

Coverage and Inheritance in The Preposition Project

Ken Litkowski
CL Research
9208 Gue Road
Damascus, Maryland USA
ken@clres.com

Orin Hargraves
Lexicographer
5130 Band Hall Hill Road
Westminster, Maryland USA
orink@carr.org

Overview of Project Objective

- Create publicly available databases that characterize English prepositions based on
 - Disambiguated prepositions in FrameNet sentences
 - Using a lexicographically-based sense inventory (Oxford Dictionary of English)
 - Within a semantic framework of traditional English grammar (Quirk et al., 1985)

The Annotation Process

- *Use the Oxford Dictionary of English (ODE) senses*
 - *373 prepositions (including 220 phrasal prepositions)*
 - *About 847 senses (grows as evidence indicates)*
- *Lexicographer describes each sense with*
 - *Semantic role name*
 - *Syntactic and semantic properties of its complement and attachment point*
 - *Syntactic function and meaning (from Quirk)*
- *Tags FrameNet instances with senses*

Tagging Progress To Date

- Instances completed for 17 prepositions
 - *About, across, against, as, at, between, by, for, from, in, into, of, on, over, through, toward(s), with*
- Covers 218 senses out of initial 847 (with 20 added by lexicographer when analyzing instances)
- 60 prepositions not present in ODE (removed from original inventory in the first edition)
 - Many variant spellings (e.g., **abaht** for **about**)
 - Mostly phrasal prepositions (e.g., **to the accompaniment of**) no longer viewed as sufficiently idiomatic to warrant a distinct entry as a phrase (e.g., under **accompaniment**)

Questions About Usability of Current Results

- With so few prepositions analyzed thus far, how complete is our effort and can we extrapolate from what is presently available?
- How thoroughly does the present data (about 25 percent of the senses) cover the semantic space of prepositions?
- Can inheritance be used to impute meaning (semantic roles and characteristics of the complement and point of attachment) to unanalyzed senses?

Coverage of the Current Analyzed Senses

- Linguistics literature replete with lists of semantic roles
 - Gildea & Jurafsky; O'Hara compilations from Penn Treebank, FrameNet, OpenCyc, Factotum; Boonthum et al. assessment of Jackendoff, Dorr's LCS preposition database, and Barker's analysis
 - Quirk et al. is arguably the most complete (with 50 paragraphs laying out the semantic space, mostly spatial and temporal meanings)
- BUT, linguistics literature is not data-driven

Correspondence of TPP Semantic Roles to Quirk Paragraphs

- TPP sense analysis identifies applicable Quirk paragraph
 - Only half of the 238 senses could be linked
 - Coverage is almost complete (47 of 50 paragraphs)
 - Unlinked senses can be grouped to identify gaps in Quirk
 - Large group of senses (~30) pertaining to quantities
 - Large group of senses (~30) where complements provide points of reference
 - Remaining senses describe fine-grained meanings
- Not only does TPP cover semantic space, but identifies gaps in coverage

Using TPP Semantic Roles to Analyze the Spectrum of Preposition Semantics

- O'Hara analysis of *at* reveals difficulty of traditional assessments of semantics
 - Uses less comprehensive sense breakdowns (Penn Treebank and FrameNet frame elements alone)
 - Exposes significant differences in frequencies and difficulty of mapping frame elements into semantic roles
- TPP data-driven approach with a well-developed sense inventory allows new senses to be identified
- TPP method of assigning semantic role names locally (instead of *a priori*) provides the data for a subsequent **rationalization** of semantic roles

Inheritance Within the Preposition Sense Inventory

- With decreasing numbers of instances for remaining prepositions and few likely for phrasal prepositions, can we extend current results?
- Litkowski (2002) laid out a graph-theoretical inheritance hierarchy
 - BUT, does it work and what does inheritance mean?
 - How can we draw the analogy to hypernymic inheritance?

Mapping the Inheritance Hierarchy

- Inheritance is based on a terminal preposition in the definition (**about**, *on the subject of*)
 - 621 senses follow this structure
 - 62 senses do not have a terminal preposition and 164 are usage notes (viewed as primitives)
- Analysis of first 13 prepositions completed
 - 11 without terminal preposition
 - 100 senses with usage note definitions
 - 411 of 621 with terminal prepositions end in one of these 13 prepositions
 - Remaining senses are mostly contained in definitional cycles described in Litkowski (20020)

The Meaning of Inheritance

- Begin by substituting the definition for the preposition
 - A book **about** ancient Greece becomes a book **on the subject of** ancient Greece
 - Consider **about** *isa* **of**: note that **of** links **subject**, not **book**
 - Inheritance plausibly involves only the semantic relation name and the properties of the complement, but not the properties of the point of attachment
- Hypothesis: *the semantic relation name and the complement properties of an inherited sense are more general than those of the inheriting sense*

Testing the Inheritance Hypothesis

- Lexicographer was asked to disambiguate 451 definitions in 411 senses with a terminal preposition among the 13 analyzed
 - Lexicographer was very reluctant, since definitions of phrasal preposition frequently substituted one phrase for another (**in the name of** defined as *for the sake of*)
 - Tagging was feasible (i.e., lexicographer was able to make a judgment based on very little context)
 - Of 451 definitions analyzed, 48 were in senses of the 13 prepositions already analyzed
- Table arraying putative hypernymic semantic relations and complement properties for each definition analyzed supports hypothesis

Making Use of Inheritance

- Verification of inheritance hypothesis allows continued use of digraph analysis approach
 - With pointers to specific senses (rather than treating nodes as undifferentiated agglomerates of all senses), digraph analysis can disentangle large definitional cycles
- Definitive digraph analysis should await further refinement of semantic role names and properties of complements and points of attachment

Modifications to the Lexicographer's Steps in The Preposition Project

- Lexicographer is now disambiguating senses having terminal prepositions as each preposition is analyzed
- Rationalization of semantic role names
 - Local decisions about names can now be examined *in toto* to identify duplicates and minor name variants
 - Similar semantic roles can be grouped (e.g., in the time space) and organized more logically and consistently
- Lexicographer is now characterizing senses (names, complement and attachment properties) for prepositions without instance sets
 - To be viewed as preliminary until instance sets are available
 - Will permit “rationalization” and inheritance to be used sooner

Refining the Lexicographic Analyses

- Application of digraph analyses will allow closer examination of semantic role names and complement properties
 - Moving toward a WordNet-like hierarchy of prepositions
- Analysis of features has begun with FrameNet instance sets (e.g., decision trees)
 - Senseval-compatible files are being fully parsed for syntactic features
 - Using WordNet tops, Roget-style thesaurus categories, Oxford WordNet hierarchy for semantic characterizations

Summary and Conclusions

- The Preposition Project is continuing to create databases of useful characterizations of preposition behavior
- Analysis of TPP's coverage suggests not only consistency with other analyses of prepositional behavior, but also that TPP is identifying gaps in coverage
- Analysis of TPP data has permitted a characterization of inheritance behavior that will allow more rapid use of results