Abstract

English prepositions are extremely frequent and extraordinarily polysemous. In some usages they contribute information about spatial, temporal, or causal roles/relations; in other cases they are institutionalized, somewhat arbitrarily, as case markers licensed by a particular governing verb, verb class, or syntactic construction. To facilitate automatic disambiguation, we propose a general-purpose, broad-coverage taxonomy of preposition functions that we call supersenses: these are coarse and unlexicalized so as to be tractable for efficient manual annotation, yet capture crucial semantic distinctions. Our resource, including extensive documentation of the supersenses, many example sentences, and mappings to other lexical resources, will be publicly released.

Prepositions are perhaps the most beguiling yet pervasive lexicosyntactic class in English. They are everywhere; their functional versatility is dizzying and largely idiosyncratic (1). They are nearly invisible, yet indispensable for situating the where, when, why, and how of events. In a way, prepositions are the bastard children of lexicon and grammar, rising to the occasion almost whenever a noun-noun or verb-noun relation is needed and neither subject nor object is appropriate. Consider the many uses of the word to, just a few of which are illustrated in (1):1

1 a. My cake is to die for.
   b. If you want I can treat you to some.
   c. How about this: you go to the store
d. to buy ingredients.
e. Then if you give the recipe to me
f. I’m happy to make the batter
g. and put it in the oven for 30 to 40 minutes
h. so you’ll arrive to the sweet smell of chocolate.
i. That sounds good to me.
j. That’s all there is to it.

Sometimes a preposition specifies a relationship between two entities or quantities, as in (1g). In other scenarios it serves a case-marking sort of function, marking a complement or adjunct—principally to a verb (1b–1e, 1h, 1i), but also to an argument-taking noun or adjective (1f). Further, it is not always possible to separate the semantic contribution of the preposition from that of other words in the sentence. As amply demonstrated in the literature, prepositions play a key role in multiword expressions (Baldwin and Kim, 2010), as in (1a, 1b, 1j).

An adequate descriptive annotation scheme for prepositions must deal with these messy facts. Following a brief discussion of existing approaches to preposition semantics (§1), this paper offers a new approach to characterizing their functions at a coarse-grained level. Our scheme is intended to apply to almost all preposition tokens, though some are excluded on the grounds that they belong to a larger multiword expression or are purely syntactic (§2). The rest of the paper is devoted to our coarse semantic categories, supersenses (§3).2 Many of these categories are based on previous proposals—primarily, Srikumar and Roth (2013a) (so-called preposition relations) and VerbNet (thematic roles; Bonial et al., 2011; Hwang, 2014, appendix C)—but we organize them into a hierarchy and motivate a number of new or altered categories that make the scheme more robust. Because prepositions are so frequent, so polysemous, and so crucial in establishing relations, we believe that a wide variety of NLP applications (including knowledge base construction, reasoning about events, summarization, paraphrasing, and translation) stand to benefit from automatic disambiguation of preposition supersenses.

---

1 Though infinitival to is traditionally not considered a preposition, we allow it to be labeled with a supersense if the infinitival clause serves as a PURPOSE (as in (1d)) or FUNCTION. See §2.

2 Supersense inventories have also been described for nouns and verbs (Ciaramita and Altun, 2006; Schneider et al., 2012; Schneider and Smith, 2015) and adjectives (Tsvetkov et al., 2014). Other inventories characterize semantic functions expressed via morphosyntax: e.g., tense/aspect (Reichart and Rappoport, 2010), definiteness (Bhatia et al., 2014, also hierarchical).
A wiki documenting our scheme in detail can be accessed at http://tiny.cc/prepwiki. It maps fine-grained preposition senses to our supersenses, along with numerous examples. The wiki is conducive to browsing and to exporting the structure and examples for use elsewhere (e.g., in an annotation tool). From our experience with pilot annotations, we believe that the scheme is fairly stable and broadly applicable.

1 Background

The descriptive challenges raised by prepositions have not gone unnoticed in the literature; see, e.g., Saint-Dizier (2006a) for an assortment of syntactic and semantic issues. Here we touch on some of the lines of inquiry, resources, and NLP approaches to preposition semantics found in previous work.

