

A VerbNet Variants

We considered 3 kinds of annotations on arguments that could be added or removed: selectional restrictions on arguments (S), thematic roles (T), and prepositional literals (P) (Figure 2). This makes 8 ways of encoding syntactic frames, in addition to using the default VerbNet encoding, which contains some thematic roles and prepositional literals. These 9 variants have anywhere from 61 to 3644 frames.

The statistics in the main paper are computed in Table S1 for all 9 variants. Besides the first two extremely minimal variants, we did not observe significant differences in fit to VerbNet. We selected the -S/+T/+P variant as it was easy to generate stimuli: having prepositional literals was useful, and it was easy to choose nouns for NPs with general thematic role restrictions, but it was often tricky to pick nouns that tested certain selectional restrictions (e.g. +INT_CONTROL in Figure 2).

Table S1: Comparisons from BHC to VerbNet for each VerbNet encoding. The encoding used in the main paper is delineated.

Encoding			Comparison statistic							
S	T	P	n^1	H_{super}	H_{standard}	H_{sub}	C_{super}	C_{standard}	C_{sub}	Entanglement
-	-	-	73	0.65	0.67	0.65	0.67	0.97	1.00	? ²
-	-	+	109	0.73	0.75	0.72	0.70	0.99	1.00	0.26
-	+	-	170	0.91	0.90	0.86	0.72	0.99	1.00	0.21
-	+	+	148	0.88	0.88	0.83	0.72	0.99	1.00	0.20
+	-	-	179	0.92	0.92	0.88	0.71	0.99	1.00	0.31
+	-	+	165	0.89	0.89	0.84	0.71	0.99	1.00	0.27
+	+	-	223	0.96	0.96	0.91	0.72	0.99	1.00	0.29
+	+	+	186	0.90	0.90	0.86	0.72	0.99	1.00	0.27
		default	163	0.87	0.87	0.83	0.71	0.98	1.00	0.26

¹ Number of clusters in the flat clustering after BHC cut.

² Due to a bug in version 3.5.5 of Dendroscope (Huson and Scornavacca, 2012), we were unable to produce this specific statistic.

B BHC Details

Here we outline in more detail BHC’s posterior predictive distribution and the process for evaluating the conditional probabilities of verbs taking syntactic frames. For full details, see Heller and Ghahramani (2005).

As an approximation to a Dirichlet Process Mixture Model, BHC defines a predictive distribution on new data. Specifically, the tree is a mixture model with each node a component:

$$p(\mathbf{x} \mid \mathcal{D}) = \sum_{k \in \mathcal{N}} \omega_k p(\mathbf{x} \mid \mathcal{D}_k), \quad (1)$$

where \mathcal{N} is the set of all nodes in the tree, ω_k is the weight of the component related to its merge likelihood, and $p(\mathbf{x} \mid \mathcal{D}_k)$ is the posterior predictive distribution associated with the data under node k (in our case, a set of independent Beta-Bernoulli models predicting the probability of occurrence of each frame).

We use this joint distribution to evaluate the conditional probability of a verb taking a specific frame when conditioned on the other frames of the verb. For verbs with multiple senses, we take the maximum of the predictions across all associated VerbNet classes, under the assumption that people interpret a verb with only one sense—its most likely one—and coerce with that sense.