Processing noncanonical word order in Czech

Ondrej Bojar[1, 1]Jiri Semecky, Shravan Vasishth[2 ,] Ivana Kruijff-Korbayova[2]
[1]Charles University, Prague, [2]Saarland University, Germany

In relatively free word order languages like Finnish and Hindi, when discourse context is not provided, deviating from canonical order results in increased processing difficulty (Hyona and Hujanen 1997, Vasishth 2003). However, it has been recently shown that in such relatively free word order languages, discourse context can facilitate the processing of noncanonical order (e.g., (Kaiser and Trueswell 2004)). These constraints also seem to apply in languages with not as free a word order, such as English (Altmann and Steedman 1988).

It is nevertheless possible that in languages like Czech, which have even freer word order than Finnish and Hindi, the relatively high frequency of noncanonical orders (Kruijff and Vasishth 2003) could have the consequence that processing is not adversely affected by noncanonical order even *without* any supporting discourse context.

To test this prediction for Czech, we conducted a self-paced reading study (n=42) involving six conditions, each a permutation of the string containing the agent (AGNT), patient (PAT) and the verb (V). These were structures like (1) below. Reading time at the main verb immediately following the permuted string was taken as a measure of processing difficulty since the integration of the permuted string into the sentence would occur at that point at the earliest.

(1) Uz mesic / se / Zdenek / knihu   / snazi / najit   / bez vysledku
      AGENT PATIENT VERB   MAIN-VERB
     Already for a month himself Zdenìk book aims to-find with no result
     'Zdenek has been trying unsuccessfully to find the book for a month already.'

The prediction of the frequency-based view is the null hypothesis: in Czech there should be no difference in processing difficulty with canonical versus non-canonical orders. The alternative possibility is the standard one assumed in the literature: noncanonical order should adversely affect processing in the absence of preceding context.

In order to argue for the null hypothesis, we did not use the standard technique of computing so-called observed power along with p-values because this has been shown to be an incorrect use of power. It is a fallacy to assume that, in the face of a null result from a t-test or ANOVA, high power (say, greater than .80) provides grounds for accepting the null hypothesis: a nonsignificant p value *entails* low observed power (see (Hoenig and Heisey 2001) for details). Consequently, we used a statistical technique called (bio-)equivalence testing (Berger and Hsu 1996); this is commonly used in the pharmaceutical industry to demonstrate, for FDA approval, effective equivalence of brand-name versus generic drugs. In equivalence tests (specifically, two one-sided t-tests or TOST), the null hypothesis is treated as the alternative hypotheis, and vice versa.

Assuming that a difference of less than 25 milliseconds amounts to effective equivalence, the results show that there is effectively no difference in processing ease with agent-before-patient versus patient-before-agent orders (only SVO vs. OVS was inconclusive using TOST). Thus, the absence of discourse context does not necessarily have an adverse effect on noncanonical order processing: a critical cross-linguistic variable is the degree of word order freedom available *a prior*i. Languages with freer word order do not suffer the effects of noncanonical order to the extent that comparatively rigid order languages do.

References
Altmann, G.T.M. and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition* 30, 191-238.
Berger, R.L. and Hsu, J.C.. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11(4), 283-302.
Hoenig and Heisey (2001). The abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis. The American Statistician, 55(1), 19-24.
Hyona, J. and Hujanen, H. (1997). Effects of case marking and word order on sentence parsing in Finnish: An eye fixation analysis. *Quarterly Journal of Experimental Psychology*, 50A(4), 841-858.
Kaiser, E. and Trueswell, J.C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, in press.
Kruijff, G.J.M. and Vasishth, S. (2003). Quantifying word order freedom in natural language: Implications for sentence processing. In Proceedings of the AMLaP conference, Clasgow, UK.
Vasishth, S. (2003). Working memory in sentence processing: Processing Hindi center embeddings. Garland Press, New York.