1.1 Linguistic Approaches

Most studies of preposition semantics are limited to so-called “lexical” (essentially, spatiotemporal) usages. The lexical-vs.-functional dimension and, relatedly, the degree of association between prepositions and other words (especially verbs) used in combination has received some theoretical attention (e.g., Bolinger, 1971; Vestergaard, 1977; Jolly, 1993; Rauh, 1993; O’Dowd, 1998; Tseng, 2000). We draw on insights from this literature where possible, but find that many of the proposed diagnostics are insufficiently clear and robust for a general-purpose preposition annotation scheme.

The structured polysemy analysis of over put forward by Brugman (1981) and elaborated by Lakoff (1987, pp. 416–461), Dewell (1994), Tyler and Evans (2003, ch. 4), and others has been influential within cognitive linguistics. Working in this tradition, Lindstromberg (2010) examines over 90 English prepositions, considering the schematic spatial situations that can be expressed as well as their non-spatial extensions. Chapter 21 gives an inventory of about 75 “non-spatial notions”—these are not unlike the categories we will adopt below, though some are quite fine-grained: e.g., BEING RESOLVED, FIXED as in pin him down vs. BEING UNRESOLVED, UNDECIDED as in everything’s still up in the air. How well annotators could be trained to agree on Lindstromberg’s detailed categorization is unknown.

Crosslinguistic variation in adpositions and spatial categorization systems has received considerable attention from theorists (Bowerman and Choi, 2001; Hagège, 2009; Regier, 1996; Xu and Kemp, 2010; Zelinsky-Wibbelt, 1993) but is of practical interest as well, especially when it comes to machine translation and second language acquisition. A corpus creation project for German preposition senses (Müller et al., 2010, 2011) is similar in spirit to the supersense approach taken below. Finally, the PrepNet resource (Saint-Dizier, 2006b) aimed to describe the semantics of prepositions across several languages; however, it seems not to have progressed beyond the preliminary stages. Thus far, our approach has focused on English, but aims to define supersense categories semantically rather than by language-specific criteria (e.g., syntactic tests) so as to encourage its adaptation to other languages in the future.

1.2 Preposition Resources

The following corpus resources contain semantic categorizations that apply to English prepositions:

The Penn Treebank. As detailed by O’Hara and Wiebe (2009), the PTB since version II (Marcus et al., 1994) has included a handful of coarse function tags (such as LOCATION and TIME) that apply to constituents, including PPs.

FrameNet. Semantic relationships in FrameNet (Baker et al., 1998) are organized according to scenes, known as frames, that can be evoked by predicates in a sentence. Each frame defines roles, or frame elements, for components of the scene that can be elaborated with arguments in the sentence. Many roles are highly specific to a single frame, while others are quite generic. Arguments are often realized as PPs, thus the frame element labels can be interpreted as disambiguating the function of the preposition.

The Preposition Project (TPP). This is an English preposition lexicon and corpus project (Litkowski and Hargraves, 2005) that adapts sense definitions from the Oxford Dictionary of English and applies them to prepositions in sentences from corpora. A dataset for the SemEval-2007 shared task on preposition WSD (Litkowski and Hargraves, 2007) was created by collecting FrameNet-annotated sentences (originally from the BNC) and annotating 34 frequent preposition types (listed in (2) below) with a total of 332 attested senses. (The SemEval-2007 sentences—of which there are over 25,000,
each with a single preposition token annotated—were handpicked by FrameNet lexicographers and so are not a statistically representative corpus sample.) TPP now incorporates additional prepositions and resources, with new annotated corpora under development (Litkowski, 2013, 2014).

**Dahlmeier et al.** To learn and evaluate their joint model of semantic roles and preposition senses, Dahlmeier et al. (2009) annotated TPP senses in the PropBank WSJ corpus for 7 high-frequency prepositions (**of**, **in**, **for**, **to**, **with**, **on**, and **at**). This amounted to 3,854 statistically representative instances in the news domain. The inter-annotator agreement rate was estimated at 86%, which suggests that clearly applicable TPP senses are available for the preponderance of tokens, but gives little insight into TPP’s suitability for rare or borderline usages.

**Tratz.** Tratz (2011, ch. 4) refined the TPP sense inventory for the SemEval-2007 corpus with the goal of improving its descriptive adequacy and measuring inter-annotator agreement for all 34 prepositions. The total number of senses was reduced from 332 to 278, though a few prepositions gained additional senses.

