The Translation Technology revolution

(Don't get the translation wrong)

Ruslan Mitkov

Abstract

We witnessed the birth of the modern computer between 1943 and 1946; it was not long after that Warren Weaver wrote his famous memorandum in 1949 suggesting that translation by machine would be possible. Weaver's dream did not quite come true: while automatic translation went on to work reasonably in some scenarios and to do well for gisting purposes, even today, against the background of the latest promising results delivered by statistical Machine Translation (MT) systems such as Google Translate and latest developments in Neural Machine Translation and in general Deep Learning for MT, automatic translation gets it often wrong and is not good enough for professional translation. Consequently, there has been a pressing need for a new generation of tools for professional translators to assist them reliably and speed up the translation process. First Krollman put forward the reuse of existing human translations in 1971. A few years later, in 1979 Arthern went further and proposed the retrieval and reuse not only of identical text fragments (exact matches) but also of similar source sentences and their translations (fuzzy matches). It took another decade before the ideas sketched by Krollman and Arthern were commercialised as a result of the development of various computer-aided translation (CAT) tools such as Translation Memory (TM) systems in the early 1990s. These translation tools revolutionised the work of translators and the last two decades saw dramatic changes in the translation workflow.

The TM memory systems indeed revolutionised the work of translators and now the translators not benefiting from these tools are a tiny minority. However, while these tools have proven to be very efficient for repetitive and voluminous texts, are they intelligent enough? Unfortunately, they operate on fuzzy (surface) matching mostly, cannot benefit from already translated texts which are synonymous to (or paraphrased versions of) the text to be translated and can be 'fooled' on numerous occasions. The presentation will start with a snapshot in time by briefly outlining the emergence of computers followed by the advent of MT, then TM tools and briefly overviewing the state of the art.

What is next in the translation world? We cannot get it wrong as we cannot let the translation go wrong: it is obvious that the next generation of TM systems will have to be more intelligent. A way forward would be to equip the TM tools with Natural Language Processing (NLP) capabilities. NLP can come to help and propose solutions towards addressing this objective. The invited talk will present recent and latest work by the speaker and his colleagues from the Research Group in Computational Linguistics at the University of Wolverhampton in achieving this. More specifically, the speaker will explain how two NLP methods/tasks, namely paraphrasing and clause splitting as well as latest Deep Learning techniques, make it possible for TM systems to identify semantically equivalent sentences which are not necessarily identical or close syntactically and enhance performance. The first evaluation results of this new generation TM matching technology are already promising…

The speaker will then move to discuss the recent advances in Neural Machine Translation and what impact it has on the translation industry and translators. He will also cover other recent translation technology trends such as automatic post-editing; his own work will also be outlined.

The speaker promises to go beyond the translation world: he is already thinking not only about the next-generation translation memory tools for translators but also about the future interpretation memory (as well as NLP-inspired) tools for interpreters.

The development of new generation translation tools and high performing Machine Translation engines is essential, but not sufficient. What is no less important is educating and training a new generation of translators and interpreters who are familiar with the latest technology for translation and interpreting, are able to employ it expertly and are even capable of developing the tools of the future. In this context, the speaker will share his experience from the first and only in the world (Erasmus Mundus) Master Programme 'Technology for Translation and Interpreting' which he designed and for which he is director. He will also provide examples of new career opportunities translation technology can offer to students studying translation and interpreting.

