EUROPEAN
ASSOCIATION
FOR MACHINE
TRANSLATION

EAMT**2018**

Proceedings of the
21st Annual Conference of
the European Association
for Machine Translation

28–30 May 2018
Universitat d'Alacant
Alacant, Spain

Edited by
Juan Antonio Pérez-Ortiz
Felipe Sánchez-Martínez
Miquel Esplà-Gomis
Maja Popović
Celia Rico
André Martins
Joachim Van den Bogaert
Mikel L. Forcada

Organised by

Universitat d'Alacant
Universidad de Alicante

transducens
research group

Use of NMT in Ubiqus Group

Paloma Valenciano
Translator / Posteditor
General Manager
Traducciones Políglota
`paloma.valenciano@poliglota.com`

Abstract

After more than 30 years' experience as a translator and as a reviser, I have recently started to post-edit. During these 10 months discovering a new approach to my profession, the experience has been highly positive.

Ubiqus, the French group to which we belong, has developed 20 engines based on OpenNMT. OpenNMT derives from an academic project initiated in 2016 by Harvard NLP; Systran joined the project and an open source toolkit was released in January 2017. The community grew when individuals as well as localization professionals contributed. Ubiqus adopted this toolkit at the very beginning of 2017 and contributed to its development as well as with some extensions, developing a layer to integrate OpenNMT in our workflow environments, including SDL Studio and with our internal ERP, which enables to provide a highly efficient end-to-end system.

I have been using the EN-ES and FR-ES engines mainly for legal texts. I very soon felt comfortable with the task, I started measuring my productivity by timing my output. I was surprised by the improvement since the very beginning, and as the NMT engine was further trained and I got more used to the post-editing task I achieved even better results, improving productivity by almost 30%.
Ubiqus has also developed a scheme for the systematic scoring of all translation jobs, U-Score, a composite indicator of the overall performance of the machine. The U-Score is obtained by aggregating the in-

formation of BLEU, TER and DL-ratio and averaging them. It then performs a transformation allowing to spread the scale a bit. The scores have been clearly improving in the last months with a constant training of the engines.
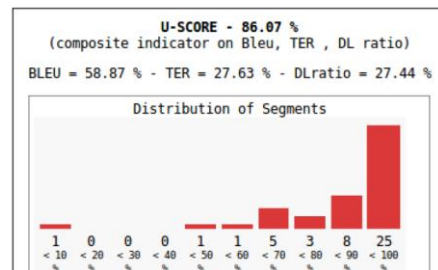

Figure 1: U-Score

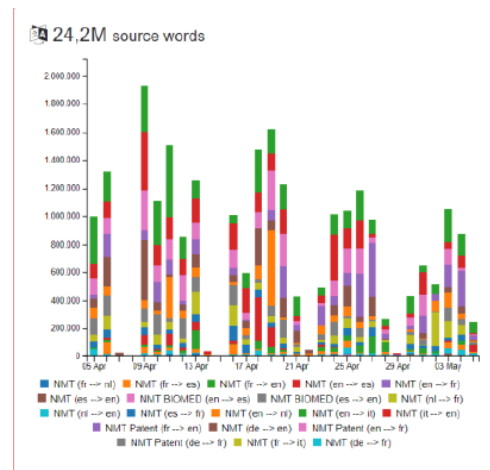Over the last 30 days, 24.2 million source words have been postedited within the Ubiqus Group using the 20 engines, which are constantly retrained.

Figure 2: No. of words postedited, April 2018