**Srikumar and Roth (S&R).** Srikumar and Roth (2013b) modeled preposition token relations, i.e., the preposition’s governor, object, and semantic label. For their experiments, Srikumar and Roth coarsen the original TPP SemEval-2007 sense annotations into 32 categories determined semi-automatically (the fine-grained senses were clustered automatically, then the clusters were manually refined and given names). Detailed in Srikumar and Roth (2013a), those categories cut across preposition types to combine related TPP senses for better data-driven generalization. Cohen’s $\kappa$ for inter-annotator agreement was 0.75, which is encouraging, though it is unclear whether the disagreements were due to systematic differences in interpretation of the scheme or to difficulty with rare preposition usages. We shall return to this scheme in §3 below.

## 2 Our Approach

With the end of free-text semantic annotation in mind, we develop and document a preposition supersense tagset. Notably, we seek to include in our resource example sentences for each known preposition-supersense pairing; these examples should be particularly useful for assisting human annotators.

Before discussing the supersense tagset, it is necessary to establish the scope of the phenomenon that our scheme aims to address.

### Preposition types

For brevity, we will sidestep the controversial aspects of defining “preposition”, and defer to Pullum and Huddleston’s (2002) broad definition of a lexical class including words such as **to**, **for**, **of**, and **up**, whether they take an object (forming a transitive PP) or act as a non-idiomatic adverbal particle (e.g., *lift the book up*).

In documenting the supersense categories thus far,
Figure 1: The full supersense hierarchy: categories in large font (CONFIGURATION, ACTOR, UNDERGOER, etc.) are top-level categories, and their subcategories extend outward. Colors emphasize the different levels of the hierarchy.

our attention has been focused on the 34 preposition types annotated in the SemEval-2007 data (§1.2):

(2) about, above, across, after, against, along, among, around, as, at, before, behind, beneath, beside, between, by, down, during, for, from, in, inside, into, like, of, off, on, onto, over, round, through, to, towards, with

Of the 332 fine-grained TPP senses for these 34 prepositions, 285 have been mapped to one or more supersense categories; preliminary annotation suggests that these account for the vast majority of preposition tokens in corpora (the remaining senses are generally infrequent). Our resource further includes the full set of TPP sense definitions, bringing it to a total of 309 preposition types and 797 senses, though most senses for these new prepositions have not yet been assigned to a supersense.

**Multiword expressions.** Multiword expressions functioning as prepositions (e.g., *out of, except for*) receive a supersense as a unit, as do PP multiword expressions (*on fire, on the run, out of one’s mind*). However, in other cases where a preposition belongs to a multiword expression, it is generally excluded from receiving a preposition supersense label. Verbal expressions like *make up ‘invent’, come to ‘regain consciousness’, and take someone for some-

---

3 A majority of TPP types are multiword prepositions (e.g., *all over*). Many of the single-word prepositions are archaic, orthographically nonstandard, or rare beyond specialized domains.
Many of the categories in those schemes overlap (or nearly overlap) with S&R labels. Others include semantic categories that are absent from S&R, but appropriate for English prepositions. Table 1 compares the three inventories. The new hierarchy, comprising 73 preposition supersenses, appears in the table, and also in figure 1.

We modified S&R categories where possible to be more closely compatible with the other schemes. On a descriptive level, this allows us to take advantage of the linguistic analyses and explanations motivating is no longer covered by INSTRUMENT. S&R’s EXPERIENCER category has been removed (it is substantially different from the supersense and VerbNet categories of the same name). OBJECT OF VERB, OPPONENT/CONTRAST, PHYSICAL SUPPORT, and SEPARATION have also been removed.

Rough mappings from remapped S&R categories to supersenses: CAUSE → CAUSER, EXPLANATION → CO-PARTICIPANTS → CO-AGENT, CO-PATIENT, CO-THEME; VIA → COURSE, TRANSIT; MEDIUM OF COMMUNICATION → VIA; NUMERIC → VALUE; PARTICIPANT/ACCOMPANIER → ACCOMPANIER; PART WHOLE → PARTITIVE, WHOLE. MEANS

