Learning and evaluating hierarchies of verb argument structure

Jesse Mu*[1], Joshua Hartshorne[2], and Timothy O'donnell[3]

[1]Department of Computer Science, Stanford University – Macau SAR China
[2]Department of Psychology, Boston College – United States
[3]Department of Linguistics, McGill University – Canada

Abstract

Why can Sally *like to read* but not **appreciate to read*? How children learn the constraints between verbs and their arguments is a central question in language acquisition. Theorists (Levin, 1993) have long noted that verbs can be organized into classes based on the possible arrangements of their arguments (syntactic frames). One especially common assertion is that verb argument structure can be at least partially described by a hierarchy: each verb belongs to a class, which itself may belong to a number of broader superclasses. However, experimental and linguistic evidence for this phenomenon has been limited to relatively small case studies (Ambridge et al., 2011; Levin and Rappaport Hovav, 2005). While a large-scale classification of verbs is encoded in VerbNet (Kipper et al., 2008), its hierarchical structure has been incrementally handcrafted over decades and not extensively tested.
In this work, we provide broad-coverage psycholinguistic evidence for a hierarchical classification of verbs. We first show that such a hierarchy is objectively well-supported by the verbs and frames in English, by training the Bayesian Hierarchical Clustering algorithm (Heller and Ghahramani, 2007) on the verb-frame assignments in VerbNet (Figure 1). Using clustering similarity measures (Table 1) and phylogenetic tree comparison tools (Figure 2), we find that BHC and the handcrafted VerbNet taxonomy converge on the same structure, thus substantiating both methods and the common hierarchy they find.

Second, we show that these hierarchies capture meaningful psychological dimensions of generalization. In a verb coercion task, we ask if speakers are more likely to extend known verbs to unattested frames if the frame is attested for verbs in a closely-related class. We find that predictions from both VerbNet and BHC correlate well with human verb-frame coercion judgments (Figure 3). Nonetheless, this correlation is far from ceiling, suggesting limitations of a single hierarchy of verb classes. We discuss alternative, more complex representations.
Finally, in follow-up (in progress) work, we turn to the problem of *learning* hierarchies of argument structure from corpora. While there is a substantial literature learning verb argument structure patterns (Barak, Goldberg, and Stevenson, 2016; Kawahara, Peterson, and Palmer, 2014; Reichart and Korhonen, 2013) , these models typically learn a flat, non-hierarchical set of verb classes, which cannot capture the human generalization behavior observed in our experiments. Thus, using the same clustering and evaluation methods above, we examine whether corpus-based representations such as word embeddings or extracted frame counts can recover VerbNet's hierarchical structure. For example, we find that applying BHC to dependency-based embeddings (Levy and Goldberg, 2014) results in a hierarchy more similar

*Speaker

to VerbNet's taxonomy than chance, but much less so than BHC trained on gold-standard verb-frame assignments. These results shed light on semantic and syntactic information needed for child learners and suggest a need for more complex models of argument structure acquisition.

Keywords: argument structure, syntax, verbs, computational model, hierarchical clustering