

Word from the editors

Constantin Orăsan¹ · Marcello Federico²

Published online: 9 February 2017
© Springer Science+Business Media Dordrecht 2017

Translation Memories (TMs) are widely used by professional translators in their day-to-day activities. The underlying idea of TMs is that a translator should benefit as much as possible from previous translations. To achieve this, whenever a segment is to be translated, identical or similar segments are first sought in a database of previous translations. These translations could be either supplied by the client, made by the translator, or both. Hence, by using TMs, translators need to translate fewer segments from scratch, and in many cases, just apply a small amount of post-editing to the retrieved suggestions. The use of TMs leads, in general, to an increase in productivity and ensures that the translation complies better with the client's specified style and terminology.

Despite the fact that the core idea of TMs is about comparing segments to be translated with segments of previous translations, most of the existing implementations hardly use any natural language processing (NLP) for this purpose. Most of them rely on simple forms of edit distance, which have the advantage of being computationally efficient, but fail in situations where two segments use different wording to express the same meaning. Companies who develop TMs have mainly focused on improving the user experience by allowing the processing of a variety of document formats, and on developing intuitive user interfaces, but paid little attention to how methods from NLP could improve the way TMs work.

✉ Constantin Orăsan
c.orasan@wlv.ac.uk

Marcello Federico
federico@fbk.eu

¹ RIILP, University of Wolverhampton, Wolverhampton, UK

² FBK Trento, Trento, Italy

This double special issue on Natural Language Processing for Translation Memories is the result of the increasing interest of NLP researchers to develop methods that are directly applicable to TMs. Two concrete examples of this are the workshops on this topic that were organised in 2015¹ and 2016.² In addition to bringing together people active in the field, the workshops gave some of the researchers featured in this special issue the opportunity to present the preliminary results of their research. This volume publishes the first four papers of the special issue, which are briefly introduced in the following paragraphs.

As mentioned above, integral to TMs is the similarity metric employed to identify similar segments. The paper *Self-selection Bias of Similarity Metrics in Translation Memory Evaluation* by Wolff et al. explores the impact of using different similarity metrics for evaluating TMs and how TM evaluation metrics are biased towards certain similarity metrics. Recommendations about how to evaluate TMs are also proposed.

One of the assumptions when using TMs is that the translations retrieved from them are of high quality. This is not always the case given the advent of crowdsourced translations and the use of automatic methods to build TMs from the Web. The paper *The First Automatic Translation Memory Cleaning Shared Task* by Barbu et al., as its name suggests, reports on the organisation and results of a shared task on the cleaning of TMs organised in conjunction with the NLP4TM 2016 workshop. In addition to the overview of the shared task, the paper analyses two surveys collected after the task in an attempt to better understand what constitutes a good TM unit for translators.

The paper *Combining Off-the-shelf Components to clean a Translation Memory* by Wolff describes a system that participated in the Automatic Translation Memory Cleaning shared task. The manuscript presents a machine learning system that combines existing NLP components to assess the quality of TMs.

The final paper in this issue, *Combining Translation Memories and Statistical Machine Translation Using Sparse Features* by Li et al., proposes a method that softly combines TM and MT. The paper proposes a method which extracts TM features at run-time and adds them directly into the MT decoder. The evaluation presented in the paper shows that by using TM-derived features it is possible to enhance both phrase-based and syntax-based MT systems.

In conclusion, we would like to warmly thank all the authors who submitted papers to our special issue. We are also grateful to the reviewers who spent time over several rounds of revisions to provide useful feedback and suggestions to the authors.

¹ 1st Workshop on Natural Language Processing for Translation Memories (NLP4TM): <http://rgcl.wlv.ac.uk/nlp4tm/>.

² 2nd Workshop on Natural Language Processing for Translation Memories (NLP4TM 2016): <http://rgcl.wlv.ac.uk/nlp4tm2016/>.