Measuring the unsaid: the SCOPIC project as a parallax cross-linguistic corpus

Mittwoch,
04.03.2020
13:45–14:45
VMP5 2101/2105

Nicholas Evans, Danielle Barth (Keynote)
Australian National University, Centre of Excellence for the Dynamics of Language
nicholas.evans@anu.edu.au, danielle.barth@anu.edu.au

In his foundational work, *Miseria y Esplendor de la Traducción*, Spanish philosopher Ortega y Gasset pointed out that 'Each people leaves some things unsaid in order to be able to say others. Because everything would be unsayable.' In other words, every formulation by a speaker, as they produce a text, is not just a choice to say something but also not to say something. Choices about what to say and what not to say **reflect** many forces – context, common ground, narrative progression, individual differences, but also asymmetries at the level of language and culture, as transmitted both through learned preferences about what to talk about, and grammatical and lexical affordances in terms of what is obligatory, compact or cumbersome in a particular language.

In corpus linguistics, measuring what is said is easy – since it is there! – while measuring what is not said is much harder, since there is in principle no limit to what could have been said. Simply comparing corpora across languages does not solve the problem either, unless there is complete matching both of genres and of topics. Nor do parallel corpora, in the strict sense of being translation equivalents, solve the problem, since there will always be a 'source language founder bias', where categories in the source language show frequencies more representatively than in the target languages.

In this talk we propose and illustrate a way around this problem, by using a 'parallax corpus', namely the SCOPIC corpus (Barth & Evans 2017) based on the 'Family Problems Picture Task' (San Roque et al. 2012). This includes over 369,000 words from 27 languages spanning every continent (and including one sign language). Crucially, it is a problem-solving task which (in contradistinction to the Pear Stories, for example) puts the speaker(s) in charge of the narrative, and allows speakers to choose what to express in their shaping of the story. At the same time, the connections of what is said to the picture stimuli allows the denoted scenes to be compared across languages, speakers, and subtasks (e.g. picture description vs first-person narrative). This allows sophisticated multidimensional comparisons of what speakers choose to say, and not to say, in a way that responds to topic, referent configuration, speaker, discourse type, and language (among others).

In this talk we first outline the logic and design of the task and the SCOPIC corpus, then demonstrate how it can be used to investigate cross-linguistic differences in formulation with regard to person reference, stance, and the depiction of attributed speech and thought.

References: Barth, D. & N. Evans. Language Documentation and Conservation Special Publication No. 12 Social Cognition Parallax Corpus (SCOPIC). http:/hdl.handle.net/10125/24742, 1–21. Ortega y Gasset (1937), Miseria y esplendor de la traducción. La Nación (Buenos Aires) May–June 1937. San Roque, L., A. Rumsey, L. Gawne, S. Spronck, D. Hoenigman, A. Carroll, Ju. Miller & N. Evans (2012). Getting the story straight: language fieldwork using a narrative problem-solving task. Language Documentation and Conservation 6, 134–173.