Natural Language Engineering **15** (3): 379–414. © 2009 Cambridge University Press
doi:10.1017/S1351324908005044 Printed in the United Kingdom

A corpus-based analysis of argument realization by preposition structures

QIBO ZHU

Statistics Canada, RHC-10Q, 150 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6, Canada and Institute of Cognitive Science, Carleton University, Room 2201, Dunton Tower, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6, Canada

e-mail: Qibo.Zhu@statcan.ca; qzhu5@connect.carleton.ca

(*Received 7 May 2008; revised 15 September 2008; first published online 3 February 2009*)

Abstract

This article studies the issue of argument realization by preposition structures. By examining the preposition structures that are marked as frame elements in FrameNet, the article attempts to give corpus-based attestations to the hypothesized link between deep semantic arguments and their surface syntactic representations. Problems addressed in this article include how argument realization by preposition structures can be predictable from the target lexical unit and the frame it evokes, and why some noncentral prepositions get selected in the argument realization options. The investigation is primarily inspired by Fillmore's work in frame semantics. The source data for this study is derived from a preposition knowledge base that we have recently built by extracting all the semantically annotated preposition structures in FrameNet. The analysis shows that while there are various semantic–syntactic mapping possibilities, for most semantic arguments, the tendency of using central prepositions in their realization expressions is very strong. This is a clear indication that some preposition structures are linked to certain semantic arguments more than they are to others. A similar experiment was conducted using the annotated PropBank corpus to corroborate the supporting evidence found in FrameNet. The results of this study, together with the syntactic–semantic mapping lists of preposition structures can provide raw linguistic data for the study of preposition semantics, lexicography, argument realization, word sense disambiguation, and natural language understanding.

1 Introduction

Preposition structures are common modifiers of noun phrases, verbal phrases, adjectival phrases, or clauses and sentences as a whole. In interacting with and modifying other linguistic elements, prepositions can play a very important role in the syntactic realizations of semantic arguments. Although traditionally, prepositions have not been much studied (Trujillo 1992; Zelinsky-Wibbett 1993), there is now an increasing amount of work devoted to the study of prepositions in linguistics and computational linguistics because of the importance of prepositions in linguistic studies and in natural language understanding and processing (O'Hara and Wiebe 2003; Litkowski and Hargraves 2007). Some of the investigations are cognitive science related studying how prepositions are represented in the mind (Herskovits

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

1986; Lakoff 1987; Langacker 1987; Jackendoff 1990; Tyler and Evans 2001, 2003). Some integrated the study of prepositions into the study of verb syntactic alternations and verb argument realizations (Levin 1993), and some are computational linguistics attempts at the disambiguation of prepositions for the purpose of natural language processing and machine translation (Sumita, Furuse and Iida 1993; Nübel 1996; Ye and Baldwin 2007). All these investigations are, to a certain extent, associated with the meaning-form correspondence between the semantic arguments and the preposition structures.

The present study aims at investigating, with corpus data evidence, the hypothesized link between semantic arguments and their syntactic expressions. The linguistic data source consists of corpus-derived sentences in the preposition knowledge base we built from the Berkeley FrameNet database. The FrameNet project (Baker, Fillmore and Lowe 1998; Fillmore and Atkins 1998) is a relatively recent undertaking to provide reliable annotated semantic–syntactic information for a good proportion of the lexicon in contemporary English. It is arguably one of the best manually annotated word semantics resources for the English language now. It is 'corpus-grounded and frame-based' (Fillmore, Ruppenhofer and Baker 2004) and has its theoretical foundation in Fillmore's early work on the study of semantic cases (Fillmore 1968) and his continuing work on semantic frames (Fillmore 1982; Petruck 1996). One of the assumptions in frame semantics is that there are generalizations we can make about the link between cognitive arguments and the linguistic constituents of the sentence. In his seminal publication, *The Case for Case*, Fillmore (1968) proposed a universal set of case relationships between semantic roles and their surface syntactic structures. To Fillmore, there are elaborate meaning-form associations in our language system. Forms are usually paired with the semantic components. Fillmore hypothesized particularly about the correspondence between semantic roles and prepositions:

> The rules for English prepositions may look something like this: the A(gent) preposition is by; the I(nstrument) preposition is by if there is no A, otherwise it is with; the O(bject) and F(actitive) prepositions are typically zero(0); the B(enefactor) preposition is for; the D(ative) preposition is typically to .... (Fillmore 1966)

When the Berkeley FrameNet project was launched in 1997, an important goal was to provide, in principle, 'an exhaustive account of the semantic and syntactic combinatorial properties' of each selected target lexical unit (Fillmore and Atkins 1998) and to represent the link between semantic and syntactic relations (Fillmore *et al.* 2004). In order to capture the semantic–syntactic correspondences, FrameNet focuses on two types of valence descriptions of lexical units (referred to as *LU* henceforth): the semantic valence description and the syntactic valence description. Semantic valence description is the description of the participating frame elements (referred to as *FE* henceforth) in different frames. Frames are cognitive structures that are legitimately instantiated in the language user's mental lexicon; a cognizer invokes a frame in order to make sense of his or her experiences. For each frame, there are different constituent frame elements. Frame elements are category labels indicating thematic role types. The syntactic valence description covers argument realizations

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

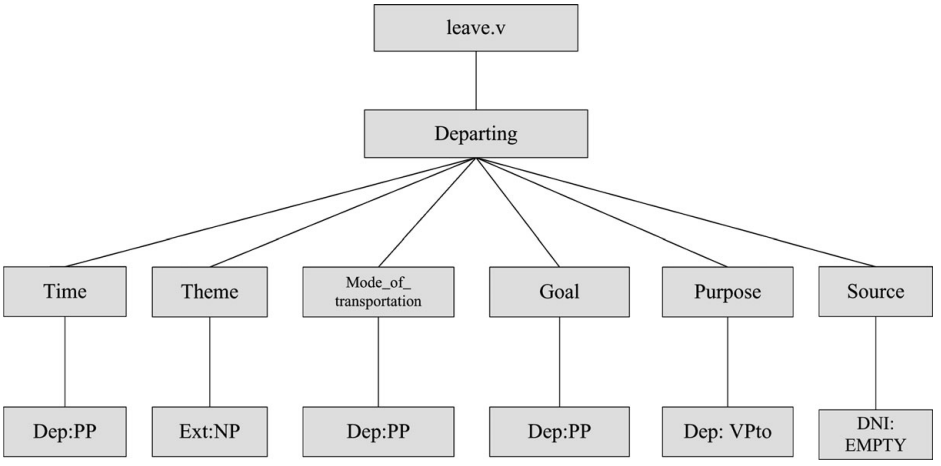


Fig. 1. Two types of valence descriptions for the lexical unit leave.v in the sentence 'On 30 May, Ho left by plane for France to work out a settlement.' The semantic valence description is a description of the semantic arguments – the frame elements. The syntactic valence description covers the argument realizations which include grammatical functions and the syntactic phrase types.

that can be decomposed to grammatical functions and the syntactic phrase types such as NP, VP, and PP etc. The two types of valence descriptions are interrelated and the basic assumption behind the reciprocal relation is that 'word meanings are organized around schematic conceptual scenarios, or frames,' and frames and frame elements underlie the use and interpretation of the lexical items (Fillmore, Johnson and Petruck 2003). Figure 1 illustrates how semantic arguments and their realizations are linked in their valence descriptions for the sentence 'On 30 May, Ho left by plane for France to work out a settlement.' In this example, the lexical unit *leave.v* invokes the frame *Departing* which in turn instantiates six participating frame elements: *Theme, Time, Mode_of_transportation, Goal, Purpose*, and *Source*. These frame elements are the semantic arguments for the frame-bearing lexical unit *leave.v*. Dep:PP, Ext:NP, Dep:PP, Dep:PP, Dep:VPto and DNI:EMPTY are the corresponding argument realizations for the six frame elements. The frame element *Source*, which indicates the starting point of a journey, is implicit in its surface syntactic realization (DNI:EMPTY), while *Time, Mode_of_transportation* and *Goal* have preposition structures in their argument realization structures (Dep:PP). The frame element *Theme* is represented by a noun phrase that serves as the subject of the sentence (Ext:NP) and the realization expression for *Purpose* is a verb infinitive phrase (Dep:VPto).

Frame elements have their roots in Fillmore's deep semantic case roles (Fillmore *et al.* 2003). In his initial model, Fillmore (1968) identified six basic semantic case roles to classify the verb types: *Agentive, Instrumental, Dative, Factitive, Locative,* and *Objective*. These six basic cases were later expanded to nine deep case structure categories: *Agent, Experiencer, Instrument, Object, Source, Goal, Location, Time*, and

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

```
<fn_prep sid="692542" file="lu10" frame="Duplication">
   <lu_type>copy.v</word_type>
   <lu_token>copied</word_token>
   <lu_def>COD:make a copy of.</word_def>
   <sentence>He copied it onto a clean sheet of paper.</sentence>
   <fe_sets>
      <fe_unit cid="1">
        <text>He</text>
        <pos>PNP</pos>
        <tp_fe_gf_pt>C:Creator:Ext:NP</tp_fe_gf_pt>
      </fe_unit>
      <fe_unit cid="2">
        <text>copied</text>
        <pos>VVD</pos>
        <tp_fe_gf_pt>:target:NONE:NONE</tp_fe_gf_pt>
      </fe_unit>
      <fe_unit cid="3">
        <text>it</text>
        <pos>PNP</pos>
        <tp_fe_gf_pt>C:Original:Obj:NP</tp_fe_gf_pt>
      </fe_unit>
      <fe_unit cid="4">
        <prep pid="1">onto</prep>
        <text>onto a clean sheet of paper</text>
        <pos>PRP:AT0:AJ0:NN1:PRF:NN1</pos>
        <tp_fe_gf_pt>E:Goal:Dep:PP</tp_fe_gf_pt>
      </fe_unit>
   </fe_sets>
</fn_prep>
```

Fig. 2. An example record from the FrameNet preposition knowledge base.

Benefactive (Fillmore 1971). In the current FrameNet framework, almost all of these deep case relations are included in the topology of frame elements.

For the proliferation of preposition constructions that are marked as frame elements in FrameNet, a phrase type such as PP, PPing, and PPinterrog is assigned to each of them (Ruppenhofer *et al.* 2006). PP is for ordinary prepositional phrases with nominal objects or for those verb particle constructions without objects. PPing is for prepositional phrases with gerundial objects. PPinterrog is assigned to prepositions that govern finite or nonfinite wh-clauses and verb phrases. Figure 2 is an example record from the preposition knowledge base, where a PP is marked as an FE.

In syntactic structures where the preposition is immediately preceded by an adverbial phrase such as *even in normal situations, still with your arms straight, far below them through the still waters, somewhere around there, soon after nine o'clock*, the intervening adverbs are included in the PP-phrase-type label as well.