<table>
<thead>
<tr>
<th>ACTOR</th>
<th>TEMPORAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>STIMULUS</td>
<td>FREQUENCY</td>
</tr>
<tr>
<td>CAUSER</td>
<td>DURATION</td>
</tr>
<tr>
<td>AGENT</td>
<td>AGE (∈ ATTRIBUTE)</td>
</tr>
<tr>
<td>CREATOR</td>
<td>TIME</td>
</tr>
<tr>
<td>UNDERGOER +1</td>
<td>RELATIVE TIME</td>
</tr>
<tr>
<td>ACCOMPANIER</td>
<td>START TIME</td>
</tr>
<tr>
<td>ACTIVITY +1</td>
<td>END TIME</td>
</tr>
<tr>
<td>BENEFICIARY</td>
<td>DEICTIC TIME</td>
</tr>
<tr>
<td>THEME</td>
<td>CLOCK TIME/CXN</td>
</tr>
<tr>
<td>CO-THEME</td>
<td>CIRCUMSTANCE</td>
</tr>
<tr>
<td>TOPIC</td>
<td>EXPLANATION</td>
</tr>
<tr>
<td>PATIENT</td>
<td>RECIPROCAL</td>
</tr>
<tr>
<td>CO-PATIENT</td>
<td>PURPOSE</td>
</tr>
<tr>
<td>EXPERIENCER</td>
<td>FUNCTION (∈ ATTRIBUTE)</td>
</tr>
<tr>
<td>PROFESSIONAL ASPECT</td>
<td>MEANS (∈ ACTIVITY)</td>
</tr>
<tr>
<td>PLACE</td>
<td>VIA (∈ PATH) +2</td>
</tr>
<tr>
<td>LOCUS</td>
<td>MEANS (∈ MEDIUM)</td>
</tr>
</tbody>
</table>

Table 1: The supersense hierarchy and its mappings to the S&R inventory, VerbNet thematic role hierarchy, and AMR non-core roles. Supersenses with multiple parents appear with one of them in parentheses; supersenses with n children listed under some other parent have a +n designation. ∈ indicates that the supersense maps to an S&R category with the same name; likewise for ∈. S&R’s EXPERIENCER category has been removed (it is substantially different from the supersense and VerbNet categories of the same name). OBJECT OF VERB, OPPONENT/CONTRAST, PHYSICAL SUPPORT, and SEPARATION have also been removed.
those categories. On a practical level, this will make it easier to combine resources (lexicons and annotated corpora enriched with semantic role labels).

Following VerbNet, our preposition supersense categories are organized into a hierarchical (multiple inheritance) taxonomy. Not only does this explicit some of the distinctions between related categories that were described textually in S&R (e.g., the relationship between \textsc{StartState} and \textsc{Source}), but it also provides a practical strategy for annotators who are unsure of how to apply a category—there is often a less specific label to fall back on.

The preposition label set proposed here is noticeably larger than the supersense inventories for other parts of speech (fn. 2). This might warrant concern that it will be too difficult for annotators to learn. However, there are arguments in favor of a larger set when it comes to prepositions. First, because prepositions range from the lexical to the grammatical, they perhaps cover a wider/higher-dimensional semantic space than verbs or nouns. Thus, more categories might be needed for comparable descriptive adequacy. Second, the hierarchy should help guide annotators to the right category or small set of related categories. They will not have to consider all of them one by one. Moreover, the presence of more and less abstract categories gives annotators flexibility when they are uncertain. Finally, because prepositions are closed-class, we envision that the annotation process will be guided (to a much greater extent than for nouns and verbs) by the word type. Having several dozen categories at multiple levels of granularity means that the number of prepositions associated with most categories is small.\textsuperscript{5} For TPP prepositions (with fine-grained senses mapped to the new scheme), it will be possible to suggest a filtered list of supersenses to the annotator, and these should suffice for the vast majority of tokens. It may even be desirable to annotate a corpus by type rather than by token, so the annotator can focus on a few supersenses at a time.

Based on preliminary rounds of annotation—a mix of type-driven and token-driven—by several annotators, we are optimistic that the general approach will be successful. The preliminary annotation has also uncovered shortcomings in the annotation guidelines that have informed revisions to the categories and hierarchy. More extensive annotation practice with the current scheme is needed to ascertain its adequacy and usability. Should the size of the hierarchy prove too unwieldy, it will be possible to remove some of the finer-grained distinctions.

Below, we examine some of the areas of the hierarchy that have been overhauled.

\subsection{3.1 Temporal Refinement}

In S&R, all temporal preposition usages fall under a single label, \textsc{Temporal}. VerbNet is slightly more discriminative, with an equivalent \textsc{Time} supercategory whose daughters are \textsc{Initial_Time}, \textsc{Final_Time}, \textsc{Duration}, and \textsc{Frequency}.

