

Natural Language Processing tools and resources for translation and interpreting applications

Introduction

Raquel Lázaro Gutiérrez

Universidad de Alcalá
raquel.lazaro@uah.es

In the last two decades we have been able to witness a technological turn in translation and interpreting studies with natural language processing (NLP) and deep learning playing more and more prominent part. There is a growing number of NLP applications which are used to support the work of translators and interpreters. The emergence of highly successful deep learning models resulted in very promising Neural Machine Translation (NMT) performance.

Jiménez Crespo (2021) reflects on the reality and discernability of a disciplinary turn in translation both as a profession and a field of research. After reviewing the concepts of “turn” and “technological turn” as defined, amongst others, by well-known translation scholars, such as Sin-Wai (2004), Cronin (2010) or O’Hagan (2013), Jiménez Crespo describes this phenomenon as “a process by which translation theories begin to incorporate the increasingly evident impact of technology, developing theoretical tools and frameworks for translation studies and related disciplines”. Human-computer interaction is nowadays a common practice and situation in professional translation (O’Brien 2012). Corpus analysis, terminology management, computer assisted translation tools, machine translation, translation project management software are at the core of the profession.

The translation technology revolution has transformed the translation profession and nowadays most professional translators employ tools such as translation memory (TM) systems in their daily work. Latest advances of Neural Machine Translation (NMT) has resulted in NMT not only becoming an integral part of most state-

of-the-art TM tools but also typical for the translation workflow of many companies and organisations.

Although translation has benefited more from technological advances, interpreting has also experienced a technological turn. Fantinuoli (2018) points out that technology has been present in professional interpretation since the beginning of simultaneous interpreting systems in 1920s. Technology mediated interpreting has also been popular in dialogue settings, and telephone interpreting dates back to the 1950s (Braun, 2015; Cabrera Méndez, 2016). However, it has not been until some years ago that soft technology has permeated interpreting practice and research. Computer assisted translation, MT and NLP tools have been adapted to be used by interpreters. One of the most important related projects is VIP (Corpas Pastor, 2021), a platform that integrates several CAI tools (terminology management, speech-to-text, note-taking).

Shlesinger (1998) already mentioned several decades ago the benefits of using corpus-based methodologies in interpreting studies, particularly to obtain information about lexical, grammatical or discursive patterns. Authors such as Van Besien (1999), Takagi et al. (2002) or Ryu et al. (2003) pioneered corpus-based studies on simultaneous conference interpreting, focusing on interpreting techniques, time span or contrastive linguistic features respectively.

More recently, corpus-based studies have reached dialogue interpreting. For instance, the ComInDat Pilot Corpus (Angermeyer, Meyer and Schmidt, 2012) comprised two subcorpora of interpreter-mediated medical interviews and court trials.

More recent are the corpora TIPp, which also contains interpreter-mediated court trials, and INTELPRAGMA / PRAGMACOR, made of telephone interpreter-mediated interactions (Himoro and Pareja-Lora, 2022). Most of the corpora of dialogue interpretations have been processed and analysed with the software EXMARaLDA.

The increasing interest in NLP and the automation of processes has brought us to multidisciplinary projects that deal with the development of models for automated oral communication. Machine interpreting has already been developed and is being improved, focusing on speed and accuracy matters (Müller et al. 2016). Either domain-specific (commercial, military, humanitarian...) or general (Skype Translator), there is still a long way to go to render machine interpreting more human-like (Braun, 2019).

The eight contributions contained in this volume are rich and varied. Machine translation is profusely tackled, with a special emphasis on under resourced languages and combinations, such as Ngambay-French or Romanian-Spanish, and literary translation, a field in which neural machine translation and generative pre-trained transformer models are contributing to improve quality and smooth the translation workflow. Pre-trained transformer models are also approached in this volume, not only to experiment in their design for specific challenges, such as idioms, but also to evaluate their usefulness at different stages of the translation workflow, such as the revision phase. In sum, this volume offers cutting-edge studies revolving around artificial intelligence, NLP and large language models in relation to both translation and interpreting, presenting innovative research results while opening new paths to further experimentation.