In the present investigation, we aim at analyzing the relationship of preposition structures and the associated semantic roles they represent. This investigation will contribute to computational linguistics in two aspects. First, we provide fresh corpus evidence and statistical data to support the hypothesized link between semantic arguments and their surface syntactic realizations. Efforts in doing so may have been hampered in the past by the lack of a fine-grained typology of semantic roles and the lack of a sizeable natural language database that is semantically annotated with

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

detailed information about preposition structures and their semantic arguments. The syntactic–semantic mapping relationship of prepositions in the hypothesis can have implications in building theoretical models of natural language understanding, in prepositional attachments disambiguation and in automatic sense labeling. Second, we extended the research from a few basic case relations in the original hypothesis to more prepositional phrase patterns and semantic arguments. The research generated a dozen or so preposition semantics feature lists that can capture the mapping relation between preposition structures and semantic roles. These lists can have diversified practical applications. They can be good resources for examining the degree of preposition centrality in the syntactic realization of their associated frame elements. The lists can also help improve natural language understanding systems and machine translation systems.

2 Analysis of argument realization options

Argument realization by preposition structures in FrameNet is a phenomenon whereby frame elements are syntactically realized by linguistic structures introduced by different prepositions. The question to ask is: are there certain central prepositions that are consistently used more than others in representing certain frame element concepts? For example, in representing the frame element *Agent*, is the *by* phrase a primary choice? Are *with* and *by* typical prepositions for the introduction of the frame element concept *Instrument*? Is *for* the most frequently used preposition for the frame element *Beneficiary*? If there are such frame elements that are closely associated with particular prepositions, what are the frame elements and the corresponding prepositions? To what degree, are some prepositions central in representing some semantic arguments?

A total of 62 812 manually annotated sentences that contain preposition structures marked as frame elements were extracted from the FrameNet database to build a preposition knowledge base. There are 144 different preposition types including complex prepositions in the preposition knowledge base. Considering the total number of annotated sentences in FrameNet (139 439), it means that in every two to three annotated sentences, there is one sentence with prepositional phrases marked as frame elements (2.21:1). It is interesting to note that argument realizations in the form of prepositional structures are so marked in FrameNet that in many sentences, the majority of the annotated semantic arguments are realized by prepositional phrases. In FrameNet, there are 314 180 syntactic phrases marked as frame elements, out of which 70 790 are instances of frame elements that are realized by preposition phrases (referred to as *PP-FE* henceforth). This means that for every nine annotated frame elements in FrameNet, around two are PP-FEs (4.44:1). It can be reasoned that by studying the syntactic realization of arguments by preposition structures, we are studying a good proportion of the linguistic structures that actually get mapped in our cognitive mechanisms.

In processing the preposition structures that are marked as frame elements, we notice that for reasons of technical commitments (such as those connected with the inheritance hierarchy) in the earlier stages of the FrameNet project, each frame

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

element in a given frame is treated as distinct from every frame element in every other frame. For example, the frame element *Location* could have a completely unique name in each frame to satisfy the original technical requirements. For some frame element names, the FrameNet team chose to give related frame elements the same names for practical reasons in annotation. Also, sometimes FrameNet uses 'dotted names,' such as *Ingestion.Theme*. The difference in the naming of some frame elements can, to a certain extent, have some effects on the statistics.

Frame elements in the FrameNet database can be partitioned into 769 ontological categories, among which 640 can be represented by prepositional phrases. The most frequent frame element types with preposition structures in their argument realizations are in the conceptual fields indicating spatial and temporal properties, agents and themes, means and manners relations. In the FrameNet sentences that are populated with PP-FEs, the number of frame elements in a frame that a lexical unit can evoke is from one to thirty. The most PP-FE abundant lexical units are verbs of motion such as *traverse, pass, mount, leap jump, hop, descend,* and *cross*. On the other end of the argument structure spectrum, fourteen lexical units take only one frame element in their subcategorization scheme, and most of these lexical units are adjectives or adverbs such as *whole*, *reportedly*, *mere*, and *broken*. While it is true that these words usually do not take as many semantic argument structures as some verbs and nouns, this in no way means that none of these lexical units can have more frame elements, particularly noncore PP-FEs.

The correspondence between frame elements and preposition structures can be one to one, one to many, or many to one. In the process of analyzing the data sets in the preposition knowledge base, we produced some feature lists that can capture such syntactic–semantic correspondence relationships. The lists can be readily integrated in natural language processing systems.

2.1 One frame element to one preposition structure mapping

Prepositions can be semantically vacuous or polysemous. In many cases such as in verb particle constructions, they have little semantic content. Because of the semantic emptiness of prepositions, many search engine indexing agents include prepositions in the bag of stop word lists. In other cases such as in locative expressions, the information content they contain is not low. For example, in 'He drove down the road' the semantic weight of the sentence lies in the preposition. Because prepositions are highly promiscuous distributionally, dictionaries usually give sets of examples demonstrating polysemous uses and various meanings of prepositions. Also, prepositions do not translate well to other languages, so they are one of the major sources of difficulties for students who are learning English as a second or foreign language (Boers and Demecheleer 1998; Celce-Murcia and Larsen-Freeman 1999).

The semantically polysemous nature of prepositions has been much studied in the past (Radden 1989; Vandeloise 1994; Pütz and Dirven 1996; Cuyckens and Radden 2002; Tyler and Evans 2003). Linguists have been analyzing the literal senses of prepositions and the metaphorical meanings they enter into. In many cases,

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

the polysemy of prepositions can be explained by radial categories (Lakoff 1987) where many related senses of a preposition are connected to a semantic common core. Prepositions get their abstract senses from a primary proto-scene (Tyler and Evans 2003) or a semantic prototype (Wesche 1986) 'by means of generalization or specialization of meaning or by metonymic or metaphoric transfer' (Cuyckens and Radden 2002: xiii). For example, for the preposition *at*, the core meaning can be a simple point of location or reference as in *at the front door*. The following are some of the concepts that can be extended from the core meaning of *at*:

- Extended locative concepts: *the party at the beach*, *the meeting at Ottawa.*
- Extended intentional concepts: *to wait at the bus stop*, *to be at school.*
- Extended temporal concepts: *to wake up at midnight*, *to be busy at the moment*.
- Extended scale concepts: *to work at 20 dollars an hour*, *to drive at 50 mph.*
- Extended causal concepts: *to leave at my signal*, *to be happy at the arrival of the rescue ship.*
- Extended directional concepts: *to throw the ball at him, to shout at him, to grab at the purse, to shoot at the rabbit*.

To a certain extent, these conceptual extensions of *at* show that the semantic contents that a preposition can convey are not simply an arbitrary cluster of unrelated senses. They come from a highly schematic common core and constitute a unitary meaning. On the basis of this, we can assume that there can be a strong linking relationship between prepositional phrases and primary underlying meanings they represent. However, do we have corpus evidence to show that prepositions are associated with core meanings, and that some meanings can be represented by primary preposition? Are there correspondence patterns between frame elements and preposition structures? For grammar or syntax reasons, if we have to select a preposition structure that can be used with a specified lexical unit to represent a frame element concept, what prepositions are we most likely to choose? In other words, can we predict the selection of prepositions on the basis of the information about the frame element and the target lexical unit?

In addressing these questions, we analyzed the argument realization patterns for FrameNet sentences and observed some interesting cases where a few lexical words combined only with selected types of prepositions to represent some frame elements in the evoked frame. The prepositions are bound to target lexical words not in a way as strong as those traditionally called prepositional–verb constructions such as *depend on, account for*, and *look at*. The link between the preposition and the target lexical word seems to be much looser syntactically, but more widespread. In the prepositional–verb constructions, the connection between the preposition and the lexical verb is usually individually based, so that we say *to listen to somebody*, but not *to hear to somebody*. But in the one frame element mapping to one preposition structure case, the same preposition is selected not only for one individual lexical item, but also for a variety of target lexical units in the same frame. For instance, in the *Seeking* frame, the preposition *for* is used with almost all of the verbs such as *comb.v*, *feel.v*, *forage.v*, *fumble.v*, *grope.v*, *hunt.v*, *listen.v*, *look.v*, *pan.v*, *ransack.v*, *rummage.v*, *scour.v*, *seek.v*, and *sift.v* for the frame element *Sought_entity*. Here

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

are more examples of such one-to-one mapping relations where one designated preposition is consistently used for the realization of a particular frame element:

- Air: of 12 [breath.n, exhalation_act.n, inhalation.n]
- Alterant: with 12 [stain.v, treat.v]
- Arguer2: with 77 [altercation.n, argue.v, argument.n, bicker.v, disagreement.n, fight.v, quarrel.n, quarrel.v, quibble.v, row.n, row.v, spat.n, tiff.n, wrangle.v]
- Artist: by 15 [direct.v, film.v, produce.v, shoot.v]
- Assailant: by 16 [ambush.v, assault.v, attack.n, attack.v]
- Authorities: by 25 [apprehend.v, arrest.v, bust.v, cop.v, detain.v, intern.v, internment.n, nab.v]
- Beneficiary: for 18 [create.v, deal.v, fix.v, resolve.v, seek.v, solve.v]
- Complaint: of 13 [complain.v, complaint.n]
- Component_Sound: of 16 [babble.n, bellow.n, blare.n, boom.n, bray.n, burble.n, buzz.n, drone.n, peal.n, rap.n, titter.n, yammer.n]
- Controlling_Act: in 12 [authoritarian.a, indulgent.a, lenient.a, liberal.a, strict.a, tolerant.a]
- Conveyed_Emotion: of 19 [frown.n, grin.n, look.n, rictus.n, smile.n, smirk.n, sneer.n]
- Crime_Jurisdiction: to 28 [extradite.v, extradition.n]
- Employee: of 31 [dismissal.n, resignation.n, retirement.n, work.v]
- Executed: of 14 [execution.n, guillotining.n]
- Fact: by 11 [account.v, explain.v]
- Hair: of 12 [braid.n, curl.n, cut.n, lock.n, plait.n, ponytail.n, topknot.n, tress.n]
- Ingestor: by 17 [consume.v, devour.v, drink.v, eat.v, ingest.v, munch.v, nibble.v, sip.v]
- Inspector: by 15 [double-check.v, examination.n, examine.v, inspect.v, inspection.n]
- Interceptor: by 12 [intercept.v, interception.n]
- Investigator: by 15 [investigate.v, investigation.n, probe.n, under investigation.prep]
- Light: of 18 [flash.n, gleam.n, glimmer.n, glitter.n, glow.n]
- Mass: of 357 [abundance.n, amount.n, avalanche.n, degree.n, deluge.n, flood.n, heap.n, load.n, mass.n, measure.n, mite.n, modicum.n, myriad.n, oodles.n, ounce.n, pile.n, pinch.n, plethora.n, quantity.n, raft.n, stream.n, ton.n, torrent.n, touch.n, trickle.n, wave.n]
- Noisy_Event: of 76 [boom.n, burble.n, buzz.n, chatter.n, clack.n, clang.n, clatter.n, click.n, clink.n, creak.n, crunch.n, drone.n, gurgle.n, hiss.n, hum.n, patter.n, rap.n, rasp.n, roar.n, rustle.n, scrape.n, screech.n, slap.n, thump.n, tinkle.n, whine.n]
- Object: of 25 [area.n, bless.v, confession.n, consecration.n, depth.n, length.n, rite.n, sacrifice.n, width.n]
- Party_2: with 28 [agree.v, agreement.n, come together.v, meet.v, treaty.n]
- Preventing_Cause: by 21 [derail.v, foil.v, nip in the bud.v, thwart.v]
- Profiled_Item: by 20 [eclipse.v, equal.v, outdo.v, outshine.v, outstrip.v, rival.v, surpass.v]

