

[Click here to be redirected to the virtual room of the Project Expo.](#)

Project abbreviation: NTEU

Project name: Neural Translation for the European Union

Project coordinator: Manuel Herranz (Pangeanic)

Project consortium: Pangeanic, Tilde, KantanMT and SEDIA.

Funding: 1,9M€

Project duration: From September 2019 until August 2021.

Main key words: Neural Machine Translation, Deep Learning, NLP.



Background of the research topic: Normally, data for translation are available in English from or to another language. With a few exceptions, all eTranslation MT engines include English as either source or target. Thus, to translate between two non-English languages, English must be used as a pivot. NTEU provides a capacity service to eTranslation by building a near-human-professional-quality neural engine farm which includes all EU language combinations.

Goal of the project: building direct machine translation engines between any of the 24 EU official languages (e.g. Spanish to German, Croatian to Italian, Greek to Polish, etc.) without pivoting through English (around 550 translation engines in total).

Project abstract: The NTEU project aims to build a neural engine farm with all the 24 European official language combinations for eTranslation, without the necessity to use a high-resourced language as a pivot (around 550 translation engines in total). NTEU provides a capacity service to eTranslation by building a near-human-professional-quality neural engine farm which includes all EU language combinations. Lower-resourced languages have been a challenge. We have experimented with techniques to supplement the original data, such as generating synthetic data by doing back-translation and transfer learning. In addition, the NTEU consortium gathers and cleanses data from all language combinations so that the engines can be retrained with other technologies in the future. The results of the MT have been manually evaluated following industry and WMT practices in an open-source platform created by the consortium. <https://nteu.eu>

Publications:

[Neural Translation for European Union \(NTEU\)](#). Mercedes García-Martínez, Laurent Bié, Aleix Cerdà, Amando Estela, Manuel Herranz, Rihards Krišlauks, Maite Melero, Tony O’Dowd, Sinéad O’Gorman, Mārcis Pinnis, Artūrs Stafanovič, Riccardo Superbo, Artūrs Vasiļevskis. Proceedings of Machine Translation Summit XVIII: Users and Providers Track. August 2021.

[Neural Translation for the European Union \(NTEU\) Project](#). Laurent Bié, Aleix Cerdà-i-Cucó, Hans Degroote, Amando Estela, Mercedes García-Martínez, Manuel Herranz, Alejandro Kohan, Maite Melero, Tony O’Dowd, Sinéad O’Gorman, Mārcis Pinnis, Roberts Rozis, Riccardo Superbo, Artūrs Vasiļevskis. Proceedings of the 22nd Annual Conference of the European Association for Machine Translation. November 2020.