Hypothesis Testing and Theory Amending: What corpus research can offer Linguistics? -- a Chinese case

Yanan Hu Dirk Geeraerts Dirk Speelman

Research Unit QLVL, Department of Linguistics, Faculty of Arts, KU Leuven

With corpus-related methodology established in linguistic studies, we hereby address the most crucial contribution that corpus research makes to Linguistics, namely hypothesis testing and theory amending. Inspired by Speelman and Geeraerts (2009), our study re-examines the (in)direct causation hypothesis formulated first by Verhagen and Kemmer (1997) on Dutch causatives *doen* and *laten*, and analyzed by Stukker (2005), from a Chinese perspective.

Starting with the assumption put forward by Ni (2012) that in Chinese *shǐ* expresses direct causation like *doen* and *ràng* expresses indirect causation like *laten* (among the seven analytic causatives *shǐ*, *lìng*, *ràng*, *jiào1*, *jiào2*, *gěi* and *yào*), we apply multinomial logistic regression analysis and multiple correspondence analysis to a data set of occurrences retrieved from two corpora, the Sheffield Corpus of Chinese (http://www.hrionline.ac.uk/scc/db/scc/index.jsp) and the UCLA Chinese Corpus (http://www.lancs.ac.uk/fass/projects/corpus/UCLA/) 1st edition. All occurrences were annotated for a series of factors cited in the literature as being related to the (in)direct causation hypothesis (inanimateness of causer, coreferentiality of causer and causee, for example). Our study, which investigates how Chinese analytic causatives behave with respect to predictions derived from the (in)direct causation hypothesis, zooms in on the following questions: 1) How are the Chinese causatives distributed? 2) Do the aforementioned factors have a significant effect on the distribution of the causatives? 3) If so, do these factors suffice to adequately model the choices between the causatives?

Since our data has a chronological dimension, covering eras of mandarin Chinese from 1100 BC to 2005 AD, we'll also look into the aforementioned questions from a diachronic point of view. We visualize the periodized distributions of the seven causative verbs so that the changes of them along dimensions pertaining to (in)directness can be seen.

Our analysis shows that the (in)direct causation hypothesis is not unimportant to the Chinese case but that at the same time it is far from powerful enough to capture all significant variation. It fails to tell apart the very pair of *shǐ* and *ràng* in particular, which opposes Ni's claim (2012). All in all, the Chinese case of analytic causatives cannot validate the (in)direct causation hypothesis but rather calls for a more refined theoretical model to reveal Chinese linguistic construal of causality.

References

Ni, Yueru. 2012. *Categories of Causative Verbs: a Corpus Study of Mandarin Chinese*. Utrecht: Utrecht University MA thesis.

Speelman, Dirk and Dirk Geeraerts. 2009. Causes for causatives: the case of Dutch 'doen' and 'laten'. In Ted Sanders and Eve Sweetser (eds.), *Causal Categories in Discourse and Cognition* 173-204. Berlin/New York: Mouton de Gruyter.

Stukker, Ninke. 2005. *Causality marking across levels of language structure*. PhD dissertation, University of Utrecht.

Verhagen, Arie and Suzanne Kemmer. 1997. Interaction and Causation: Causative Constructions in Modern Standard Dutch. *Journal of Pragmatics* 27. 61–82.