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

- Purpose_Of_Recipient: for 11 [provide.v, supply.v]
- Sentence: to 15 [sentence.v]
- Set: of 13 [proliferation.n]
- Sought_Entity: for 86 [comb.v, feel.v, forage.v, fumble.v, grope.v, hunt.v, listen.v, look.v, pan.v, ransack.v, rummage.v, scour.v, seek.v, sift.v]
- Themes: of 31 [clash.n, clatter.n, click.n, clink.n]
- Transferors: between 11 [transfer.n]
- Vantage_Point: from 35 [conceal.v, concealed.a, cover.v, hidden.a, hide.v, mask.v, obscure.v, obscured.a, screen.v, screened.a]
- Weapon: with 12 [armed.a, assault.v, attack.v]

For each entry in the above list, a frame element name is followed by the primary preposition that is used to represent the frame element, the frequency (f) of this preposition in the realization of the FE concept, and the lexical units that are used with the primary preposition. The threshold for selection is $f > 10$. In these instances, there are outstanding patterns of correspondence between FEs and preposition structures. We observe, for example, that if we have to choose a preposition structure for the syntactic realization of the frame element *Weapon*, we are much more inclined to pick *with* to head the prepositional phrase. Thus, we say *to be armed with a weapon* and *to attack somebody with a weapon*. Similarly, we use *for* for the frame element *Beneficiary*. For the FE *Arguer-2*, the second party involved in an argument, we use the preposition *with*, no matter if the lexical unit is a noun such as *argument.n* and *altercation.n*, or a verb such as *bicker.v* and *wrangle.v*. For the two transferors involved in a transfer, we use *between* as in 'It seems that a similar ambiguity is common in Britain in relation to [THEME money] [TARGET transfers] [TRANSFERORS between kin], although we do lack up-to-date evidence on this point.' At the same time, evidence in this list shows that most of the *by* cases are not interesting in their own right. They are limited to passive uses of the verb, and it just happens to be the case that the FE that would have been Subject (Ext) in the active gets marked as *by* in the passive – but not at all because of the participation in a particular frame.

2.2 One frame element to multiple preposition structures mapping

Prepositions represent a closed class. On the one hand we have the richness of meaning; on the other hand we have the paucity of linguistic structures including prepositions to represent them. Thus, we can have a single preposition representing different conceptual relations, and a semantic role realized by multiple prepositions. While there are 214 frame element types that exhibit the one FE one preposition mapping relationship, there are almost twice as many ($f = 426$) FE types that manifest syntactic realization alternations when they use prepositional phrase representations. This means that most PP-FEs in FrameNet are represented by multiple preposition structures. Table 1 demonstrates the most frequent frame elements that have preposition phrases in their surface structures. The frequency threshold is $f \geq 500$ for these PP-FEs.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Table 1. *Frame element and preposition distributions in the most frequent PP-FE structures in FrameNet (threshold $f \geq 500$)*

FE	Prep_Tokens	Prep_Types	Mean	Std. Dev.
Goal	6 181	59	104.7627	307.0189
Path	4 687	41	114.3171	197.3984
Topic	2 873	28	102.6071	299.3783
Source	2 859	28	102.1071	315.3027
Time	2 513	38	66.1316	156.4402
Place	2 212	49	45.1429	170.9795
Reason	1 748	32	54.625	195.285
Theme	1 469	13	113	234.3918
Individuals	1 449	2	724.5	1 023.184
Manner	1 419	42	33.7857	92.3087
Stimulus	1 278	17	75.1765	132.4534
Area	1 203	38	31.6579	70.3148
Addressee	1 143	18	63.5	169.104
Means	1 046	22	47.5455	120.7491
Whole	1 036	9	115.1111	322.9928
Content	906	21	43.1429	92.8333
Ground	854	40	21.35	55.8664
Location	789	46	17.1522	39.0839
Stuff	695	2	347.5	490.025
Substance	681	4	170.25	329.8579
Category	644	15	42.9333	103.1569
Message	617	25	24.68	53.8166
Relative_location	613	33	18.5758	42.7434
Evaluee	597	19	31.4211	65.0404
Speaker	569	7	81.2857	125.1222
Direction	505	32	15.7813	16.0417

The FE column lists the frame element names. In FrameNet, target lexical units are classified on the basis of the semantic frames they belong to. In each semantic frame, there are participating FEs. The inventory and taxonomy of frames and FEs are created by the FrameNet group at Berkeley, and they are based on Fillmore's theory of frame semantics (Fillmore 1982). Some FEs are general across frames such as *Source* and *Goal*, although definitions of the FE in those frames can be different to a certain degree. Others are unique names of thematic roles specific to frames. The FEs in Table 1 are the most frequent semantic roles that are represented by prepositional phrases. Prep_Tokens refers to the total number of phrases that are headed by prepositions; Prep_Types is the number of different prepositions that are used to head the prepositional phrases. For example, the FE *Individual* is used across nine frames (Aggregate, Congregating, Dispersal, Gathering_up, Make_acquaintance, Quantity, Relation_between_individuals, Social_connection, and Unemployment). In 1 449 (Prep_Token) cases, this FE is represented by prepositional phrases, and for these 1 449 occurrences, only 2 (Prep_Types) prepositions are used. Std. Dev. stands for standard deviation.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

From the distribution of prepositions that head the preposition structures in argument realizations, we can see the diversity of syntactic configurations in selecting prepositions to represent frame element concepts. *Goal* in Table 1 is the frame element that is most frequently represented by prepositions (Prep_Tokens = 6 181). At the same time, it is also the frame element that can be represented by most of the preposition types (Prep_Types = 59). *Direction* is the frame element with the most even distribution (mean = 15.78 and Std. Dev. = 16.04) in the use of preposition structures in argument realizations. Although frame elements can have multiple argument realization possibilities, there is a strong tendency for the deep semantic arguments to consistently employ a few key prepositions much more than they do other prepositions in their realization expressions. We consider these key prepositions as central prepositions in representing the associated frame elements. For some frame elements in the one frame element to multiple preposition structure mapping scheme, the level of dependency on certain prepositions to represent the deep semantic arguments is almost as high as those listed in the one-to-one mapping conditions. In some cases, the centrality of prepositions in argument realizations is so notable that even though there are other prepositions structures in the realization patterns, the noncentral preposition structures are only few and far between. Now let us decompose the information in Table 1 and have a closer look at the preposition distribution information for each of the listed frame elements.

- Goal: to 0.305, into 0.209, in 0.112, on 0.076, over 0.039, onto 0.027, at 0.024, on to 0.022, against 0.017, down 0.016.
- Path: through 0.180, across 0.122, down 0.121, along 0.102, up 0.089, towards 0.080, over 0.063, past 0.041, round 0.026, in 0.025, around 0.023, by 0.018.
- Topic: about 0.524, on 0.183, of 0.133, over 0.042, in 0.020, concerning 0.016.
- Source: from 0.567, off 0.143, out 0.112, out of 0.084, of 0.050, away from 0.016.
- Time: in 0.327, on 0.177, at 0.122, after 0.084, during 0.060, before 0.056, by 0.027, as 0.020, over 0.016.
- Place: in 0.482, at 0.244, on 0.121.
- Reason: for 0.630, as 0.114, of 0.070, with 0.033, on 0.027, from 0.021, because of 0.019, at 0.019, in 0.019.
- Theme: with 0.509, of 0.338, by 0.077, in 0.052.
- Individuals: of 0.999.
- Manner: in 0.302, like 0.255, with 0.186, on 0.039, without 0.039, as 0.038, at 0.030, by 0.026.
- Stimulus: by 0.368, at 0.255, with 0.133, of 0.085, about 0.056, in 0.038, for 0.025, over 0.015.
- Area: around 0.281, about 0.190, in 0.145, round 0.086, on 0.064, over 0.045, of 0.042, among 0.027, at 0.015.
- Addressee: to 0.621, at 0.192, with 0.042, of 0.033, from 0.031, in 0.022, on 0.019, for 0.017.
- Means: by 0.511, with 0.226, through 0.066, in 0.060, on 0.058, of 0.034.
- Whole: of 0.942, into 0.031.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

- Content: of 0.474, for 0.119, about 0.094, in 0.063, at 0.046, against 0.032, as to 0.030, from 0.025, to 0.022, with 0.022.
- Ground: in 0.383, on 0.143, of 0.131, through 0.060, against 0.027, at 0.026, from 0.022, around 0.019, under 0.019.
- Location: in 0.285, on 0.143, at 0.142, by 0.043, around 0.042, of 0.042, over 0.030, behind 0.025, under 0.024, outside 0.023, across 0.020, against 0.020, round 0.018, beside 0.016.
- Stuff: of 0.999.
- Substance: of 0.977, on 0.016.
- Category: as 0.540, of 0.363, into 0.034, in 0.017, with 0.017.
- Message: of 0.389, for 0.232, to 0.086, in 0.063, about 0.044, with 0.044, against 0.028, on 0.024, into 0.019, as 0.016.
- Relative_location: in 0.359, on 0.162, at 0.140, of 0.082, between 0.036, near 0.031, around 0.026, above 0.023, outside 0.018.
- Evaluee: of 0.464, for 0.159, on 0.136, at 0.044, in 0.040, about 0.037, to 0.034, against 0.027, with 0.015.
- Speaker: by 0.578, from 0.288, of 0.125.
- Direction: through 0.115, towards 0.109, down 0.079, in 0.073, over 0.073, into 0.071, around 0.053, across 0.051, round 0.044, up 0.040, out 0.034, to 0.034, about 0.030, from 0.028, along 0.022, behind 0.022, inside 0.022, past 0.018.