References

Alfred. V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling, volume 1*. Prentice-Hall, Englewood Cliffs, NJ.

Philip Angermeyer, Bernd Meyer, B., and Thomas Schmidt. 2012. Sharing community interpreting corpora. In Thomas Schmidt, T., and Kai Wörner, K. (eds.) *Multilingual Corpora and Multilingual Corpus Analysis*. Amsterdam: John Benjamins, pages 275-294.

Sabine Braun. 2015. Remote interpreting. In Holly Mikkelsen, and Renée Jourdenais (Eds) *The*

Routledge Handbook of Interpreting. London/New York: Routledge, pages 352-367.

- Sabine Braun. 2019. Technology and interpreting. In Minako O'Hagan (Ed.) *The Routledge Handbook of Translation and Technology*. London: Routledge, pages 271-288.
- Gabriel Cabrera Méndez. 2016. *Mamá, quiero ser intérprete*. Madrid: Pie de Página Editorial.
- Gloria Corpas Pastor. 2021. Technology Solutions for Interpreters: The VIP System. *Hermēneus. Revista de Traducción e Interpretación*, 23: 91-123.
- Michael Cronin. 2010. Globalization and translation. In Yves Gambier, and Luc van Doorslaer (eds.) *Handbook of Translation Studies. Volume 1*. Amsterdam: John Benjamins, pages 134-140.
- Claudio Fantinuoli. 2018. Interpreting and technology. The upcoming technological turn. In Claudio Fantinuoli (ed.) 2018. *Interpreting and Technology*. Berlin: Language Science Press, pages 1-12.
- Marcelo Yuji Himoro, and Antonio Pareja-Lora. 2022. Codificación y anotación preliminar de un corpus oral multilingüe de conversaciones telefónicas interpretadas para el estudio de los ataques a la imagen. *Pragmalingüística*, (30): 413-432.
- Miguel Ángel Jiménez Crespo. 2021. Localization. In Esperança Bielsa, and Dionysios Kapsaskis (eds.) *The Routledge Handbook of Translation and Globalization*. London: Routledge, pages 392-405.
- Markus Müller, Thai Son Nguyen, Jan Niehues, Eunah Cho, Bastian Krüger, Thanh-Le Ha, Kevin Kilgour, Matthias Sperber, Mohammed Mediani, Sebastian Stüker, and Alex Waibel. 2016. Lecture Translator - Speech translation framework for simultaneous lecture translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, San Diego, California. Association for Computational Linguistics, pages 82-86.
- Sharon O'Brien. 2012. Translation as human-computer interaction. *Translation Spaces*, 1: 101-122.
- Minako O'Hagan. 2013. The Impact of New Technologies on Translation Studies: A technological turn? In Carmen Millán, and Francesca Bartrina. *The Routledge Handbook of Translation Studies*, London: Routledge, pages 503-518.
- Koichiro Ryu, Shigeki Matsubara, Nobuo Kawaguchi, and Yasuyoshi Inagaki, Y. 2003. Bilingual Speech Dialogue Corpus for Simultaneous Machine Interpretation Research. In *Proceedings of the 6th Oriental COCOSDA workshop in Singapore on Oct 1-3 2003*, pages 164-168.

Miriam Shlesinger. 1998. Corpus-Based Interpreting Studies as an Offshoot of Corpus-Based Translation Studies. *Meta*, 43(4): 486-493.

Chan Sin-Wai. 2004. *A Dictionary of Translation Technology*. Hong Kong: The Chinese University Press.

Akira Takagi, Shigeki Matsubara, and Yasuyoshi Inagaki. 2002. A corpus-based analysis of simultaneous interpretation. In *Proceedings of International Joint Conference of the 5th Symposium on Natural Language Processing (SNLP-2002)*, pages 167-174.

Fred Van Besien. 1999. Anticipation in Simultaneous Interpretation. *Meta* 44(2): 250-259.