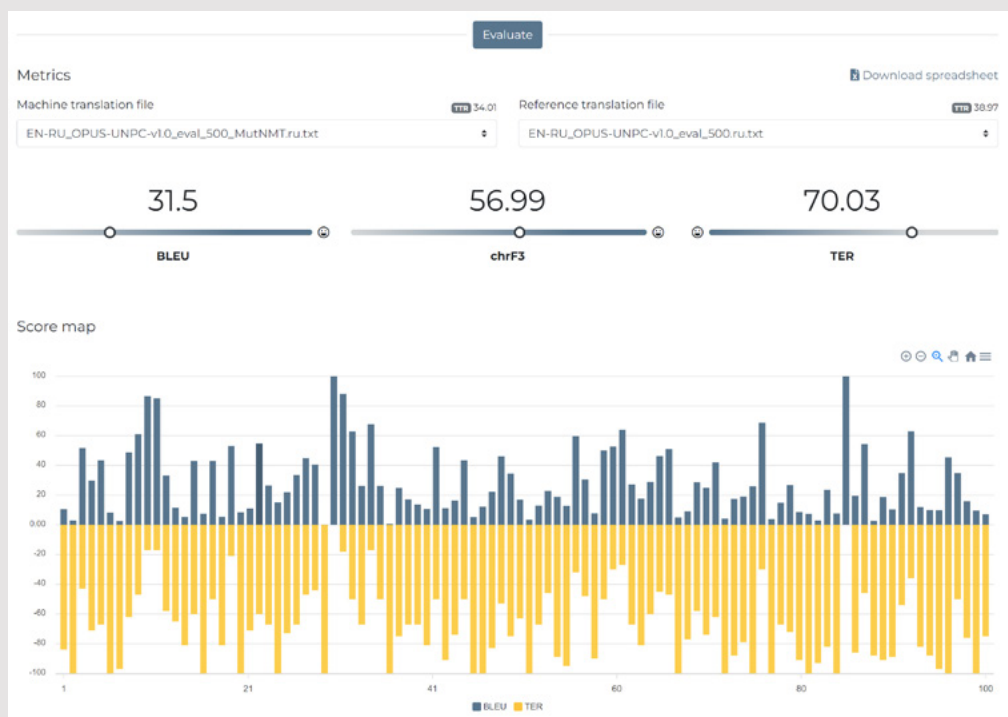
Figure 3: MutNMT's interface

Figure 4: MT evaluation scores and map generated by MutNMT

"Translate" button is pressed. The system starts by "tokenizing" the input sentence (splitting it into words, punctuation marks, and sometimes subword segments). The engine then produces "N-Best" candidate translations, from which the most probable one is selected. These steps are visualized for your attention and learning. You can also compare the output of several selected engines for a given language pair.

Evaluate: Computes several popular metrics (e.g., BLEU, chrF3, and TER) by comparing the output of a chosen engine with a reference translation produced, hopefully, by a professional human translator. You need to upload a source file (up to 500 sentences, in plain text format, one sentence per line) as well as the MutNMT output and reference files, which must be perfectly aligned with the source. Please note that this test set should not be used for training the engine! In addition to the document-level scores, MutNMT generates a sentence-by-sentence BLEU/TER "score map" for the first 100 test sentences. (See Figure 4.) You can display each of them to see what may be wrong with the MT output. As a bonus, you could use the Evaluate feature to score any MT output (e.g., from Google Translate, ModernMT, or your own custom engine) to get an almost scientific sense of its quality—just by uploading three text files and pressing the "Evaluate" button.

Uploading Corpora and Training Engines (for More Advanced Users)

The five features of MutNMT discussed here allow anyone to open the "black box" of NMT and get inside it. Those who feel comfortable with it and are prepared to do more work can request "Expert" status to be able to upload new corpora and train new engines. This is very exciting, but also time- and resource-consuming. There are lots of multilingual public corpora available in different formats, including TMX and parallel text files (e.g., see the OPUS site in the sidebar). And if you have a good translation memory with >100K units, you could try to train an engine on it.[6] Corpora for a given language pair can be combined for training, for a total of 500K sentence pairs. In addition, you'll need to create smaller separate corpora (3-5K sentence pairs) for "Validation" and "Testing." I suggest adding one more for "Evaluation" (500 sentence pairs).

Assuming these data don't overlap with the training set or among themselves, you'll be fully equipped for the entire process. In MT research and development, it's standard practice to produce validation and testing data by splitting them off from the large training corpus. But some public corpora are highly repetitive, so you would need to ensure that there's no overlap among the resulting subsets, otherwise you may get inflated scores but poor quality. In any case, the corpora must be fully aligned, cleaned, and otherwise pre-processed to be used with MutNMT.

"Expert" users may be further promoted to "Admin" status if they decide to use MutNMT in teaching or otherwise partner with MultiTraiNMT in an official capacity. For further tips, please read the materials referenced in the notes section and the sidebar and watch very helpful videos on the MultiTraiNMT YouTube channel.

Unpacking the Black Box

The best way to learn MT is to unpack its black box. It's becoming increasingly possible thanks to efforts like MultiTraiNMT. Getting under the hood of NMT is very empowering! ⒶⓉⒶ

For More Information

Machine Translation Training for Multilingual Citizens
MultiTraiNMT's home page.

MutNMT
A web application to train NMT engines for didactic purposes.

MutNMT: Basic and Advanced Features
Description and instructions for MutNMT.

MultiTraiNMT Erasmus Project
MultiTraiNMT's YouTube Channel

Kenny, Dorothy, editor. *Machine Translation for Everyone* (Language Science Press, 2022). An open-access coursebook released as part of the MultiTraiNMT project.

Learning Activity Explorer
From MultiTraiNMT: features over 200 activities for MT learners.

H5P
A plugin for learning and publishing systems used in the MultiTraiNMT activities.

OPUS
A growing collection of open multilingual corpora.

NOTES

1. Such as the Machine Translation Literacy initiative led by Lynne Bowker. (Be sure to check out the Twitter page as well!)

2. Ramírez-Sánchez, Gema, et al. "MultiTraiNMT: Training Materials to Approach Neural Machine Translation from Scratch." *Translation and Interpreting Technology Online* (July 2021).

3. Forcada, Mikel L., et al. "MultiTraiNMT Erasmus+ Project: Machine Translation Training for Multilingual Citizens." *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation* (2022), 291-292.

4. Up to 0.5M sentences, which is the limit set by the developers. For comparison, corpora used to train commercial MT engines may have 100M+ sentence pairs.

5. MutNMT is based on JoeyNMT, an educational NMT framework with a simplistic architecture and many inherent limitations.

6. But bear in mind that "Sharing" it will make it public. Even if you keep it in your individual collection, it's a good idea to double check with the developers on the confidentiality of the uploaded data.

Yuri Balashov, CT is a professor of philosophy and a faculty fellow in the Institute for Artificial Intelligence at the University of Georgia. He is also an ATA-certified English>Russian translator. He is working on a project exploring the cognitive, linguistic, and philosophical dimensions of human and machine translation. *balashov.yuri@gmail.com*

If you have any ideas and/or suggestions regarding helpful resources or tools you would like to see featured, please e-mail Jost Zetzsche at *jzetzsche@internationalwriters.com*.