We have refined this further (figure 2) after coming to the conclusion that the major temporal prepositions

\textsuperscript{5}Currently, only 9 preposition types are mapped to more than 10 supersenses: \textit{for} and \textit{by} (20 each), \textit{of} (18), \textit{to} and \textit{in} (16), \textit{with} (15), \textit{at} and \textit{on} (13), and \textit{from} (11). 20 have 4–9 supersenses.
cluster neatly into finer-grained subcategories. Relations that situate a time as before or after another time are under RELATIVE TIME; special cases are START TIME, END TIME, times implicitly situated relative to the present (DEICTIC TIME), and constructions for telling time that express an offset in minutes relative to the hour (CLOCK TIME). We also follow AMR’s lead in creating a dedicated AGE category, which inherits from TEMPORAL and ATTRIBUTE.

Given that most of the prepositions in figure 2 are only associated with one or two temporal supersenses (only in and at are known to occur with three), we do not expect that the subcategories will impose too much of a burden on annotators.

### 3.2 Paths

Extensive discussion has gone into a section of the hierarchy for paths, which were not accounted for to our satisfaction in any of the existing schemes (due to unclear boundaries between the categories). Our analysis draws upon recent studies of caused motion constructions in the context of improving their treatment in VerbNet. Those studies address the basic scenarios of CHANGE OF LOCATION, CHANGE OF STATE, TRANSFER OF POSSESSION, TRANSFER OF INFORMATION, and CHANGE IN VALUE ON A SCALE with regard to their syntactic and semantic argument structures (Hwang et al., 2014; Hwang, 2014, ch. 5). Figure 3 shows our subhierarchy for paths, which is closely related to the approach adopted for VerbNet, but in some respects more detailed. Taking PATH to be the intermediate part of literal or abstract/metaphoric motion, we distinguish subtypes:

- **Traversed**: A stretch of physical space that the mover inhabits during the middle of motion (not necessarily where the event as a whole is located, which would be marked with a simple LOCATION preposition). This category is a subtype of LOCATION as it describes the “where” of the intermediate phase of motion. It is further refined into:
  - 1D Trajectory: A 1-dimensional region of space that is traversed, such as by following a path or passing a landmark. E.g.: walk along the river, over the bridge, past the castle
  - 2D Area: The 2-dimensional region of space that is “covered”, though there is less of a notion of completeness than with a 1-dimensional trajectory: I walked about/around the room
  - 3D Medium: Volumetric material that the figure moves through, and which may exert a facilitatory or opposing force on the figure: I waded through the swamp
- **Direction**: This covers prepositions marking how the motion of the figure, or the figure itself, is aimed/oriented (by contrast with DESTINATION, where the preposition expressly indicates an intended endpoint of motion): walk toward the door, kick at the wall, toss the ball up.
- **Contour**: This describes the shape, but not the location, of a path; it is also a kind of MANNER: walk in a zigzag.
- **Extent**: Also a subtype of VALUE, this is the size of a path—the physical distance traversed or the amount of change on a scale: ran for miles.
- **Via**: Prepositions in this category mark something that is used for translocation, transfer, or communication between two points/parties. It is a subtype of PATH because it pertains to the intermediate phase of (literal or figurative) motion, and also a subtype of INSTRUMENT because it is something used in order to facilitate that motion. S&R used the label Via for the spatial domain and MEDIUM OF COMMUNICATION for communication devices; we instead use the Via supersense directly for cases that are not physical motion, e.g.: talk by phone; talk on/over the phone; make an appearance on TV; order by credit card via/on the Internet; I got the word out via a friend. Enablers expressed metaphorically as paths, e.g. Hackers accessed the system via a security hole, are included as well. There are two subcases:
  - **Transit**: The vehicle/mode of conveyance that facilitates physical motion traversing a path. It is also a subtype of LOCATION because it specifies where the figure was during the motion: go by plane.
  - **Course**: The roadway or route that facilitates physical motion traversing a path. It is also a subtype of 1D Trajectory because it specifies a 1-dimensional path for the figure’s motion: drive via back roads.
For spatial usages of certain prepositions that portray static scenes as motion ("fictive motion"; Talmy, 1996), an argument could be made for either the locative or path categories. Our conventions are:

