
Entropy predicts uncertainty in subcategorization frame distributions

Aaron Steven White
Johns Hopkins University
 aswhite@jhu.edu

Kyle Rawlins
Johns Hopkins University
 kgr@jhu.edu

Freitag
 10.03.2017
 12:00 – 12:30
 B3 2, 0.03

AG1

Syntactic bootstrapping approaches to verb-learning posit that learners use a verb's subcategorization frame (SCF) distribution to learn its meaning (Landau & Gleitman 1985). Most computational models of syntactic bootstrapping assume that this is carried out via some form of verb clustering based on by-verb SCF distribution entropy (Alishahi & Stevenson 2008, White 2015). Recent work in online sentence processing raises a potential problem for this view: while comprehenders are sensitive to surprisal (Hale 2001) and entropy reduction (Hale 2006) in online processing, they are not sensitive to a verb's SCF distribution entropy (Linzen & Jaeger 2015).

We show that, while SCF distribution entropy is not deployed online, it is nonetheless encoded, yielding traces in the variability found across participants' responses on offline measures. To establish this, we derive a measure of SCF uncertainty using White & Rawlins's (2016) large-scale acceptability judgment dataset and a measure of SCF entropy using Korhonen et al.'s (2006) VALEX dataset for 1000 different verbs. We show that, controlling for frequency, there is a reliable positive relationship between uncertainty and entropy, suggesting that computational models of syntactic bootstrapping are licensed in assuming access to SCF entropy.

References: • Alishahi, A., and S. Stevenson. 2008. A computational model of early argument structure acquisition. *Cognitive Science* 32:789–834. • Hale, J. 2001. A Probabilistic Earley Parser As a Psycholinguistic Model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*. • Hale, J. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30:643–672. • Korhonen, A., Y. Krymolowski, and T. Briscoe. 2006. A large subcategorization lexicon for natural language processing applications. In *Proceedings of LREC*. • Landau, B., and L.R. Gleitman. 1985. *Language and experience: Evidence from the blind child*. Harvard University Press. • Linzen, T., and T.F. Jaeger. 2016. Uncertainty and expectation in sentence processing: evidence from subcategorization distributions. *Cognitive science* 40(6): 1382–1411. • White, A.S. 2015. *Information and incrementality in syntactic bootstrapping*. Doctoral Dissertation, University of Maryland. • White, A.S., and K. Rawlins. 2016. A computational model of S-selection. In *Semantics and Linguistic Theory* 26.