Abs-32
Ana Frankenberg-Garcia – ISLA Campus Lisboa
Sentence boundaries in translation

As predicted by Mona Baker almost two decades ago (Baker 1993), the rise of corpus linguistics has had a great impact on translation studies over the past years. Bilingual comparable corpora have been used in translation training (e.g. Kübler 2011) and to explore functional equivalents between languages (e.g. Tognini Bonelli & Manca 2004). Comparable corpora of translated and non-translated texts have been used to study what sets translated language apart (e.g. Laviosa 1998, Olohan & Baker 2000, Tirkkonen-Condit 2004, Puurtinen 2004, Frankenberg-Garcia 2008). And parallel corpora have been used in various studies comparing source-texts and translations (e.g. Frankenberg-Garcia 2005, Johansson 2007, Frankenberg-Garcia 2009, Pérez Blanco 2009, Lefer 2010, Bernardini 2010, Saldanha 2011). However, most existing corpus-based analyses focusing on the contrasts between source texts and translations are either purely lexical or are constrained by sentence boundaries. Probably the main reason underlying this limitation is the fact that, in parallel corpora, texts are usually segmented at the level of the sentence due to the relative ease with which sentence boundaries can be identified automatically. The alignment of source texts and translations is then usually carried out such that whenever there is not a one-to-one correspondence between source-text and translation segments, they are aligned either on a one-to-many or on a many-to-one basis, blurring out the details of what happens when source-text sentences are not preserved in translation. With the focus of this 6th Corpus Linguistics conference being on discourse, the study of which transcends the level of the sentence, this paper aims to explore shifts in sentence boundaries that occur in the process of translation.

The unique, manually post-edited alignment of COMPARA - a three-million-word parallel corpus of English and Portuguese fiction (Frankenberg-Garcia & Santos 2003) - enables one to retrieve automatically all sentences that were joined together and all sentences that were split apart in translation. Using this free online corpus, the present exploratory study seeks to answer questions such as (1) To what extent do translators preserve sentence boundaries? (2) What is more common in translation: sentence-splitting or sentence-joining? (3) Do English and Portuguese language translators differ in terms of preserving sentence boundaries? (4) Could author or translator style bear an influence on the extent to which sentence boundaries are preserved or not? (5) What kind of sentences do translators split and what kind sentences do they join together? It is believed that the findings observed can have implications not only for translator training, but also for bilingual text alignment and the development of machine translation and translation memory systems.

Baker, M. (1993) 'Corpus linguistics and translation studies. Implications and applications.' In M. Baker, G. Francis and E. Tognini-Bonelli (eds.) Text and Technology: In Honour of John Sinclair. Amsterdam & Philadelphia: John Benjamins, 233-250.

Bernardini, S. (2010) 'Parallel corpora and the search for translation norms/universals'. Plenary presented at MATS 2010, Methodological Advances in Corpus-Based Translation Studies, University of Gent, 8-9 January 2010. Slides available online at http://veto.hogent.be/actua/mats2010/presentations.cfm.

Frankenberg-Garcia, A. (2005) 'A corpus-based study of loan words in original and translated texts'. In P. Danielsson & M. Wagenmakers (eds.) Proceedings of the Corpus Linguistics 2005 conference, Birmingham, UK, 14-17 July. Available online at http://www.corpus.bham.ac.uk/pclc/index.shtml.

Frankenberg-Garcia, A. (2008) ' "Suggesting rather special facts": a corpus-based study of distinctive lexical distributions in translated texts'. Corpora, 3/2, 195-211.

Frankenberg-Garcia, A. (2009) 'Are translations longer than source texts? A corpus-based study of explicitation'. In Beeby, A., Rodríguez, P. & Sánchez-Gijón, P. (eds.) Corpus use and learning to translate (CULT): An Introduction. Amsterdam & Philadelphia: John Benjamins, 47-58.

Frankenberg-Garcia, A. & D. Santos (2003) 'Introducing COMPARA, the Portuguese-English parallel translation corpus'. In F. Zanettin, S. Bernardini & D. Stewart (eds.) Corpora in Translation Education. Manchester: St. Jerome Publishing, 71-87.

Johansson, S. 2007. Seeing through Multilingual Corpora: On the use of corpora in contrastive studies. Amsterdam & Philadelphia. John Benjamins.

Kübler, N. (2011). Working with corpora for translation teaching in a French-speaking setting. In A. Frankenberg-Garcia, L. Flowerdew & G. Aston (eds.) New Trends in Corpora and Language Learning. London: Continuum, 62-80.

Laviosa, S. (1998) 'Core patterns of lexical use in a comparable corpus of English narrative prose'. Meta, 43/4, 557-570. Available online
http://www.erudit.org/revue/meta/1998/v43/n4/003425ar.pdf

Lefer, M-A. (2010) 'Morphological productivity and untranslatability: a corpus-based approach.' Paper presented at MATS 2010, Methodological Advances in Corpus-Based Translation Studies, University of Gent, 8-9 January 2010. Slides available online at
http://veto.hogent.be/actua/mats2010/presentations.cfm.

Olohan, M. & M. Baker (2000) 'Reporting that in translated English: Evidence for subconscious processes of explicitation?' Across Languages and Cultures 1/2, 141-158.

Pérez Blanco, M. (2009) 'Translating stance adverbials from English into Spanish: a corpus-based study'. International Journal of Translation, 21, 41-55.

Puurtinen, T. (2004) 'Clause connectives in Finnish children's literature'. In A. Mauranen and P. Kujamäki (eds.) Translation Universals, Do They Exist? Amsterdam & Philadelphia: John Benjamins, 165-176.

Saldanha, G. (2011) 'Translator style: methodological considerations..The Translator, 17(1), 25-50.

Tirkkonen-Condit, S. (2004) 'Unique items – over – or under-represented in translated language?' In A. Mauranen and P. Kujamäki (eds.) Translation Universals, Do They Exist? Amsterdam & Philadelphia: John Benjamins, 177-184.

Tognini Bonelli, E. & E. Manca, Elena (2004) 'Welcoming Children, Pets and Guests: Towards a Functional Equivalence in the Languages of "Agriturismo" and "Farmhouse Holidays"'. TradTerm, 10, 295-312.