

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

Project abbreviation: OPUS-MT

Project name: Open Translation Models, Tools and Services

Project coordinator: Jörg Tiedemann

Project consortium: University of Helsinki

Funding: ELG pilot project

Project duration: August 2020 – August 2021

Main key words: machine translation (MT), neural machine translation (NMT), computer-aided translation (CAT), translation services, low resource languages

Background of the research topic: Machine translation is an important application that connects people with different backgrounds and makes information accessible across language boundaries. Automatic translation is also one of the most prominent research topics in language technology. However, most models and tools are unavailable for research and development and translation services are dominated by commercial providers. It is important to release translation models to reduce dependence on commercial products, increase transparency and drastically increase reuse, replicability and sustainability of machine translation approaches.

Goal of the project: The project aims at the development of open models and services for machine translation to be used in research and development. The main focus is set on Nordic European languages and European minority languages. We also want to support the workflow of professional translators with a seamless integration of MT in their production environment.

Project abstract: OPUS-MT will produce state-of-the-art neural machine translation models that can freely be shared, re-used and integrated in open web services and professional translation workflows. The project will focus on European minority languages and their improved support through multilingual NMT models and transfer learning. Furthermore, OPUS-MT will deliver easily deployable translation services and tools for quick domain-adaptation and on-demand personalisation. We will emphasise open resources that can freely be distributed and used in research and professional applications. For the latter we want to offer local solutions that are independent of on-line services to avoid security risks with open data transfer.

Publications:

[Boosting Neural Machine Translation from Finnish to Northern Sámi with Rule-Based Backtranslation](#), Aulamo, M., Virpioja, S., Scherrer, Y. & Tiedemann, J., May 2021, *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*. Dobnik, S. & Øvrelid, L. (eds.). Linköping: [Linköping University Electronic Press](#), p. 351-356 6 p. (Linköping Electronic Conference Proceedings ; no. 78)(NEALT Proceedings Series ; no. 45).

[OPUS-MT -- Building open translation services for the World](#), Tiedemann, J. & Thottingal, S., 1 Nov 2020, *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. Martins [et al.], A. (ed.). Geneva: [European Association for Machine Translation](#), p. 479-480 2 p.

[The Tatoeba Translation Challenge - Realistic Data Sets for Low Resource and Multilingual MT](#), Tiedemann, J., 1 Nov 2020, *Proceedings of the Fifth Conference on Machine Translation*. Barrault [et al.], L. (ed.). Stroudsburg: [The Association for Computational Linguistics](#), p. 1174-1182 9 p.



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

The Helsinki submission to the AmericasNLP shared task, Vázquez, R., Scherrer, Y., Virpioja, S. & Tiedemann, J., 1 Jun 2021, *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*. Mager [et al.], M. (ed.). Stroudsburg: [The Association for Computational Linguistics](#), p. 255-264 10 p.