In the above list, each frame element name is followed by a head preposition and the probability of its occurrence. For the frame element names, they are taken independently of the frames to which they were assigned, which in our view, corresponds more to the newer view of these things in the current FrameNet database. The probability selection threshold is $T \geq 0.015$. We can observe the obvious centrality in the use of prepositions for the realization of some FEs. For example, although there are five prepositions (*as, of, into, in, with*) for the frame element *Category*, the key prepositions are actually only the first two prepositions: *as* and *of*. These two prepositions subsume more than 90% of the uses of preposition structures in the realization patterns of this FE. For the frame element *Reason*, the probability of using the central preposition *for* is 0.630, much higher than the probability of using the second or the third most frequently used preposition (*as* 0.114, *of* 0.070). For the frame element *Addressee*, the central prepositions *to*, *at*, and *of* account for more than 80% of the uses. There is an obvious selection preference of *of* for the following four frame elements: *Whole* (*of* 0.942) in the frames of Part_whole, Part_orientational, Part_ordered_segment, Part_inner_outer, Part_edge, Calendric_unit; *Stuff* (*of* 0.999) in the frames of Measure_linear_extent, Measure_mass, Measure_volume; *Substance* (*of* 0.977) in the frames of Part_piece and Shapes; *Individuals* (*of* 0.999) in the frames of Aggregate and Quantity. The probabilities of using *of* for the frame elements *Individuals* and *Stuff* are so high (0.999) that they almost exclude the use of other prepositions.

Figure 3 shows the distribution of central prepositions and the degree of preposition centrality in the syntactic realization of semantic arguments. The figure includes the top fifteen frame elements that use preposition structures in their

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

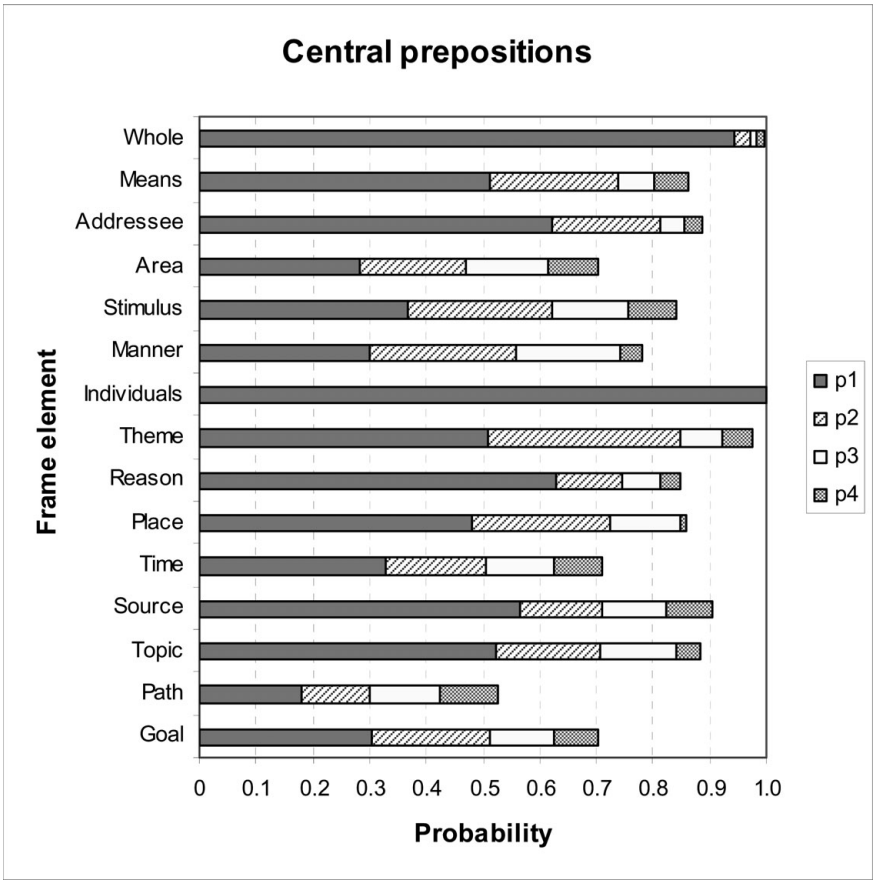


Fig. 3. Degree of preposition centrality for the fifteen top most frequent PP-FEs. p1, p2, p3, and p4 represent the first, second, third, and fourth most frequently used preposition types for the specified frame element.

argument realizations. The four sections (p1, p2, p3, and p4) for each bar represent the first four most central prepositions for each frame element. For most of the frame elements, the first four prepositions cover more than 70% of the preposition uses for the associated semantic argument. The frame element *Path* is the only exception. In ten out of the fifteen frame elements, the first four central prepositions for each FE cover a PP-FE proportion that exceeds 80%. This means that in most cases, if we have information about the frame element and the lexical word, we would be able to predict the most likely prepositions to use.

It is worth noticing that some frame elements with a cross-frame occurrence nature such as *Direction*, *Path*, and *Location* have a rather noticeable tendency of using multiple realization configurations with different preposition structures. Here we use entropy to measure the uniformity of distribution of preposition frequencies to help us decide whether there is a tendency of using central prepositions for the realization of some semantic arguments.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Table 2. *Preposition density variance in the most frequent PP-FE structures in FrameNet (threshold $f \geq 500$)*

FE	Entropy	P (FE | Prep)	Max_P	Prep	Example
Individuals	0.0082	0.020469	0.9993	of	army [of civil servants]
Stuff	0.0157	0.009818	0.9986	of	mugs [of hot chocolate]
Substance	0.1888	0.00962	0.9765	of	coils [of hair]
Whole	0.4314	0.014635	0.9421	of	top [of the tree]
Speaker	1.4258	0.008038	0.5782	by	appeals [by Soviet scientists]
Category	1.6092	0.009097	0.5404	as	counted [as fully employed]
Theme	1.725	0.020752	0.5092	with	coated [with a waxy substance]
Addressee	1.9086	0.016146	0.6212	to	reprimand [to those in contempt]
Source	2.0747	0.040387	0.5666	from	jumped [from the car]
Reason	2.1905	0.024693	0.6299	for	admiration [for her courage]
Means	2.239	0.014776	0.5115	by	suicide [by taking poison]
Topic	2.3125	0.040585	0.5238	about	lectures [about the expedition]
Place	2.5424	0.031247	0.4819	in	snowstorms [in the Alps]
Stimulus	2.5623	0.018053	0.3678	by	traumatized [by what we saw]
Evaluee	2.6166	0.008433	0.464	of	wicked [of you] to say that
Content	2.8375	0.012798	0.4735	of	dislike [of this type of work]
Message	2.8695	0.008716	0.389	of	admission [of defeat]
Manner	2.9694	0.020045	0.3016	in	conversing [in low voices]
Relative_location	3.2467	0.008659	0.3589	in	cart [in the street]
Time	3.2503	0.035499	0.3271	in	storm [in 1910]
Area	3.3794	0.016994	0.281	around	traveled [around Europe]
Ground	3.38	0.012064	0.3829	in	fumbled [in her handbag]
Goal	3.5587	0.087315	0.3055	to	proceeded [to the station exit]
Location	3.827	0.011146	0.2852	in	seated [in an armchair]
Path	3.8525	0.06621	0.1796	through	escape [through the kitchen]
Direction	4.3217	0.007134	0.1149	through	glanced [through the curtains]

Table 2 shows the entropy information about the degree of multiplicity and the density variance in the use of prepositions to represent FEs. It is for the PP-FEs that have an occurrence frequency of more than 500. The FE column is for frame

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

element names. Given a preposition structure that is marked as a frame element in FrameNet, the probability of this preposition structure representing the specified FE is P(FE|Prep). Max_P is the maximum probability of a preposition type denoting the semantic argument of this specified FE. Prep refers to the preposition that bares the maximum probability. It can also be interpreted as the most frequently used preposition representing the FE concept. The Example column lists condensed examples from the preposition knowledge base, demonstrating how the most central preposition for each of the frame elements is used in FrameNet. Words such as *army*, *mugs*, and *coils* that are outside the square brackets in the Example column are frame-bearing target lexical units. They are the target lexical units that evoke the specific frame elements in the FE column and that are used with the prepositions in the Prep (preposition) column. The words included in the square brackets are the preposition structures that represent the corresponding frame element concept in the FE column. The table is sorted by the entropy values.

A high entropy value coupled with a low maximum probability value means that the specified frame element has more varied preposition structures in its syntactic realization expressions. This is where the frame element lacks the centrality of preposition structures in representing the frame element concepts. *Direction* and *Path* are the two most obvious examples. Both of the frame elements indicate general spatial concepts. The spatial concepts can be associated with all directions in the real world such as *through, across, towards, down, along, up* etc. It is hard to say that one direction is more important than the others. That is why the uses of prepositions are more or less evenly distributed, and the prominence of using key prepositions for these frame elements is thus obscured.

2.3 Explaining multiple FE-preposition correspondence options

The argument realization options of both one-to-one mapping and one-to-many mapping clearly support the hypothesis that there is a link between the frame elements and the central prepositions in the surface syntactic realizations. The syntactic realization of arguments is predictable to a large extent from the frame element concept and from the target lexical units. In the FrameNet sentences analyzed, the tendency to assimilate the core meanings of individual prepositions with some basic semantic notions such as *Agent, Manner, Means, Instrument, Theme, Reason, Topic* etc. is rather marked. Now, let us focus on why at the same time noncentral prepositions are selected in some surface syntactic representations. Our analysis will be based on the factors that can have an impact on the selection of prepositions in the semantic argument realization process.

We isolated, from the preposition knowledge base, sentences that can demonstrate the use of multiple prepositions in representing the same frame element, and generated a mapping comparison list. This list contains 27 036 sentences. A close examination of the sentences in the list reveals some major factors that can affect the selection of different preposition structures as the surface manifestations of arguments.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

2.3.1 Interchangeable prepositions

One reason for the one-to-many frame element to preposition structures mapping is that in some cases, preposition representation variations do not disrupt the meaning of the sentence. For some frame elements, rules for selecting prepositions are in fact not so rigid among certain pairs of prepositions. The prepositions in these meaning-form mapping relations are more or less interchangeable without affecting the overall meaning of the argument they represent. And this, in a way, can generate statistical differences in the counting of different preposition types. For example, in many sentences, when *to* is replaced with *towards*, *about* with *around*, *among* with *amongst*, and *on* with *upon*, the basic meaning of the sentence is still preserved. Sometimes, this interchangeability is limited to preposition pairs that are used for representing a certain frame element concept, and not necessarily elsewhere (see Examples (4)–(8)). The choice of prepositions is not always semantically determined as the choice of some spatial or temporal prepositions such as *before noon* versus *after noon*, and *inside the box* versus *outside the box*. In the daily use of English, we also find examples where one preposition is replaced by another preposition or prepositional phrase without making much difference in the truth value of the expressions. We can say *to use boxes for chairs* instead of *to use boxes instead of chairs*. We can use *to act for my client* to replace *to act in the interest of my client*. We can also substitute *to vote for him* with *to vote in favour of him* and *to give a banquet for somebody* with *to give a banquet in honor of somebody*. The motivation for the choice of prepositions in constructions that allow freely interchangeable prepositions is not changing the sense but presenting the same information from different perspectives for pragmatic or stylish reasons. However, variations concerning the use of different prepositions for the same senses can create deviations from those structurally identified as central prepositions for the senses at issue. Prepositions that are similar in semantic content but different in the structural form can pose crucial challenges for the framework that counts only the word types in deciding the centrality of preposition usage. In addition, when selecting different prepositions with about the same senses, the 'aboutness' can vary in degrees. Some interpretations can get very complicated as to whether the use of similar prepositions involves fully or partially overlapping applications of prepositional senses in specific contexts. Those complex cases may demand some more elaborate abstractions or different categories in classification.