Bionote

Prof Dr Ruslan Mitkov has been working in Natural Language Processing (NLP), Computational Linguistics, Corpus Linguistics, Machine Translation, Translation Technology and related areas since the early 1980s. Whereas Prof Mitkov is best known for his seminal contributions to the areas of anaphora resolution and automatic generation of multiple-choice tests, his extensively cited research (more than 270 publications including 20 books, 35 journal articles and 40 book chapters) also covers topics such as deep learning for NLP, machine translation, translation memory and translation technology in general, bilingual term extraction, automatic identification of cognates and false friends, natural language generation, automatic summarisation, computer-aided language processing, centering, evaluation, corpus annotation, NLP-driven corpus-based study of translation universals, text simplification, NLP for people with language disorders and computational phraseology. In addition, Ruslan Mitkov is well known for his vision in research based on innovative ideas and drive towards research output which seeks to enhance the work efficiency of different professions (e.g. for teachers, translators and interpreters) or seeks to improve the quality of life (e.g. for people with language disabilities) and which has significant impact beyond academia. Mitkov is author of the monograph *Anaphora resolution* (Longman) and Editor of the most successful Oxford University Press Handbook - *The Oxford Handbook of Computational Linguistics* whose second and substantially revised edition was published in June 2022. Current prestigious projects include his role as Executive Editor of the Journal of Natural Language Engineering published by Cambridge University Press and Editor-in-Chief of the Natural Language Processing book series of John Benjamins publishers. Dr Mitkov is also working on the forthcoming Oxford Dictionary of Computational Linguistics (Oxford University Press, co-authored with Patrick Hanks) and the *Oxford Handbook of Phraseology* Linguistics (Oxford University Press, co-authored with Gloria Corpas and Jean-Pierre Colson). Prof Mitkov has been invited as a keynote speaker at more than 200 international conferences. He has acted as Chair or Programme Chair of more than 65 international conferences on Natural Language Processing (NLP), Machine Translation, Translation Technology, Translation Studies, Corpus Linguistics and Anaphora Resolution. He is asked on a regular basis to review for leading international funding bodies and organisations and to act as a referee for applications for Professorships both in North America and Europe. Ruslan Mitkov is regularly asked to review for leading journals, publishers and conferences and serve as a member of Programme Committees or Editorial Boards. Prof Mitkov has been an external examiner of many doctoral theses and curricula in the UK and abroad, including Master's programmes related to NLP, Translation and Translation Technology. Prof Mitkov is Coordinator (Director) of the first and only Erasmus Mundus Master's Programme in Technology for Translation and Interpreting - an innovative and inspirational programme, with a strong research focus but an equally strong emphasis on business; leading companies in the global translation and language industry participate as associated partners. Dr Mitkov has considerable external funding to his credit (more than £ 20,000,000) and has been Principal Investigator of 25 projects, funded by UK research councils, by the EC as well as by companies and users from the UK and USA. Ruslan Mitkov received his MSc from the Humboldt University in Berlin, his PhD from the Technical University in Dresden and worked as a Research Professor at the Institute of Mathematics, Bulgarian Academy of Sciences, Sofia. Mitkov is Professor of Computational Linguistics and Language Engineering at the University of Wolverhampton which

he joined in 1995 and where he set up the Research Group in Computational Linguistics. His Research Group has emerged as an internationally leading unit in applied Natural Language Processing and members of the group have won awards at different NLP/shared-task competitions and conferences. In addition to being Head of the Research Group in Computational Linguistics, Prof Mitkov is also Director of the Research Institute in Information and Language Processing and Director of the Responsible Digital Humanities Lab. The Research Institute consists of the Research Group in Computational Linguistics and the Research Group in Statistical Cybermetrics, which is another top performer internationally. Ruslan Mitkov is Vice President of ASLING, an international Association for promoting Language Technology. Dr Mitkov is a Fellow of the Alexander von Humboldt Foundation, Germany, was a Marie Curie Fellow, Distinguished Visiting Professor at the University of Franche-Comté in Besançon, France and Distinguished Visiting Researcher at the University of Malaga, Spain; he also serves/has served as Vice-Chair for the prestigious EC funding programmes 'Future and Emerging Technologies' and 'EIC Pathfinder Open'. In September 2022 the renowned National Board of Medical Examiners (USA) presented Prof Mitkov with a certificate of distinguished collaboration which resulted in lasting impact on the strategic planning and decision making of the US organisation and their employment of NLP solutions to assessment for the last 17 years. In recognition of his outstanding professional/research achievements, Prof Mitkov was awarded the title of Doctor Honoris Causa at Plovdiv University in November 2011. At the end of October 2014 Dr Mitkov was also conferred Professor Honoris Causa at Veliko Tarnovo University and on 25 October 2022 Prof R Mitkov received the title 'Doctor Honoris Cause' for the third time, this time awarded by New Bulgarian University, Sofia.