CrossMark

Word from the editors

Constantin Orăsan[1] · Marcello Federico[2]

Translation Memories (TM) play an important role in the translation workflow of many professional translators. 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. If a match is found, the translator is asked to validate or post-edit the retrieved translation. The use of TMs, in general, leads to an increase in translators' productivity and more consistent translations. A recent survey about translators' needs regarding electronic tools revealed that 90% of them are familiar with TMs and that over 75% of them are using TMs to translate documents (Zaretskaya et al. 2018).

Surprisingly most existing TM tools hardly rely at all on natural language processing (NLP) to help translators. Companies who develop TM tools usually focus their efforts on improving the user experience by allowing the processing of a variety of document formats and on developing intuitive user interfaces. Unfortunately, they pay little attention to how they could enhance these tools with methods from NLP. Most TM tools still rely on simple forms of edit distance for matching a segment to be translated to segments already in the database. This approach has the advantage of being computationally efficient. Unfortunately, it fails when required to identify two segments which might have the same meaning despite using different words. A semantically enhanced edit-distance method that can identify such situations with the

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

  Marcello Federico
  federico@fbk.eu

[1] MC139, RIILP, University of Wolverhampton, Wulfruna St, Wolverhampton WV1 1LY, UK

[2] c/o FBK HLT-MT, Via Sommarive, 18, Povo, 38123 Trento, Italy

⁂ Springer

help of a paraphrase database was proposed by Gupta et al. (2016), but not integrated in a professional translation tool.

This double special issue on Natural Language Processing for TM is partially a result of two workshops organised in 2015[1] and 2016[2] dedicated to this topic. It also reflects the increasing interest of NLP researchers to develop methods that are directly applicable to TMs.

The core idea of TMs is comparing segments to be translated with segments of previous translations. As most of the existing implementations hardly use any NLP for this purpose, we expected to receive a number of submissions attempting to enhance the existing matching and retrieval methods with methods from language processing. Interestingly enough, the topic addressed by most of the papers submitted is that of cleaning of TMs and reflects the need of resources in the translation process. These papers are a direct result of the the 1st Automatic Translation Memories Cleaning Shared Task[3] which was organised in conjunction with the 2nd Workshop on Natural Language Processing for Translation Memories (NLP4TM 2016). The first part of the special issue presented an overview of the shared task and lessons learnt (Barbu et al. 2016) together with one of the participating systems (Wolff 2016) which combined existing NLP components to assess the quality of TMs.

The first paper in this issue, *Automatic translation memory cleaning* by Negri at al., proposes supervised and unsupervised machine-learning approaches for cleaning of TMs. The supervised method obtains very good results on the data set released by the 1st Automatic Translation Memories Cleaning Shared Task, whilst evaluation of the unsupervised method shows its feasibility in cases where there is no training data available.

The paper *Improving retrieval performance of translation memories using morphosyntactic analyses and generalized suffix arrays* by Weitz addresses the problem of improving the retrieval from TMs by analysing the (morpho)syntactic structure of segments and identifying the longest common substring between two segments by means of generalised suffix arrays. Evaluation on German–English datasets reveals an increase in the precision of the retrieval, but a small reduction in recall. However, this reduction can always be addressed by requesting more segments from the database. The proposed method is very fast which means it can be used in real translation scenarios without any negative impact on the productivity of translators.

The final paper, *A system for terminology extraction and translation equivalent detection in real time: Efficient use of statistical machine translation phrase tables* by Oliver, presents a system for automatic terminology extraction and automatic detection of the equivalent terms in target language. The method relies on the output of Moses, a well-known statistical machine translation tool, and was evaluated using OmegaT, an open-source translation memory tool.

---

[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/.

[3] http://rgcl.wlv.ac.uk/nlp4tm2016/shared-task/.

This double special issue would not have been possible without the help of the members of the guest editorial board. During the reviewing process they provided very valuable feedback to authors and helped us to select the most appropriate papers for the issue. The board members listed in alphabetical order are: Wilker Aziz, Bogdan Babych, Eduard Barbu, Maud Ehrmann, Kevin Flanagan, Mikel Forcada, Gabriela Gonzalez, Meritxell Gonzalez, Rohit Gupta, Maxim Khalilov, László Laki, Qun Liu, Yanjun Ma, Matteo Negri, Carla Parra Escartín, Michael Paul, Uwe Reinke, Germán Sanchis-Trilles, Christophe Servan, Liling Tan, Marco Turchi, Mihaela Vela, Friedel Wolff, Marcos Zampieri and Jian Zhang.

Last but not least, we would like to warmly thank all the authors who submitted papers to our special issue. Without them this special issue would have not been possible.

References

Barbu E, Parra Escartín C, Bentivogli L, Negri M, Turchi M, Orasan C, Federico M (2016) The first automatic translation memory cleaning shared task. Mach Transl 30(3–4):145–166. https://doi.org/10.1007/s10590-016-9183-x

Gupta R, Orăsan C, Zampieri M, Vela M, van Genabith J, Mitkov R (2016) Improving translation memory matching and retrieval using paraphrases. Mach Transl 30(1):19–40. https://doi.org/10.1007/s10590-016-9180-0

Wolff F (2016) Combining off-the-shelf components to clean a translation memory. Mach Transl 30(3–4):167–181. https://doi.org/10.1007/s10590-016-9186-7

Zaretskaya A, Corpas Pastor G, Seghiri M (2018) User perspective on translation tools: findings of a user survey. In: Duran I, Corpas Pastor G (eds) Trends in e-tools and resources for translators and interpreters. Brill, Leiden