

Subcategorization frame entropy in online verb learning

Aaron Steven White*, Valentine Hacquard**, Philip Resnik** & Jeffrey Lidz**

*Johns Hopkins University, **University of Maryland, College Park

Contact: aswhite@jhu.edu

Background. Syntactic bootstrapping approaches to verb learning posit that learners use a verb's subcategorization frame (SCF) distribution to learn its meaning [3]. Thus, they require that this distribution can be inferred from the input. [1] raise a potential problem for syntactic bootstrapping: child-directed speech (CDS) has lower by-verb SCF entropy than adult-directed speech. This suggests that CDS is less likely to provide each verb's full SCF distribution in reasonable samples.

Findings. Here, we show (i) that online verb-learning is actually *improved* by the presentation of lower entropy subcategorization frame distributions when the number of samples presented from that distribution are small and (ii) that higher entropy only matters as sample size grows. This suggests that, rather than being a hindrance, low entropy may be necessary for getting syntactic bootstrapping off the ground, while high entropy feeds later fine-tuning of the learned meanings.

Corpus study. We first show that different utterance contexts in which learners find themselves give rise to different SCF entropy. Verb-SCF pairs were automatically extracted from each transcript in the Gleason corpus [4], which contains dependency-parsed transcripts of both play and dinner contexts for 24 children. A stratified nonparametric bootstrap was used to first match sample size by utterance context via subsampling within child, followed by resampling within child. The by-verb SCF entropy was calculated on each resampled dataset, and a mixed effects zero-inflated gamma (ZIG) model was fit to those entropies with fixed effects for context and random intercepts for child. Both fixed effects components of the ZIG suggest higher entropy in dinner contexts.

Norming task. Our norming task, based on the "one-shot" Human Simulation Paradigm (HSP) [2,6,7] task in [5,8], aims to measure the semantic informativity of particular sentences in CDS. We use this measure in constructing the main study and to establish that item informativity, as opposed to informativity of the whole distribution, is the same across contexts. All sentences containing the 10 most frequent clause-embedding verbs in Gleason were extracted. For each verb and context, 20 sentences were sampled and the verb replaced with a blank. Standard HSP nonce variants were also created. 677 participants were recruited through Amazon Mechanical Turk (AMT) to fill in the blank in each sentence. Logistic mixed models (random inter.: PARTICIPANT and VERB) of response accuracy were constructed; LEXICAL CONTEXT (*nonce, real*) and TRUE VERB LOG FREQUENCY are significant in LLRTs ($ps < 0.001$), but UTTERANCE CONTEXT (*dinner, play*) is not.

Main study. Our main study trains participants on a novel verb seen in sets of the above syntactic contexts and afterward asks them to make similarity judgments between the novel verb and known verbs (*novel judgments*). These judgments are then compared to similarity judgments for the 10 real verbs used to construct the sets (*real judgments*), made publicly available by [9]. For each sentence from the norming, a nonce verb was inserted in place of the blank. Each verb's sentences were then divided into high and low informativity sets based on norming accuracy (cf. [5,9]) and another norming task. These were further subdivided into large (10) and small (5) sets. 4800 participants were recruited through AMT and shown one condition of the 160 created by crossing UTTERANCE CONTEXT, LEXICAL CONTEXT, VERB, INFORMATIVITY, and SET SIZE. They were then asked to judge the similarity between the trained nonce word and the 31 attitude verbs investigated in [9]. A linear mixed model was fit with correlation between z-scored *nonce* and *real judgments* as the dependent variable and the above factors plus TRUE VERB LOG FREQUENCY as predictors. We find (i) in small high informativity sets, play correlations are higher than dinner but get no higher with large sets, and (ii) in large high informativity sets, dinner correlations are higher than play.

Selected references. [1] Buttery & Korhonen 2005. [2] Gillette et al. 1999. [3] Landau & Gleitman 1985. [4] Masur & Gleason 1980. [5] Medina et al. 2011. [6] Papafragou et al. 2007. [7] Snedeker & Gleitman 2004. [8] Trueswell et al. 2013. [9] White et al. 2015.