Abs-259

Dorothy Kenny (Dublin City University)

A corpus-based study of the discourse of contemporary machine translation

Human translators have co-existed peacefully with automatic translation for decades, but recent advances in statistical machine translation (SMT, as used in systems like Google Translate) are forcing a reappraisal of this relationship, for a number of reasons. One is that SMT is based on the recycling of human translations, so never before has machine translation been so reliant on human translation. Another is that free online machine translation and collaborative translation platforms allow non-professionals to participate in translation projects in ways that were not possible previously.

In this paper I address a number of issues that arise in this context, focusing on the discourse of statistical machine translation. Drawing on ideas from Conceptual Metaphor Theory, lexical semantics, critical discourse analysis and corpus-based studies of metaphor (especially Deignan 2005, Goatly 2007), amongst others, I examine how key participants in SMT conceptualize contemporary machine translation, and attempt to explain some areas of divergence between the view of translation adopted in machine translation and more traditional translation studies. I argue that the "code-breaking metaphor", heavily referenced in machine translation circles, is based on metonymy rather than being a "pure" metaphor, and that it exhibits a number of the features that Deignan (ibid.) associates with metonymy-based metaphor (eg lack of systematic mapping between source and target domain; limited overlap between collocates in the source and target domain). I ultimately argue that "code-breaking" is not a convincing metaphor with which to explain translation, although it has significant ideological import. I also discuss how other metaphors used by computer scientists to describe translation (eg when translation is understood as "food") entail ideological positions that obscure the input of human beings to the translation process, a fact that raises a number of ethical issues for researchers and teachers of translation technology alike. Finally, I consider how certain sources address non-expert addressees in particular in such a way as to position translation outside the realm of paid work (referring to the addressee's 'friends' rather than his/her colleagues, for example).

The empirical work presented is based on two resources: firstly a multi-modal corpus of high-profile explanations of statistical machine translation (expert-to-lay communication), taken, for example, from Microsoft and Google blogs and video presentations, press reports, and press releases from major industrial sources; and secondly a corpus of journal articles and refereed conference papers (expert-to-expert communication), from the journal Machine Translation and the archive of the European Association for Machine Translation.

References

Deignan, Alice (2005) *Metaphor and Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins.

Goatly, Andrew (2007) *Washing the Brain - Metaphor and Hidden Ideology*. Amsterdam/Philadelphia: John Benjamins.