Just a Dash of AI to Streamline the Translation Workflows?

BY STEVE DEPT

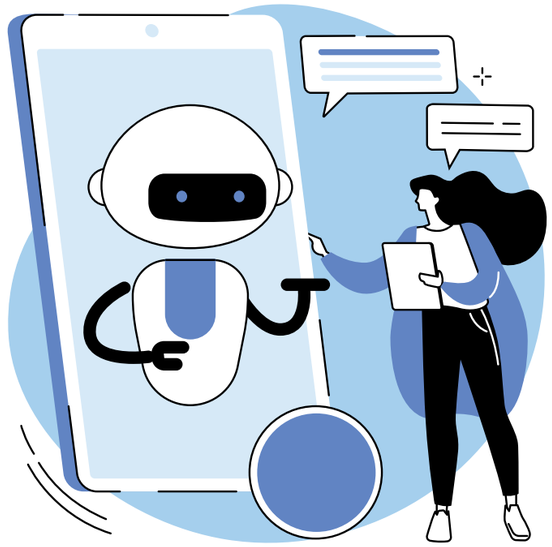
After experimenting with a rule-based MT (machine translation) engine called SYSTRAN in the nineties, we were convinced that translation quality control would soon involve a new type of revision: post-editing of MT. The term post-editing, involving a comparison of the source and target versions, had been around for about as long as MT, which is already in its seventh decade.

Post-editing was already on our radar when cApStAn was set up in 2000 to evaluate and improve translation through the lens of semantic equivalence and cross-language comparability. We did not implement it in our first major project (the OECD Programme for International Student Assessment, or PISA), but our second large contract was for RPE (rapid post-editing) of internal working documents of the European Union. This project highlighted the immense variability in the quality of the MT output, with unfamiliar types of errors (for example, "Vancouver, B.C." became "Vancouver, Before Christ" in the translated version). When the source version was in English, we noticed that non-native English speakers produced source material that generated significantly better translation results (and less gibberish) than native speakers of English.

Intelligibility of idiomatic language is debatable, and one cannot just inject artificial intelligence (AI) in a translation workflow. One needs to rethink the process, to redefine requirements, to determine new quality indicators, and to monitor the outputs differently. This is not utopian, but it does call for a collective reset and the willingness to face a steep learning curve. At this stage, there is no guarantee of a return on investment in terms of instrument quality. We are still a long way from running a stimulus and an item through a translation engine, then letting a human reviser fix the errors, assuming efficiency gains without loss of quality. In effect, one would need to engineer a new workflow rather than just tweak it. Training item developers to use a form of controlled writing that the MT engine would have better chances to interpret correctly should be considered, as well as piloting the assessment in all or most of the translated languages. The MT engine generates a prediction based on the bilingual content to which it has access and on which it has been trained. The prediction then needs to be validated, or modified, or replaced by a more suitable translation.

How is this relevant to IEA studies?

When IEA enlisted cApStAn's help to verify the extent to which national versions of TIMSS 2003 complied with translation and adaptation guidelines, to report deviations and to propose corrective action, we looked at the feasibility of using CAT (computer-assisted translation) tools. CAT tools are not MT, mind you. CAT tools are productivity tools that generate translation memories, so that bilingual pairs of sentences are stored in a database and automatically called up when a similar sentence needs to be translated. The quality of a translation memory solely depends on how good the translator is, while the quality of MT depends on an algorithm and on how good the datasets are, with which the MT engine is trained. At the time, it was difficult to make a case for the MT approach because it represented a paradigm shift that was not compatible with the time constraints of the study.



Today, the quality of MT engines has improved: state-of-the-art technology is no longer rule-based or statistical, we now have NMT (neural machine translation), which uses deep learning to improve its output. To implement this successfully, by achieving at least the same quality with less resources, the main ingredients are to increase the lead time, to do more language engineering upfront, and more testing. Translation verification could be replaced by post-editing, which could potentially lead to a shorter time frame for a semi-automated translation process. Exciting times ahead!

"One needs to rethink the process."