(1) Frame: Contingency
   1. **depends** [DETERMINANT on the frequency and abundance of species only]
   2. **depends** [DETERMINANT upon the snail reproducing]

(2) Frame: Categorization
   1. **count** [CATEGORY among Sinatra 's guests]
   2. **count** [CATEGORY amongst my favorite places on earth]

(3) Frame: Cause_motion
   1. **yanked** [PATH toward the car]
   2. **yanked** [PATH towards him]

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

(4) Frame: Emotion_directed
   1. **overjoyed** [STIMULUS at the election result]
   2. **overjoyed** [STIMULUS with the suggestion]

(5) Frame: Quarreling
   1. **disagreement** [ISSUE about the way in which these goals should be achieved]
   2. **disagreement** [ISSUE on how that is best achieved]
   3. **disagreement** [ISSUE over club rules]

(6) Frame: Candidness
   1. **honesty** [ADDRESSEE to Tim]
   2. **honesty** [ADDRESSEE with your partner]

(7) Frame: Emotion_directed
   1. **annoyed** [STIMULUS at his bad manners]
   2. **annoyed** [STIMULUS by my silence]

(8) Frame: Abounding_with
   1. **full** [THEME of treasures]
   2. **full** [THEME with all the Christmas mail]

2.3.2 LU-governed prepositions

In FrameNet, the majority of prepositions are used to indicate a syntactic relationship between the target lexical unit and its complement, or sometimes between the LU and its modifiers. However, it is possible that a preposition is extremely dependent on the target lexical word so that it becomes semantically meaningless, and its semantic value can be interpreted only with respect to the specific target lexical unit. In these sentences, target lexical words can pose constraints on the selection of the accompanying preposition structures. This means that different prepositions are required for the realization of the same frame element concept because of the target lexical unit in the sentence. For example, in 'Our primary mission is to educate people so that they can arrive at the right conclusion themselves,' we say *arrive at* the right conclusion, but not *arrive in* the right conclusion. The governing lexical head word can be a verb such as in *suffer from, succeed in, respond to, prevent from, pray for, participate in, contribute to, comment on, concentrate on, consist of, count on, believe in, belong to*. This is what Fillmore called 'governed' prepositions (Fillmore 1968). Prepositions in these constructions are firmly wedded to the head words as if they have practically become one word. Although these constructions can be construed as complete units in the lexicon, they are not analyzed as multi-word lexical units such as the noun–noun compound *wheel chair* in FrameNet (Ruppenhofer *et al.* 2006). Such prepositions are regarded as head prepositions in the preposition phrases as can be seen from these annotation divisions: 'The passengers LOOKED [at the information monitors],' 'Let me know if you COME [across that reference],' 'But all that DEPENDS [on one person].'

Sometimes, the LU that selects the preposition is an adjective as in *absent from, aware of, afraid of, attracted to, ashamed of, blessed with, contrary to, good at, guilty*

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

of. The lexical head word can also be a noun such as in *comments on, reason for, the solution to a problem, one's reaction to something, alternative to, addiction to, belief in, dedication to, delay in, preference for*. Of course, the dependence upon the target lexical word can be on various points in the granularity scale. Depending on the target lexical word at stake, it can be difficult sometimes to find a clear cut distinction between prepositions that are controlled by the target LU and those that are not. In Examples (9)–(12), we can see how the dependence on target lexical unit for the selection of prepositions can disrupt the centrality pattern of using key prepositions to represent FEs.

(9) Frame: Cause_to_be_wet

1. **soaked** [LIQUID in ether]
2. **saturate** [LIQUID with deuterium]

(10) Frame: Hindering

1. **inhibit** [ACTION from launching out on their own]
2. **hinder** [ACTION in their beautiful lives]
3. **interferes** [ACTION with perception]

(11) Frame: Cure

1. **treated** [AFFLICTION for the disease]
2. **resuscitated** [AFFLICTION from a heart attack]

(12) Frame: Compliance

1. **submit** [NORM to any authority]
2. **comply** [NORM with express procedural requirements]

2.3.3 Prepositions for sense distinctions in some general FE concepts

Some frame elements have a rich variety of conceptual parameters within them. For example, the frame element *Direction* can branch off in all directions such as *up* and *down*, *to* and *from*, *inside* and *outside*, *around* and *across* etc. Similarly for the concept of *Time*, we may have *before* or *after*, *during a period of time* or *at a point of time*. One of the functions of the preposition structure is to differentiate the distinctive parameters within the general concept so that *burn near the house* is different from *burn inside the house*, and *look past the room* is not the same with *look round the room*. In Examples (13)–(17), the exchanging of prepositions does not constitute a syntactic breach, but the alternation may generate different senses within the same semantic argument framework. This type of one-to-many FE-preposition correspondence usually involves some frequent spatial and temporal frame element concepts such as *Goal*, *Path*, *Direction*, *Area*, *Location*, *Time*, and *Duration* where preposition structures can encompass specifications of semantic information needed in the representation of thematic arguments.

(13) Frame: Placing

1. **stuffed** [GOAL down the front of our pyjamas]
2. **stuffing** [GOAL in a cupboard]

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

3. **stuff** [GOAL inside my jacket]
4. **stuffed** [GOAL into an envelope]
5. **stuffed** [GOAL under her pillow]

(14) Frame: Dispersal

1. **scattered** [GOAL_AREA across his desk]
2. **scattered** [GOAL_AREA around the country]
3. **scatter** [GOAL_AREA in the graveyard]
4. **scattering** [GOAL_AREA into the dust]
5. **scatter** [GOAL_AREA on to their scallops]
6. **scattered** [GOAL_AREA on the floor]
7. **scatter** [GOAL_AREA over the fish]
8. **scattering** [GOAL_AREA upon the heads of cheerful spectators]

(15) Frame: Natural_features

1. **burn** [RELATIVE_LOCATION at the side of the field]
2. **burn** [RELATIVE_LOCATION below the house]
3. **Burn** [RELATIVE_LOCATION near Loch Doon in south-west Scotland]

(16) Frame: Perception_active

1. **peek** [DIRECTION into someone 's kettle]
2. **peek** [DIRECTION round the back]
3. **peek** [DIRECTION through the window]
4. **peek** [DIRECTION under the cloche]

(17) Frame: Sleep

1. **dozed** [DURATION through the night]
2. **dozed** [DURATION until morning]

2.3.4 Prepositions selected by object NPs

The selection of prepositions to represent a certain frame element can depend on the noun phrase that serves as the object of the preposition. In many cases, because of the noun complement in the prepositional phrase, only certain prepositions are possible. For example, the preposition *in* is used for not only particular seasons, but also for seasons in general (*swim in summer*, *launched in the summer of 1989*). However, for different parts of the day such as night, morning, afternoon, and evening, a different preposition is used when referring to a particular night, morning, afternoon, or evening. We say *work at night*, but *misbehavior on that night*; we say *call in the evening*, but *the call on the evening of 28 April 1981*. Although *in the afternoon*, *at night*, *on Monday morning*, and *on that evening* are all preposition structures indicating the *Time* frame element concept, different prepositions are chosen because of the complement noun phrase in the preposition structure. It would not sound right when the prepositions are switched in these phrases, unless there are special texts to clarify the justification. Preposition object constraints like this can be a contributing factor for the diversity of semantic–syntactic mapping options.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

(18) Frame: Arrest

1. **arrest** [TIME at 11pm on 23 December]
2. **arrest** [TIME on Thursday morning]

(19) Frame: Contacting

1. **ringing** [TIME at regular intervals]
2. **ring** [TIME in an hour or so]
3. **rang** [TIME on the Saturday at 9.30 in the morning]

(20) Frame: Killing

1. **drowned** [TIME at birth]
2. **beheading** [TIME in 1766]
3. **assassinated** [TIME on 14th]

(21) Frame: Hostile_encounter

1. **battle** [TIME in 1388]
2. **confrontation** [TIME on Boxing Night]

(22) Frame: Waking_up

1. **awoke** [TIME at dawn]
2. **awoke** [TIME in the morning]
3. **awoke** [TIME on Monday]

2.3.5 Frame restrictions to the selection of prepositions

Cross-frame FEs such as *Goal*, *Theme*, and *Topic* are not uncommon in FrameNet. One interesting phenomenon is that the same frame element in different parent frames can have different syntactic realizations by prepositions structures. Examples (23)–(25) demonstrate this interesting phenomenon. The representation of the same frame element *Theme* involves the use of different prepositions in different frames. In the Filling frame, we say *to load something with something*, but in the Emptying frame we do not say *to clear something *with something*; instead we say *to clear the cart of the food items*. Similarly, we do not use *with* in the Removing frame to represent the frame element *Theme*. Thus, instead of saying *evacuation with refugees by aircraft carriers*, we say *evacuation of refugees by aircraft carriers*. In these examples, the frame itself can be the semantic determinant in the syntactic representations, and thus can account for some one-to-many meaning-form mapping options in the process of argument realization.

One notable difference between the LU-governed prepositions discussed in Section 2.3.2 and the frame-restricted prepositions here is that the former is based on the individual lexical unit and the structures are often idiosyncratic, while for the latter, the choice of the preposition can affect most of the lexical units in the frame. For example, in the Notification_of_charges frame, we have *two men accused of the murder of a police informer*, *to charge a father with attempted murder*, and *the promoter was indicted on federal wire fraud charges*. For all that the three verbs *accuse, charge,* and *indict* are categorized in the same frame, three different prepositions are selected because of the control of three individual verbs. Also, for the LU-governed preposition structures, when different head words call for different prepositions, we

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

do not omit one of them in parallel structures as in 'He was fascinated by and enamored of this beguiling woman.'