- With a figure whose shape/spatial extent is being described with respect to a landmark:
  - \textsc{1Dtrajectory} for the extent of a 1-dimensional shape: \textit{a cable runs above the duct; the bridge [that goes] across the river}
  - \textsc{2darea} for the extent of a 2-dimensional shape: \textit{Her hair was in plaits about her head}
  - \textsc{initiallocation} for the "starting point": \textit{a road which runs from Exop into the hills; single wires leading off the main lines}
  - \textsc{destination} for the "ending point": \textit{every driveway to the castle was crowded}

- For the spatial orientation of a figure: \textsc{direction}: \textit{they faced away from each other}
- Suggesting the spatial path that may be traversed to access a place starting from a reference point (such as the speaker's location): \textsc{location}: \textit{in a little street off Whitehall; He must have parked around the front of the motel; the auditorium is through a set of double doors}
- For a physical path of perception (line of sight, hearing, etc.): \textsc{1dtrajectory}: \textit{Lily peeped around the open curtain; glance over her shoulder}
- For a perspective in perception or communication: \textsc{location}: \textit{I can see Russia from my house; views over Hyde Park; she rang him at home}

3.3 Communication

English systematically invokes language of motion and transfer to describe communication (Reddy, 1979). S&R includes a specific \textsc{mediumofcommunication} category, but its boundaries are not entirely clear. Similarly, AMR incorporates a \textsc{medium} role, though this conflates communicative mediums with what we have called \textsc{3dmedium} above. Instead, our definition of \textsc{via} (§3.2) includes instruments of communication but is slightly more general.

There are also cases where the preposition marks an entity involved in communication, without framing that entity as an intermediary between two parties:

(3) a. I got the scoop \textit{from} a friend/the Internet.
   b. I put it down \textit{on} paper.
   c. The answer is somewhere \textit{in} this book/room.
   d. The rumor spread \textit{around} the school.

Rather than create a proliferation of communication-specific categories, we apply the abstract categories \textsc{locus, source, and goal} for abstract communication, and \textsc{location, initiallocation, and destination} for communication with a concrete component (such as writing).

3.4 Accompaniment vs. Joint Participation

The preposition with is frustratingly promiscuous. It often marks an entity that is associated with a main entity or event; what is frustrating is that the nature of the association seems to lie on a continuum from physical copresence to active counterpart in an event:

(4) a. Tim prefers [tea \textit{with} crumpets].
   b. Tim sat \textit{with} his computer.
   c. Tim walked \textit{with} Lori.
   d. Tim talked \textit{with/to} Lori.
   e. Tim fought \textit{against/with} Lori.
   f. Tim fought \textit{against/#with} the idea.

S&R has \textsc{participant/accompanier} and \textsc{opponent/contrast}, but these miss the highly frequent case of talk \textit{with}, which involves a cooperative rather than adversarial activity. VerbNet, on the other hand, has roles \textsc{co-agent, co-theme, and co-patient} for “events with symmetrical participants”.\footnote{VerbNet defines \textsc{co-agent} as “Agent who is acting in coordination or reciprocally with another agent while participating in the same event” (VerbNet, p. 20).} We adopt the following supersense conventions:

- \textsc{accompanier} applies for (4a–4c), where the two participants are physically colocated or performing the same action in separate (but possibly inferentially related) events. Adding \textit{together} seems more natural for these: \textit{Tim walked/took together with Lori}.
- \textsc{co-agent, co-patient, and co-theme}, as in VerbNet, apply where both participants are engaged in the same event in the same basic capacity (4d, 4e).
- \textsc{theme} applies for (4f), where the thing being fought is not fighting back.

3.5 Values and Comparisons

Many prepositions can be used to express a quantitative value (measuring attributes such as a quantity, distance, or cost), to compare to another value, or to compare to something qualitatively. S&R define a broad category called \textsc{numeric} for preposition senses that mark quantitative values and classify some qualitative comparison senses as \textsc{other}. We have developed a finer-grained scheme.
Comparison/Contrast applies to qualitative or quantitative analogies, comparisons, and differentiations: e.g., he used to have a car like mine; he was screaming like a banshee; the club’s nothing to what it once was; the benefits must be weighed against the costs; the difference between income and expenditure; these fees are quite distinct from expenses. Where these are relative to a specific scale or ranking, the subtype Scalar/Rank is used. Qualitative Scalar/Rank examples include: place duty before all else; at a level above the common people; warm weather for the time of year.

