

# ProMut: The evolution of NMT didactic tools

**Pilar Sánchez-Gijón**

Universitat Autònoma de Barcelona  
pilar.sanchez.gijon@uab.cat

**Gema Ramírez-Sánchez**

Prompsit Language Engineering  
gramirez@prompsit.com

## Abstract

Neural Machine Translation intensifies educational challenges in translation technologies. The MultiTraiNMT project, focused on the creation of open access materials for teaching and learning about machine translation, developed MutNMT, an open-source, didactic platform for training and evaluating NMT systems. Building upon it, the LT-LiDER project introduces the ProMut platform, which implements three main novel features: migration of the core NMT framework from JoeyNMT to MarianNMT, close integration with OPUS datasets, engines and connectors and the addition of a researcher profile for larger datasets and extended training processes and evaluation.

## 1 Introduction

The integration of language and translation technologies into the education of future professionals has consistently posed significant challenges since these technologies first emerged. The advent of machine translation—especially with the rise of neural machine translation (NMT) systems—has added further layers of complexity, impacting both the training of students and the skill development of educators. These challenges continue to be highly relevant today.

Machine translation has become a cornerstone in the translation industry, driven by ongoing advancements that introduce new functionalities and refine existing systems. Consequently, having access to up-to-date, well-maintained tools is crucial. Such tools must not only clarify the workflows and tasks involved in incorporating neural machine translation into professional practice but also encourage meaningful interactions that enhance users' understanding of these systems.

## 2 Background

Within the broader context of rapid NMT adoption, the Machine Translation Training for Multilingual

Citizens project<sup>1</sup> was conceived to provide the resources needed for both trainers and students to effectively learn about and operate NMT systems. A central achievement of this project was the development of MutNMT, a platform enabling the management and creation of NMT engines. MutNMT also integrated features for evaluating translation quality, enhancing its applicability as both a training and research tool. As open-source software, MutNMT is freely accessible on GitHub<sup>2</sup> alongside comprehensive documentation. Additionally, it is hosted on the servers of the Autonomous University of Barcelona (UAB), where it supports a community of over 700 registered users.

MutNMT allows users to upload data corpora for training engines in any language combination. Based on the Joey NMT framework (Kreutzer et al., 2019), it was developed specifically for pedagogical purposes, with limitations on the volume of training data and the complexity of the training process. Once an engine is trained, it can be used directly within the platform for translation tasks. Users can also leverage various metrics to evaluate translation quality or to compare results with other MT systems. Building on this foundation, the LT-LiDER project<sup>3</sup> expanded the focus to include the development of digital literacy competencies among professionals, trainers, and trainees in translation and multilingual communication. One of its outputs is ProMut, a didactic tool aimed at enabling advanced engagement with and management of NMT systems. ProMut offers functionalities for creating engines and evaluating translation quality, while also broadening the system's capabilities and contexts of use. By deepening technical understanding and diversifying application scenarios, ProMut stands as a robust resource for education and training in translation technologies.

<sup>1</sup><http://multitrain.eu>

<sup>2</sup><https://github.com/Prompsit/mutnmt>

<sup>3</sup><http://lt-lider.eu>

### 3 From MutNMT to ProMut

MutNMT is currently an application focused on teaching users of translation technologies the essentials about data management, training, usage and evaluation of machine translation systems. The online application implements different profiles with different rights regarding the ability to train a new engine within the tool. Beginners are not allowed to do it, but they can interact with the rest of the functionality of the tool (see and upload datasets, see engine’s training info, translate, inspect, evaluate). Training, though, is only allowed to Expert and Admin profiles and is limited to the following parameters:

- 1-hour of training time slots with the possibility to stop or continue training for 1 more hour successively.
- A maximum of 500k sentence pairs in the training set, 5k for validation and test.
- Training parameters with default values.
- Use of engines only inside the application by lack of integration with external tools.

In order to open up the use of MutNMT to a wider range of usages related with MT research, the LT-LiDER project generously expands the aforementioned limits through ProMut:

- ProMut includes a new profile, the Researcher profile, for which the limits in training time and data sizes are extended along with more flexibility in choosing training parameters.
- Corpus management allows the upload of large corpora and connection to the OPUS (Tiedemann, 2012) parallel data repository.
- ProMut implements the ability to download pre-trained engines from OPUS-MT-train<sup>4</sup> and the possibility to fine-tune these or others already trained within the application.
- To this end, a replacement of the core engine from JoeyNMT to MarianNMT (Junczys-Dowmunt et al., 2018) was implemented to enable compatibility with OPUS-MT models and, at the same time, enable compatibility with OPUS-CAT,<sup>5</sup> allowing to connect ProMut engines to a variety of computer-assisted translation (CAT) tools.

- In evaluation, COMET (Rei et al., 2022) has been added as a new evaluation metric that complements the previous n-gram and character-based metrics.
- ProMut also provides updates to previously-available functionalities (e.g. evaluation and training detailed plots and histograms) as well as technical and user documentation.

ProMut, currently in internal testing by the LT-Lider partners, will be available by May 2025 along with the code under a free/open-source licence.

### Acknowledgments

This project has been funded with support from the European Commission. We thank Paula Guerrero Castelló for the contributions, ideas, and comments to the current MutNMT and ProMut frameworks.

### References

- Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, and Alexandra Birch. 2018. *Marian: Fast neural machine translation in C++*. In *Proceedings of ACL 2018, System Demonstrations*, Melbourne, Australia.
- Julia Kreutzer, Jasmijn Bastings, and Stefan Riezler. 2019. *Joey NMT: A minimalist NMT toolkit for novices*. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 109–114, Hong Kong, China. Association for Computational Linguistics.
- Ricardo Rei, José G. C. de Souza, Duarte Alves, Chrysoula Zerva, Ana C Farinha, Taisiya Glushkova, Alon Lavie, Luisa Coheur, and André F. T. Martins. 2022. *COMET-22: Unbabel-IST 2022 submission for the metrics shared task*. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 578–585, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.
- Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey. European Language Resources Association (ELRA).

<sup>4</sup><https://github.com/Helsinki-NLP/Opus-MT-train>

<sup>5</sup><https://github.com/Helsinki-NLP/OPUS-CAT>