(23) Frame: Filling

1. **adorn** [THEME with fake diamonds]
2. **coated** [THEME with a single molecular layer of a dye]
3. **cover** [THEME with a foot of topsoil]
4. **crammed** [THEME with expensive furniture]
5. **drape** [THEME with nets]
6. **embellished** [THEME with precious gems]
7. **filled** [THEME with belay nuts]
8. **flood** [THEME with nylon sheets made by one of its subsidiaries]
9. **glazed** [THEME with a plain color or a printed design]
10. **loading** [THEME with all manner of bounties cast at their feet by providence]
11. **scattered** [THEME with damaged tanks, …]
12. **showered** [THEME with coins and rubbish]
13. **smeared** [THEME with butter]
14. **splashed** [THEME with cold water]
15. **spread** [THEME with a blue and crimson Indian blanket]
16. **squirt** [THEME with blinding sprays]
17. **surfaced** [THEME with stone]

(24) Frame: Emptying

1. **cleared** [THEME of leaves]
2. **denuded** [THEME of wood]
3. **divested** [THEME of the jacket]
4. **emptied** [THEME of black coffee]
5. **purge** [THEME of all unnecessary files]
6. **stripped** [THEME of everything valuable]
7. **draining** [THEME of fruit juice] [MANNER in one gulp]

(25) Frame: Removing

1. **withdrawa**l [THEME of funding] [AGENT by the Japanese government]
2. **ejection** [THEME of Gandhi] [GOAL on to a station platform]
3. **evacuation** [THEME of would-be emigrants] [SOURCE from the embassy]
4. **eviction** [THEME of a family] [SOURCE from a house in Creggan]
5. **expulsion** [THEME of the Jesuits] [SOURCE from Portugal]
6. **removal** [THEME of Franca] [SOURCE from the upper storey]

2.3.6 Prepositions in idiomatic expressions

Some prepositional phrases become rigid idiomatic expressions. These rigid expressions involve uninterrupted sequences of words headed by a preposition, such as *in the long run, from scratch, by all means, by word of mouth, to all intents and purposes, to date*. They can include prepositions and nouns (*under consideration, by chance, by accident, off the track)*, prepositions and adjectives (*from afar, at last, for good*),

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

or prepositions and adverbs (*at once*). They can involve a complex preposition (*out of the blue*) or more than one preposition (*from head to toe, from time to time*). In these preposition constructions, the combination of the head preposition with some of the complement words in the prepositional phrase has become established as a unit in representing frame element concepts. The meaning of the preposition phrase cannot be construed without reference to each of the main constituents including the heading preposition. Examples (26)–(30) list some rigid or idiomatic expressions that are not governed by target LUs, nor are they selected by the complement nouns. They are integrated with the participating constituents as fixed expressions and they can only be properly understood in connection with the rest of the phrase. In general, these prepositional phrases cannot be broken into smaller fragments; they are lexical units in and of themselves. The degree of idiomaticity can also be varied. Some structures are more rigid and fixed, while some others are not so frozen and can allow for some minor changes in certain participating constituents. For example, in *on the upper floor*, it can also be *on the lower floor*. Similarly *at the thought of*, can be *at the sight of*, and *on George's part*, can be *on my teacher's part*. These prepositional phrases do not seem to follow any simple construction rule; and the idiomatic nature of the prepositional phrase does not allow for the alternation of the head preposition. Changing the preposition would either alter the meaning, or render it ungrammatical. This use of prepositions can explain why in some cases a noncentral preposition is preferred over the key prepositions in the surface representations of arguments. The selected preposition may not be a central preposition in representing the frame element, but the fixed prepositional expression has unique subcategorization properties that usually cannot be captured by replacing the head preposition.

(26) Frame: Residence

1. **lodged** [LOCATION at our house]
2. **lodged** [LOCATION on the upper floor]

(27) Frame: Commerce_buy

1. **buy** [SELLER over the counter]
2. **purchased** [SELLER at any computer supply store]
3. **purchase** [SELLER from one supplier]

(28) Frame: Emotion_directed

1. **depressed** [STIMULUS at the thought of the ordeal to come]
2. **depressed** [STIMULUS by the whole affair]

(29) Frame: Self_motion

1. **plodded** [COTHEME in his wake]
2. **hurried** [COTHEME after the others]
3. **scramble** [COTHEME behind it]
4. **dances** [COTHEME with the rest]

(30) Frame: Placing

1. **positioned** [GOAL close to the speaker]
2. **position** [GOAL within easy reach of the doors]

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

2.3.7 Interaction of linguistic constituents restraining the selection of some prepositions

One of the factors accountable for the variance in semantic–syntactic mapping patterns is the interaction between different linguistic constituents of the sentence. In some sentences, selection of prepositions can be influenced by syntactic constructions other than the target lexical word or the object of the preposition. For example, a present perfect tense may take a *since* or *by* phrase to indicate the time concept. In most cases the passive voice predicate may require that the initiator, not necessarily the performer, of the action be introduced by the preposition *by* such as in 'I am angered by his words.' Also, if other parts of the sentence indicate that two parties are involved in a dispute, only *between* can be chosen in the preposition structure option *dispute among/between*. Examples (31)–(33) demonstrate the fact that linguistic element interactions can in fact license the use of only certain prepositions and not others.

(31) Frame: Desiring

1. [EXPERIENCER the man] [EXPERIENCER who] **yearned** with all his heart [DURATION for all his life] [EVENT to be a woman …]
2. [EVENT the drink] [EXPERIENCER he] had **craved** [DURATION since noon]

(32) Frame: Response

1. [AGENT the firemen] must have **responded** [TIME by now]
2. [AGENT Saudi Arabia] **responded** [TIME on Feb. 3] [RESPONSE by banning the recruitment of Thai workers]

(33) Frame: Quarreling

1. **dispute** [ARGUERS among professionals of quality who had any discernment in such matters]
2. **dispute** [ARGUERS amongst theoreticians and dozens of political parties and sects]
3. **disputes** [ARGUERS between labor and employers]

3 Related work and other corpus-based evidence

Many researchers such as Stockwell, Schachter and Partee (1973) and Langendonck (1974) have supported or adopted the hypothesis that deep semantic roles are connected with some typical prepositions. Generally speaking, verbs in English take semantic arguments and many of the semantic arguments can be realized by prepositional phrases. Scholars also hypothesized that the lexical meaning of a verb can, to a large extent, determine its syntactic behavior (Talmy 1985; Gropen *et al.* 1989; Levin 1993; Goldberg 1995). The mapping relationship between verb meaning and their syntactic behavior has been extensively studied in Levin's landmark research on verb classifications, *English Verb Classes and Alternations* (Levin 1993). In this influential work that has served as the basis for explorations and discussions in artificial intelligence and natural language processing, Levin defined approximately

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Table 3. *Sample annotations for admit.01 with the pattern sub-ARG0 S-ARG1-to in PropBank*

Verb	Pattern	Annotation
admit.01	sub-ARG0 S-ARG1-to	[ARGM-TMP After his arrest], [ARG0 the forger] [rel admitted] to [ARG1-to *-1 faking and selling other paintings up and down the Eastern seaboard] .
admit.01	sub-ARG0 S-ARG1-to	Attorneys for Messrs. Burke and Evans and Ms. Clarke said that although [ARG0 their clients] [rel admitted] to [ARG1-to *-3 making some misrepresentations in the sales], they had believed that the works were authorized *-1 by Mr. Dali, who *T*-2 died in January .
admit.01	sub-ARG0 S-ARG1-to	[ARG0 Both residents' organizations] [rel admit] to [ARG1-to *-1 receiving some funding from abroad] .

200 verb classes, which she argues reflect important semantic regularities. It is assumed that verbs sharing similar semantic properties can have similar syntactic projections. They can be organized into classes on the basis of their meaning components and syntactic manifestations. From a practical point of view, what this means to computational linguistics and corpus linguistics is that semantic meaning can, to a certain extent, be inferred with knowledge of the classified verb classes and their syntactic behavior patterns. As long as the study of preposition semantics is concerned, this can indicate that with enough information about the surface syntactic structures of verbs and their accompanying prepositional phrases, we should be able to glean information about the underlying semantic argument features for the verb preposition constructions. For example, when dative prepositions are used with dative verbs and when benefactive verbs take benefactive prepositional phrase complements, we can generally predict the underlying semantic roles represented by the prepositional phrases in these constructions. For the past decade or so, Levin's classification of verb classes has been successfully applied in word sense disambiguation (Dorr and Jones 1996), information retrieval (Levow, Dorr and Lin 2000), machine translation (Palmer and Wu 1995; Dorr 1997), document classification (Klavans and Kan 1998), and subcategorization acquisition (Korhonen 2002).

In trying to find out if empirical data supports the hypothesis that semantic roles can be typically marked by certain prepositions, we also experimented with data sets from the PropBank project. PropBank (Kingsbury and Palmer 2002; Palmer, Kingsbury and Gildea 2005) includes a corpus of texts annotated with information about semantic roles (see Table 3). It adds a layer of predicate-complement relations to the syntactic structures of the Penn Treebank corpus.

In contrast to FrameNet, PropBank is designed to cover every instance of a verb in the Wall Street Journal articles of the Penn Treebank corpus. Semantic roles

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

in PropBank are defined on a per-verb basis. The annotation of this corpus is based on Levin verb classes (Levin 1993) defined in the VerbNet lexicon (Kipper, Dang and Palmer 2000; Kipper 2005). Because existing approaches thus far differ widely with respect to frame types and their granularity (Brent 1993; Manning 1993; Ushioda *et al.* 1993; Baker and Ruppenhofer 2002), PropBank chooses to neutralize the taxonomy in frames and semantic roles for argument realization. The expected arguments of each of Levin's senses are simply numbered from *ARG0* to *ARG5*, and argument labels are not necessarily significant across verb classes. In general, some regularity can be observed for lower numbered arguments. For example, subjects of transitive verbs are assigned *ARG0*. When a preposition is used to mark *ARG0*, it is usually *by* in the passive voice construction. *ARG1*s are usually direct objects and they correspond to the roles of *Theme* and *Patient*. *ARG2*s generally mark indirect object and are more varied in representing semantic roles: some are *Benefactive*, some are *Instrument* or *Attribute*, and some are indicators of the end state. Many *ARG3*s can also be assigned the same semantic roles; they can be *Benefactive*, *Instrument*, or *Attribute*. It is interesting to note that when a source–goal concept is expressed using prepositions in PropBank, *ARG3* is usually the source indicating the start point, and *ARG4*, the goal showing the end point. Prepositional phrases can be used to mark semantic arguments from *ARG0* to *ARG5* and some *ARGM* categories that reflect the original taxonomy in the Treebank annotations: *DIR* for direction, *MNR* for manner, *LOC* for location, *TMP* for time, *CAU* for cause, *ADV* for adverbials, and *PCN* for purpose.