Value captures points on a formal scale—prices start at $10; the drunken yobbo who turned up by the cartload; my car does ten miles to the gallon—plus prepositions used as mathematical operators.

Scalar/Rank and Value share a subtype, ValueComparison, for comparisons/differentiations on a formal scale—the hill was above/below sea level. A subtype of this, Approximator, is for cases such as We have over/about/around/in the vicinity of 3 eggs left and We have between 3 and 6 eggs left. Prepositional expressions under, more than, less than, greater than, fewer than, at least, and at most fit into this category as well. Note that these can all be paraphrased with mathematical operators: \( a < b \leq c \). Approximator applies regardless of the semantic type of the thing measured (whether it is a spatial extent, temporal duration, monetary value, etc.).

### 3.6 Manner and Means

In our supersense hierarchy, we place Manner as a parent of Instrument (see figure 3). We also propose to distinguish Means for prepositions that mark an action that facilitates a goal (S&R include these under Instrument). We define Means as a subtype of both Instrument and Activity.

Manner and its subcategories are for prepositions that mark the “how” of an event: How did she lecture? With enthusiasm (Manner); How did he break up the anthill? With a stick (Instrument);

*How did they retaliate? With vicious shootings (Means); How did we coordinate? Over Skype (Via); How did you drive? In a zigzag (Contour).*

### 3.7 Other Major Changes

Space does not permit a full accounting of our modifications to the S&R scheme, which also include:

- **Explanation** and Reciprocal, two new causal categories with names borrowed from FrameNet. Explanation is for secondary events introduced as contributing to the occurrence of the main event (e.g., he lied out of dishonesty/for fear of rejection), with special cases Purpose (what somebody wants to happen) and Reciprocal (what is being reacted to: he was admired/thanked/punished for his deeds).
- **Creator**, a new subtype of Agent that captures usages such as stories by/of A.A. Milne.
- **State**, covering (e.g.) on morphine/loff work, as a new supertype of Start/State and End/State.
- **Configuration**, a new top-level category for senses marking static configurational relationships between two entities (typically nominals). Subtypes: Whole (renamed from S&R’s Part/Whole), Species, Possessor, and new categories Partitive, Superset, and Elements.
- **Location** prepositions can be used with a verb of motion to indicate a resulting location: put the hat on the stool; go inside the house. S&R list such usages under Destination. We instead deem the preposition’s meaning as coerced by the verb, and label the preposition as Location (simplifying documentation and annotation). We reserve the Destination supersense for to, into, etc., which exclusively mark endpoints of motion when used spatially.

### 4 Conclusion

English prepositions are a challenging class, given that there are so many of them and they are put to so many uses. We have built on prior work to propose a new hierarchical taxonomy of preposition supersenses, so that their semantics can be modeled in a coarse WSD framework. Our resource documents each supersense with detailed explanations, fine-grained dictionary senses, example sentences, and (where possible) mappings to other resources. The taxonomy will hopefully port well to adpositions and case systems in other languages, though we have not investigated that yet. We have successfully piloted English corpus annotation with our resource, and a full-fledged annotation effort is underway.
Acknowledgments

Improved coarse semantic categories for prepositions are the result of ongoing collaborations; they reflect the efforts of ourselves and others including Tim O’Gorman, Katie Conger, Meredith Green, Archna Bhatia, Carlos Ramírez, Yulia Tsvetkov, Michael Mordovanec, Matt Gardner, Spencer Onuffer, and Nora Kazour, as well as helpful conversations with Ken Litkowski, Orin Hargraves, Michael Ellsworth, Ed Hovy, Lori Levin, the Berkeley FrameNet group, and the AMR design group. We also thank anonymous reviewers for helpful comments. This research was supported in part by a Google research grant for Q/A PropBank Annotation, NSF CAREER grant IIS-1054319, and DARPA grant FA8750-12-2-0342 funded under the DEFT program.

References


Jena D. Hwang, Annie Zaenen, and Martha Palmer. 2014. Criteria for identifying and annotating caused motion constructions in corpus data. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno,


Leonard Talmy. 1996. Fictive motion in language and “ception”. In Paul Bloom, Mary A. Peterson, Nadel


