The VerbCorner Project

Toward an Empirically-Based Semantic Decomposition of Verbs

Joshua K. Hartshorne¹, Claire Bonial², Martha Palmer²

1. Brain & Cognitive Sciences, Massachusetts Institute of Technology       2. Institute of Cognitive Science, University of Colorado at Boulder

VerbNet

A database of 8,537 verbs categorized into 273 classes based on syntactic frames

Semantic Predicates

- Emotional_Stress(E,X,Y)
- In_Reaction,Trial(X)
- Cause(X,E)
- Process(X,Y)
- Discharge(E,X,Y)
- Movement(X,Y)
- Duration
- Non(X)
- Post(X,Y)
- End(X)
- Start(X)
- Position(X,Y)
- Location(X,Y)
- Enter,Inside(X,Y,Z)
- Contain(X,Y,Z)
- Has_Possession(X,Y,Z)
- Transfer(X,Y)
- Future_Possession(X,Y,Z)

Applications in syntactic parsing and shallow semantics.

User Experience

User Home Page

The "Equilibrium" Task

Social networking elements

Tasks

Design
- Each verb presented in each licensed frame
- Subset of items "over-sampled"
- Novel NPs and additional VPs
  - (avoids world knowledge contamination)
- Fanciful backstory to prime intuitions and focus on entailment

Tasks (semantic predicate explored)
- A Good World (positive/negative valence)
- Entropy (change of physical state)
- Equilibrium (application of force)
- Explode on Contact (contact)
- Philosophical Zombie Hunter (mental state)
- Simon Says Freeze (movement)

Preliminary Results (First 6 Weeks)

Case Study: Equilibrium

Pilot experiment:
- 138 sentences (15 verbs + licensed frames)
- Kripp’s alpha = .76 for 8 annotators
- Modal response matched expected 137/138 sent.
- “Expected” is response matching researcher intuitions about involvement of semantic predicate in event semantics

On VerbCorner
- 194 over-sampled sentences (15-20 ann./each)
- Modal response matched expected 184/194 sent.
- In remainder, weak modal response

Prospects & Future Directions

- Result: Naïve participants can and will provide high-quality semantic annotations
- Future: Using annotator’s history to weight contributions
- Improve user experience
- Full validation of predicates, potentially modify VN semantics according to results
- Separate citizen science project to validate verb-frame pairings

Limitation:

- Minimal empirical basis for semantic predicates.
- 146 semantic predicates X 1363 (sub)class-frame combinations => massive coding project

Solution: Crowdsourcing with Citizen Scientists

- Recruit volunteer “citizen scientists” to help with coding
- Distinguished from GWAPs: Citizen scientists are interested in (a) learning & (b) advancing science & are willing to tackle more difficult projects

Acknowledgements

We gratefully acknowledge the support of the Ruth L. Kirschstein NRSA program at the National Institute of Health and the National Science Foundation Grant NSF-IIS-1116782, A Bayesian Approach to Dynamic Lexical Resources for Flexible Language Processing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Now widely used in astronomy and zoology (cf. Zooniverse.org)