Basically what we want to do is to extract all the prepositional phrases that are marked with an argument type in PropBank, and analyze them on a verb class basis. We are interested in the Penn Treebank corpus evidence regarding what typical prepositions can be used with what classes of verbs to realize what argument roles in the verb constructions. To find out whether there are predictable linking relationships between prepositions and classes of verbs in representing certain semantic roles, we used the PropBank annotations from http://www.cs.rochester.edu/~gildea/PropBank/Sort/ (henceforth referred to as *annotated PropBank corpus*) as the seed corpus and proceeded to build a PropBank preposition database. First we assembled all the annotated sentences from the annotated PropBank corpus. By processing the files from which we collected the annotated sentences, we were able to assign to each of the annotated sentences, some key information such as verb type (admit.01, admit02, admit.xx), and argument pattern at the sentence level (sub-ARG0 S-ARG1-to). From this collection of annotated sentences, a further filtering process separated 31 563 sentences each of which has at least one preposition structure marked with a PropBank sense tag such as *ARG3* or *ARGM-TMP*. For our purpose, what was lacking in this annotated corpus was the verb classification information that indicates that a certain verb instance belongs to a certain verb class. Without this piece of information, verb constructions in PropBank cannot be grouped and classified into classes, and thus verb-class-based statistical analysis cannot be carried out. The VerbNet and PropBank mapping list from http://verbs.colorado.edu/semlink/ (referred to as *semlink list*) contains verb-class classification information that can bridge this gap.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

We designed a meta-information unit that can serve as a unique identity for each of the annotated sentences in the annotated corpus and for each of the record in the semlink list. This is composed of the Wall Street Journal article file name, the target verb, and the sequential number of sentence in which this sentence can be traced in the Wall Street Journal file (for example, admit:wsj0003:45). While this denominator identity information is present in the semlink list, it is not available in the annotated PropBank corpus. By comparing sentences in the annotated PropBank corpus (31 563 sentences) and the Wall Street Journal sentences (49 208 in all) on a tag-free, brute-force matching basis, we obtained the unique identity information for each annotated sentence in the annotated PropBank corpus. After that, we did some prior pruning of the records in the semlink list to eliminate those lines that have no reference to sentences with preposition annotations. The process also left out those lines that had no information about verb classes. When this is done, we simply extracted verb class information from the semlink list and assigned them to the corresponding sentences in the annotated PropBank corpus. Thus, we built our preposition database with 15 837 sentences and 18 329 prepositional phrases that are marked as semantic roles in PropBank.

Table 4 contains 16 semantic roles that are most frequently represented by prepositions in PropBank (frequency $f>130$). Among these, 11 roles show a marked tendency of using central prepositions, with the first four most frequent prepositions covering more than 80% percent of all the prepositions used for the sense. In 6 of them, there is a particular preposition to mark the semantic role; the coverage of the primary preposition exceeds 90% of all the uses of prepositions for that sense. The binding relationship in PropBank between central prepositions and the senses they represent seems to be as strong as in FrameNet (see Fig. 3) where for 10 out of 15 frame elements, the first 4 central prepositions cover a PP-FE proportion of more than 80%. However, for only 2 out of the 15 most frequent FEs in FrameNet, the coverage of the most frequent preposition is more than 90%. One reason to account for this difference is that PropBank examples are grouped using Levin's inventory of verb classes; FrameNet frame elements and Levin verb classes are under two semantic taxonomy schemes (Baker and Ruppenhofer 2002) and thus the semantic role assignments are different in these two resources. Although some PropBank semantic roles and some FrameNet frame elements overlap to certain degrees such as *MNR* and *Manner*, *TMP* and *Time*, there are sense categories that are only present in one of the systems but not the other. For instance, FrameNet does not have an FE for prepositional phrase used as *ADV*, and PropBank does not contain *Individuals* and *Whole* in this list of 16 most frequent sense categories using prepositions, partly due to the fact that FrameNet includes nouns and adjectives as target LUs. The ranking is also different because some of the general class roles in PropBank such as *Location* might cover more than FrameNet does, while some FrameNet FEs such as *Source* and *Goal* are decomposed to argument roles such as *ARG3* and *ARG4* in individual verb classes in PropBank. However, despite the difference in semantic role categorization, both systems demonstrate a clear tendency of using central prepositions in representing semantic contents as can be witnessed from Table 4 and Figure 3. This adds corpus-based empirical support to

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Table 4. *Semantic roles that are most frequently represented by prepositions in PropBank (frequency $f > 130$)*

Argument and verb class	Tokens	Types	Mean	Std. Dev.	Four most frequent prepositions (%)	Other preps (%)
ARGM–TMP (cross-classes)	3 378	44	76.7727	175.0011	in: 31.6, after: 10.1, for: 9.1, at: 6.4	42.8
ARGM–LOC (cross-classes)	2 159	45	47.9778	199.6701	in: 60.8, at: 11.4, on: 10.9, under: 5.1	11.8
ARGM–MNR (cross-classes)	1 125	44	25.5682	51.1056	with: 19.5, in: 16.6, by: 16.3, as: 7.0	40.6
ARGM–ADV (cross-classes)	902	53	17.0189	35.6826	as: 23.6, in: 13.1, with: 10.3, according to: 5.8	47.2
ARG4–45.06 (‘decrease’ verbs)	658	4	164.5000	322.3347	to: 98.5, below: 0.6, above: 0.6, into: 0.3	0.0
ARGM–DIR (crossx-classes)	450	38	11.8421	17.0334	into: 13.8, to: 12.4, from: 10.4, through: 8.2	55.2
ARG2–29.02 (‘characterize’ verbs)	334	10	33.4000	96.5035	as: 92.2, with: 2.1, in: 1.5, on: 0.9	3.3
ARG3–45.06 (‘decrease’ verbs)	257	1	257.0000	0.0000	from: 100.0	0.0
ARGM–CAU (cross-classes)	256	18	14.2222	22.1427	because of: 35.2, as: 17.2, on: 11.3, for: 7.4	28.9
ARGM–PNC (cross-classes)	248	13	19.0769	46.6341	for: 66.9, in: 22.6, as: 4.0, on: 1.6	4.9
ARG2–54.04 (‘price’ verbs)	232	7	33.1429	81.1263	at: 93.5, to: 3.9, as: 0.9, on: 0.4	1.3
ARG2–13.02 (‘contribute’ verbs)	205	5	41.0000	87.7696	to: 96.6, in: 1.5, throughout: 1.0, over: 0.5	0.4
ARG4–45.04 (‘alternating change of state’ verbs)	184	5	36.8000	73.0014	to: 90.8, into: 7.6, onto: 0.5, in: 0.5	0.6
ARG2–48.01.1 (‘appear’ verbs)	142	6	23.6667	37.4682	from: 66.2, in: 27.5, on: 2.1, into: 2.1	2.1
ARG2–29.05 (‘conjecture’ verbs)	132	6	22.0000	32.0062	as: 65.2, about: 12.1, for: 11.4, of: 9.1	2.2
ARGM–EXT (cross-classes)	131	9	14.5556	32.5427	at: 77.1, by: 7.6, less: 3.8, in: 3.8	7.7

Table 5. *Top thirty most frequent verb preposition combinations representing specific semantic roles in PropBank*

Argument and verb class	Verb	Preposition	Count
ARG4–45.06 ('decrease' verbs)	fall.01	to	141
ARG2–13.02 ('contribute' verbs)	contribute.XX	to	113
ARG2–87 (extended-class)	focus.XX	on	106
ARG2–54.04 ('price' verbs)	value.XX	at	98
ARG2–13.04.2 ('equip' verbs)	invest.XX	in	94
ARG2–54.04 ('price' verbs)	price.XX	at	93
ARG4–45.06 ('decrease' verbs)	drop.01	to	90
ARGM–LOC (cross-classes)	say.XX	in	88
ARG0–54.04 ('price' verbs)	rate.XX	by	87
ARG2–29.05 ('conjecture' verbs)	know.XX	as	86
ARG2–62 (extended-class)	aim.02	at	80
ARG4–45.06 ('decrease' verbs)	gain.01	to	77
ARG2–81 (extended-class)	accuse.XX	of	74
ARG2–11.01–1 ('send' verbs)	send.XX	to	72
ARG1–47.01–1 ('exist' verbs)	depend.XX	on	72
ARG2–11.01 ('send' verbs)	send.XX	to	72
ARG2–29.02 ('characterize' verbs)	view.02	as	67
ARG2–48.01.1 ('appear' verbs)	stem.01	from	65
ARG3–45.06 ('decrease' verbs)	fall.01	from	65
ARG4–45.06 ('decrease' verbs)	climb.02	to	62
ARGM–EXT (cross-classes)	close.02	at	62
ARG2–67 (extended-class)	prevent.XX	from	58
ARG2–22.02 ('amalgamate' verbs)	associate.XX	with	54
ARG4–45.06 ('decrease' verbs)	jump.01	to	52
ARG4–45.06 ('decrease' verbs)	decline.01	to	49
ARG1–37.07 ('say' verbs)	respond.XX	to	49
ARG2–47.01–1 ('exist' verbs)	wait.01	for	48
ARG0–27 ('engender' verbs)	cause.XX	by	48
ARG1–29.06 ('masquerade' verbs)	serve.01	as	47
ARG2–31.01 ('amuse' verbs)	hurt.01	by	44

the hypothesis that predictions can be made about semantic meaning on the basis of the information about the lexical unit and the preposition structures.

Table 5 lists the top thirty most frequent verb preposition combinations representing specific semantic roles in PropBank. In this list, many verb preposition combinations belong to Levin verb class 45.06 ('decrease' verbs). This reminds us that PropBank is based on the Wall Street Journal articles where verbs describing financial activities (such as *decline to, jump to, climb to, fall from, fall to, drop to, gain to*) abound. The list clearly indicates that verbs can be used with typical prepositions to consistently represent certain semantic arguments. For example, *to* is a typical preposition that can be used with 'decrease' class of verbs (45.06) to indicate the 'end point' concept. The semantic content of the preposition construction can be inferred from the verb class and the type of preposition the verb keeps company with. In general, this list agrees with our intuition in such a manner that by looking at the

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Table 6. *Most frequent verb constructions that reveal one preposition to one semantic role mapping relationship in PropBank*

Argument and verb class (Avc)	P(Avc|Prep)	Prep	Verb preposition construction frequency
ARG2–81	0.004037	of	[accuse.XX, of, 74];
ARG2–22.01–1	0.002892	with	[combine.XX, with, 34]; [merge.XX, with, 17]; [mix.01, with, 1]; [blend.XX, with, 1];
ARG1–87	0.002401	on	[concentrate.01, on, 44];
ARG1–10.06	0.001855	of	[deprive.XX, of, 12]; [divest.XX, of, 10]; [strip.XX, of, 5]; [relieve.01, of, 4]; [cleanse.XX, of, 2]; [cheat.01, of, 1];
ARG1–09.10	0.001691	for	[file.02, for, 30]; [fork.01, for, 1];
ARG2–79	0.001473	to	[devote.XX, to, 17]; [commit.01, to, 6]; [dedicate.XX, to, 4];
ARG3–11.01–1	0.001419	from	[slip.01, from, 23]; [ship.XX, from, 2]; [send.03, from, 1];
ARG3–45.06	0.014021	from	[fall.01, from, 65]; [drop.01, from, 32]; [climb.02, from, 29]; [decline.01, from, 25]; [jump.01, from, 17]; [grow.01, from, 16]; [surge.XX, from, 14]; [plunge.01, from, 13]; [jump.XX, from, 12]; [soar.01, from, 11]; [gain.01, from, 11]; [dip.01, from, 3]; [mushroom.XX, from, 3]; [tumble.XX, from, 3]; [plummet.XX, from, 2]; [decrease.XX, from, 2]; [diminish.XX, from, 1]; [climb.XX, from, 1];
ARG1–29.04	0.001255	in	[believe.XX, in, 23];

ARG0s using the preposition 'by' as the primary preposition are not included in the list. Threshold: $f \geq 20$, entropy $= 0$, max_p $= 1$.

collocation construction of the verb and the prepositional phrase, we are likely to come up with the semantic role they convey.

