

Giving the Power to Bilingual Speakers

Position paper for the APE Workshop

Ariadna Font Llitjós

Language Technologies Institute
Carnegie Mellon University
Pittsburgh, 15217
aria@cs.cmu.edu

Post-editing has traditionally been used to improve the quality of MT output to make it usable. More recently, post-editing has also been shown to be useful for automatic evaluation of MT systems. As an alternative to current automatic evaluation metrics, such as BLEU and NIST, which are criticized for not having a high correlation with human judgments, minimal post-editing of MT output provides a good measure of how good a system translation is, and as an end result it also provides us with human reference translations that are relevant to MT systems. Such system friendly translations can then be used to automatically evaluate systems much more fairly, since systems are not penalized for having picked a different sense of the word or having picked a different, yet correct, syntactic construction (Snover *et al.*, 2006).

Mostly for practical reasons, post-editing has been a monolingual task done by linguists or editors without the source language information available. However, not having access to the original meaning can hinder the post-editing activity, and can even yield incorrect reference translations. In order to palliate this problem, for the GALE evaluation, multiple translators had to agree on a gold standard reference translation that conveys the exact same meaning as the source language sentence. This gold standard is then used by post-editors as if it were the original source language sentence when post-editing the raw MT output. Nevertheless, not only is the crafting of such a gold standard a difficult task, but in addition, there can be no guarantees that the real meaning of the source language sentence is fully captured by the gold standard translation.

For language pairs with a large number of speakers, and certainly for the pairs researchers are working on for GALE (Olive, 2005), Chinese-English and Arabic-English, it would certainly be easier to get bilingual speakers to be post-editors, than having to generate a gold standard for each translation.

In this context, I use the term bilingual speaker loosely. When I say bilingual speakers, I do not mean people that were born and raised bilingually and have native skills in both languages, but rather people that are native in one of the two languages and are fluent in the second language. A good example of what I mean by bilingual speakers here could be second generation Chinese and Arabic people that were born and live in the United States or another English-speaking country.

At this point, I would like to propose a new bilingual post-editing paradigm, one where post-editing does not have to be an exclusive activity of linguists or editors. Initial user studies have shown that with user-friendly online post-editing tool, non-expert bilingual speakers can provide reliable post-editing information, given the source language sentence as well as the corresponding MT output (Font Llitjós and Carbonell 2004).

And what is more, rather than just improving MT output, we should use post-editing to improve MT system themselves by feeding it back into the systems. Researchers have proposed automated post-processing of raw MT output to alleviate the tedious task of manual post-editing by correcting the most frequent errors beforehand (Knight & Chander, 1994). However, to this day, there exists no solution to fully automate the post-editing process.

In order to improve MT systems, word alignment information also needs to be presented for correction. In the case of Transfer-based MT (TBMT) systems, bilingual post-editing can be recycled into the system by tracking the errors and fixing the incorrect rules, thus improving the lexicon and the grammar of the MT system (Font Llitjós *et al.* 2005). For data-driven systems,

such as Example-Based MT (EBMT) and Statistical MT (SMT), bilingual post-editing provides a great way to obtain annotated data, for which new, smarter training algorithms can be developed. SMT systems need exponentially more data to improve a very small fraction, but for most languages such a brute force approach will simply not work. The statistical MT community needs to start thinking of smarter ways to use post-editing information to improve their systems.

Discriminative training could make use of the existence of negative as well as (relevant) positive examples, getting away from the bad translation in addition to getting closer to the good one. At the very least, SMT folks should be able to do a better job at tuning their systems parameters given the right optimization function, namely a relevant reference translation result of the minimal post-editing process, instead of an independent reference translation, possibly very different from what the system produced.

Bilingual post-editing combined with TBMT is greatly suited for resource-scarce situations, where there exists no parallel data and linguists or editors are often not available, since most native speakers of a resource-poor language (such as Quechua or Mapudungun) have a good working knowledge of a resource-rich language (such as English or Spanish).

While I hope this approach will be useful to correct and improve MT systems for very long sentences as the quality of our systems improve, current MT output of very long sentences is too garbled, and so it is not realistic to expect post-editors to do a good job at correcting word alignments. Currently, this approach is better suited for limited domain applications, such as the travel domain or the medical domain, where the length of the MT output typically does not exceed 15 words.

Online cooperation of users to improve existing databases has recently experienced a boom, good examples of this phenomenon are Wikipedia, LEO dictionary and the Rosetta Project. Translation users would surely feel inclined to improve MT output using an online user-friendly tool similar to the one proposed by Font Llitjós and Carbonell (2004), since they would see the benefit of it, namely getting better translations next time they use the system. Making such a post-editing tool available to thousands of bilingual speakers around the world through the reaching power of the Internet, would mean that MT systems would have access to a large number of correction feedback relevant to errors they make. In this context, we are not talking about just a few hundred data points that would surely be washed out by the millions of sentences of training data, but rather we are talking about tens of thousands of correct data points that could surely have a positive effect on the quality of even SMT systems.

In sum, if we relax the assumptions of traditional post-editing and incorporate alignment correction, in addition to word correction, and we make it widely available to a new set of non-expert *post-editors*, in this case bilingual speakers, then we can collect large amounts of relevant correction data that can be used to improve even Statistical MT systems.

References

Font Llitjós, A.; J. Carbonell and A. Lavie (2005). A Framework for Interactive and Automatic Refinement of Transfer-based Machine Translation. EAMT.

Font Llitjós, A. & J. Carbonell (2004). *The Translation Correction Tool: English-Spanish user studies*. LREC.

LEO dictionary: <http://dict.leo.org/>

Joseph Olive. 2005. *Global Autonomous Language Exploitation (GALE)*. DARPA/IPTO Proposer Information Pamphlet.

The Rosetta Project: <http://www.rosettaproject.org/>

Snover, M; B. Dorr, R. Schwartz, L. Micciulla, J. Makhoul. A Study of Translation Error Rate with Targeted Human Annotation. AMTA 2006.

Wikipedia: <http://en.wikipedia.org/>