Introduction

Why can Sally like to read but not *like to read?\footnote{Most verbs can only be used with a few \textit{syntactic frames}}

Most verbs can only be used with a few \textit{syntactic frames}. Theories of argument structure \cite{pinson1989complex} agree on \textit{hierarchical verb classes} with similar semantics \& syntax.

Questions:

Q1 Is a class hierarchy supported by the patterns of verbs and frames?\footnote{We compared this to the hierarchy obtained on new data.}

Q2 Does a class hierarchy predict generalization and productivity?

Two hierarchies of verbs

VerbNet\footnote{6334 verbs x 1613 frames, incrementally hand-crafted starting with Levin (1993)}

Bayesian Hierarchical Clustering\footnote{(Kipper et al., 2008)}

VerbNet classes, under the assumption that people have different meanings depending on what we consider to be VerbNet’s flat representation of frames.

Comparison: Tanglegram

Tanglegrams\footnote{(e.g. Scornaecca et al., 2011)} visualize similarity of hierarchies by minimizing crossings of connectors between leaves.

Entanglement\footnote{(Ghahramani, 2005)}

normalized measure of total crossings of tanglegram (random = 0.66)

A1 Entanglement: low, good qualitative alignment, similar structure.

Comparison: H and C

Homogeneity $H$ and Completeness $C$.\footnote{(Rosenberg and Hirschberg, 2007)}

information-theoretic measures of clustering similarity.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Granularity</th>
<th>BH C Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>$H_{\text{gran}}$</td>
<td>88.31</td>
</tr>
<tr>
<td></td>
<td>$H_{\text{rand}}$</td>
<td>88.19</td>
</tr>
<tr>
<td></td>
<td>$H_{\text{class}}$</td>
<td>83.34</td>
</tr>
<tr>
<td>$C$</td>
<td>$C_{\text{gran}}$</td>
<td>72.31</td>
</tr>
<tr>
<td></td>
<td>$C_{\text{rand}}$</td>
<td>99.14</td>
</tr>
<tr>
<td></td>
<td>$C_{\text{class}}$</td>
<td>41.37</td>
</tr>
</tbody>
</table>

$H = 1 \Rightarrow$ all BH clusters have only members of a single VerbNet class.\footnote{$H = 1 \Rightarrow$ all BH clusters have only members of a single VerbNet class.}

$C = 1 \Rightarrow$ members of a VerbNet class always in the same BH cluster.

A1 High similarity across VerbNet, BH converges on similar structure.

Coercion: prediction

Do hierarchies predict meaningful dimensions of generalization?

VerbNet

Given a frame, categorize a verb as:
- Exact, if the verb can take the frame
- Sibling, if a verb’s super/sub class can take the frame
- None, otherwise

BHC

Posterior predictive distribution

$p(x | D) = \sum_{k \in S} p(x | D_k)$

node weight (likelihood + node size)

indep. Beta-Bernoullis predicting frames

Segments that could be added or removed: selectional restrictions (e.g., prepositional literals).\footnote{Vikash Mansinghka, Patrick Shafto, Eric Jonas, Cap Petschulat, Max Gasner, and Joshua B. Tenenbaum. (2011). Learnability and Cognition. MIT Press, Cambridge, MA.}

Varying frame representations

What’s in a syntactic frame?\footnote{(cf. Hale and Keyser, 2002; Culicover and Jackendoff, 2005)}

Varying representation of frames by adding/removing annotations does not significantly change results (see paper for details).

Conclusion

A1 A shallow + disconnected hierarchy is well-supported by the distribution of verbs and frames in language.

A2 Hierarchies capture important features of argument structure by predicting generalization in a coercion task.

However:

1. Substantial variability in coercions unexplained by BHC or VerbNet
2. Varying frames doesn’t improve BHC $\Rightarrow$ limitation of simple hierarchy.

Use models that can instantiate more sophisticated theorems\footnote{e.g. CrossCat (Mansinghka et al., 2016) for cross-cutting categories}.

Acknowledgements

We thank Richard Savage for help with BHC and Tal Galieli with help with entanglement and the R package Biocplex. JM is supported by an undergraduate research fellowship from the Boston College Psychology Department and a Barry M. Goldwater scholarship.