

Modern MT: A New Open-Source Machine Translation Platform for the Translation Industry

U. GERMANN¹, E. BARBU², L. BENTIVOGLI³, N. BERTOLDI³,
N. BOGOYCHEV¹, C. BUCK¹, D. CAROSELLI², L. CARVALHO⁴,
A. CATTELAN², R. CATTONI³, M. CETTOLO³, M. FEDERICO³,
B. HADDOW¹, D. MADL¹, L. MASTROSTEFANO², P. MATHUR³,
A. RUOPP⁴, A. SAMIOTOU⁴, V. SUDHARSHAN⁴,
M. TROMBETTI², J. van der MEER⁴

¹ University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom

² Translated srl, Via Nepal, 29, 00144 Rome, Italy

³ Fondazione Bruno Kessler, Via Sommarive, 18, 38123 Povo, Italy

⁴ TAUS B.V., Oostende 9, 1483 AB De Rijp, Netherlands

ugermann@inf.ed.ac.uk

Abstract. *Modern MT* (www.modernmt.eu) is a three-year Horizon 2020 *innovation action* (2015–2017) to develop new open-source machine translation technology for use in translation production environments, both fully automatic and as a back-end in interactive post-editing scenarios. Led by Translated srl, the project consortium also includes the Fondazione Bruno Kessler (FBK), the University of Edinburgh, and TAUS B.V. *Modern MT* has received funding from the *European Union's Horizon 2020 research and innovation programme* under Grant Agreement No645487 (call ICT-17-2014).

Project Description

Modern MT aims to improve the state of the art in open source machine translation software by developing cloud-ready software that offers

- A **simple installation** procedure for a ready-to-go, REST-based translation service.
- **Very fast set-up times** for systems built from scratch using existing parallel corpora (e.g., translation memories). The goal is to process incoming data at approximately the speed at which it is uploaded.
- **Immediate integration of new data** (e.g., from newly post-edited MT output). Rebuilding or retuning the system will not be necessary.
- **Instant domain adaptation** by considering translation context beyond the individual sentence, without the need for domain-specific custom engines.
- **High scalability** with respect to throughput, concurrent users, and the amount of data the system can handle.

A first version of the software is available at <https://github.com/ModernMT/MMT>.

Modern MT is also actively **collecting and curating parallel data** for internal use and public release from web crawls and contributions from translation stakeholders, to improve MT quality for everyone.