Sentences that mark the relationship of one particular preposition to one distinctive semantic role mapping are the best examples to demonstrate the predicative link between them. Table 6 (threshold: $f \geq 20$) introduces cases where the semantic roles for the class of verbs are marked by only one designated preposition. Every member verb in a verb class is using the same preposition to represent a particular semantic role. For example, for all the verbs listed for class 11.01 ('send' verbs), *from* is the primary preposition to indicate the argument sense of *ARG3*. 'Mix' verbs such as *combine, merge, mix*, and *blend* all use *with* to indicate the argument role of *ARG2* in verb class 22.01–1. These are added evidences to support the link between particular prepositions and the senses they represent when they are used with certain classes of verbs.

Although verbs that belong to the classes in Table 7 use more than one preposition to represent a specific semantic argument, the frequencies of the less frequent prepositions are almost negligible when compared with that of the primary

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Table 7. *Verb constructions that demonstrate the use of primary prepositions in representing semantic arguments in PropBank*

Argument and verb class (Avc)	Entropy	P(Avc| Prep)	Max_P	Prep	Verb preposition construction frequency
ARG4–11.02–1-1	0.1793	0.002019	0.9730	to	[slide.02, to, 33]; [bounce.02, to, 3]; [slide.02, below, 1];
ARG2–13.05.2–1	0.1793	0.002019	0.9730	from	[obtain.XX, from, 18]; [purchase.XX, from, 10]; [acquire.XX, from, 8]; [obtain.XX, in, 1];
ARG4–22.01–2	0.1720	0.002128	0.9744	to	[add.03, to, 38]; [add.02, for, 1];
ARG2–22.01–2	0.1687	0.002182	0.9750	to	[add.02, to, 38]; [add.03, to, 1]; [pool.01, into, 1];
ARG2–63	0.1654	0.002237	0.9756	on	[impose.XX, on, 40]; [impose.XX, against, 1];
ARG4–11.01–1	0.1565	0.002401	0.9773	to	[slip.01, to, 40]; [send.03, to, 3]; [slip.01, below, 1];
ARG2–13.01–1	0.1537	0.002455	0.9778	to	[lend.XX, to, 19]; [give.01, to, 14]; [lease.02, to, 7]; [refund.XX, to, 2]; [peddle.XX, to, 1]; [loan.01, to, 1]; [lend.XX, into, 1];
ARG1–37.07	0.1414	0.002728	0.9800	to	[respond.XX, to, 49]; [hint.XX, at, 1]; [remark.XX, on, 1];
ARG4–45.06	0.1367	0.035899	0.9848	to	[drop.01, to, 90]; [gain.01, to, 77]; [climb.02, to, 62]; [jump.01, to, 52]; [decline.01, to, 49]; [surge.XX, to, 35]; [jump.XX, to, 29]; [soar.01, to, 27]; [plunge.01, to, 27]; [tumble.XX, to, 21]; [grow.01, to, 21]; [plummet.XX, to, 8]; [dip.01, to, 7]; [climb.XX, to, 5]; [fall.01, below, 3]; [surge.XX, above, 2]; [mushroom.XX, to, 2]; [decrease.XX, to, 2]; [plunge.01, below, 1]; [tumble.XX, into, 1]; [climb.XX, above, 1]; [plunge.XX, into, 1]; [climb.02, above, 1];
ARG2–62	0.0960	0.004419	0.9877	at	[aim.02, at, 80]; [plan.XX, on, 1];
ARG2–87	0.0758	0.005892	0.9907	on	[focus.XX, on, 106]; [focus.XX, upon, 1]; [center.XX, on, 1];

Threshold: $f \geq 10$, entropy ≤ 0.2, max_p ≥ 0.9.

preposition. The tendency of using one particular preposition is still very strong. For example, *to* is a very frequent preposition in this table. When the 'give' verbs in class 13.01–1 are used in a dative sense, they select '*to*' to head a preposition phrase. For 'decrease' verbs such as *climb, jump, decline*, and *drop* in class 45.06, the meaning of the verbs determines that a prepositional phrase headed by '*to*' should be used to indicate the end point or end position. In cases like 'obtain' verbs, the source of

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Table 8. *Worst case verb classes in which multiple prepositions are used to represent the semantic roles with no notable tendency of favoring one particular preposition*

Argument and verb class	Entropy	P(Avc|Pre)	Max_P
ARG1–51.03.2 ('run' verbs)	3.2721	0.002619	0.2708
ARG2–47.06 ('balance' verbs)	3.2700	0.001910	0.3143
ARG2–45.04 ('alternating change of state' verbs)	3.2113	0.003655	0.2090
ARG2–17.01 ('throw' verbs)	2.7929	0.001528	0.3214
ARG2–09.07–1 ('spray/load' verbs)	2.7863	0.002455	0.3111
ARG2–36.01 ('correspond' verbs)	2.6669	0.003274	0.2167
ARG1–51.01–2 ('advance' verbs)	2.6031	0.001364	0.3200
ARG2–22.04 ('tape' verbs)	2.5850	0.000655	0.3333
ARG2–09.01 ('put' verbs)	2.5623	0.001528	0.4643
ARG2–47.02 ('billow' verbs)	2.5498	0.001037	0.3158
ARG2–11.04 ('carry' verbs)	2.5000	0.000436	0.2500

In this list, general, cross-class senses such as DIR, MNR, TMP, and CAU are not included. Threshold: entropy $\geq$ 2.5.

obtain, purchase, and *acquire* is almost always '*from*.' For all the verb preposition constructions in this list, the entropy value is not high but the maximum probability is close to 1. This means that although we use other prepositions representing the same sense, in an overwhelming majority of cases, one particular preposition is used much more often than the others. We can view this as strong support to the hypothesis that the semantic meaning can be inferred from the use of prepositions in the verb construction.

In Table 8, the rather high entropy values as well as the very low maximum probability values for the first most frequent prepositions show that in selecting prepositional phrases to represent semantic contents, these classes of verbs usually resist the use of central prepositions. The tendency to use primary prepositions for the semantic role realizations is very much blurred in these verb classes. For the most part, this means that the preposition frequencies are more or less evenly distributed. This obviously runs counter to what is assumed in our hypothesis. The first numbered Levin class of verbs in this list is 51.03.2 which belongs to the 'run' verbs. Although the preposition frequency is not very high, it seems that we cannot tell which preposition is significantly more important than the others in representing the particular semantic content. In these verb constructions, *ARG0* is usually the self_mover that is a living being moving under its own power, not with a separate vehicle. *ARG1* is the location where the self_mover moves. In VerbNet role categorization, this would usually be included in the 'Theme Rel Location(-preposition)' theta role pattern. Typical verbs preposition patterns in the class would include *to wander among, to walk around, to gallop down, to stumble over, to slog through, to travel to, to climb up*, and so on. With locative concepts, there are usually a variety of prepositions that can be used. This is where we have to be careful: we should not over-generalize about the use of central prepositions to represent

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

semantic contents, particularly if the prepositions are used to express locative or temporal concepts. For these concepts, it is difficult to predict which particular preposition to use when we are given information about the lexical unit and the semantic meaning (such as *Time* or *Location*) to convey.

4 Conclusions

In this article, we report on corpus evidence that supports the hypothesis about the semantic–syntactic correspondence between semantic arguments and their realizations by preposition structures. The analysis on corpus-based data from FrameNet and PropBank can be construed as a link in the study of frame semantics (Fillmore 1966) and the study of argument realization (Levin and Rappaport Hovav 2005). The statistical data shown in this investigation reveals a very strong tendency of preposition centrality in the syntactic realization of frame element concepts. It can be concluded that some information about the lexical unit together with information about the semantic role concepts can allow us to make predictions about the use of preposition structures. This corroborates the previous hypothesis about the binding relationship between semantic roles and their representations by preposition structures (Fillmore 1966). At the same time, some findings in the analysis provide food for thought for future studies. For example, for some frame elements, the centrality of preposition structures is not so marked. Preposition phrases selected to represent the spatial concepts seem to be evenly distributed over structures headed by a variety of prepositions. To avoid overgeneralization about the syntactic correlates of semantic arguments, the article also looked at some key factors that could explain the use of noncentral preposition structures in the syntactic realizations of arguments.

The findings about regularities in the use of the target lexical word and the particular prepositions can have implications in the linguistic study of argument realization and preposition semantics. Much of the work in mapping the deep semantic meaning to multiform prepositional expressions will shed light on natural language processing applications. There has been very good recent work towards the analysis of prepositions that aims at natural language processing tasks (LitKowski 2002; O'Hara and Wiebe 2003; Saint-Dizier 2005; Litkowski 2007). This corpus-based analysis will add knowledge to work done in the field. For example in prepositional phrase attachment disambiguation (Cortes and Vapnik 1995; Olteanu and Moldovan 2005), a corpus-based systematic description of the regularities in preposition and semantic role mapping can lend empirical support to rule-based approaches in disambiguating preposition phrases. In automatic semantic labeling (Gildea and Jurafsky 2002), once the mapping between LU-preposition constructions and the semantic roles they convey is defined, it will help towards an automatic argument parsing and tagging system for sentences with preposition structures. In the area of automatic semantic annotation of large corpora, or extending existing annotated databanks (Kipper, Snyder and Palmer 2004), the current analysis which is corpus-based itself can in turn help create a more robust system for better semantic role predictions and assignments based on the syntactic LU-preposition structures

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

and the semantic role relations. Also, for building a semantic interlingua that can be employed in NLP modules (Boas 2005), this study and its resultant data sets can be an important and useful source. For example, in machine translation, information with sufficient coverage of prepositions and the valences they are corresponding to will help the system understand the source language sentence, and choose the right prepositional phrases or alternative expressions to convey the meaning of the source language sentence.

Acknowledgements

My thanks go to the FrameNet team at University of California at Berkeley. I am grateful to Professor Charles J. Fillmore for his guidance and for his valuable comments on this article, to Dr. Collin Baker for his support in building the preposition knowledge base and the discussions in analyzing the data, and to Dr. Miriam R. L. Petruck and Michael Ellsworth for their assistance. I feel thankful to Professor Ash Asudeh, my supervisor at Carleton University, for his careful readings of the article, to Professor Andrew Brook and Professor Jo-Anne LeFevre of Carleton University for their suggestions in the process of preparing this article, to the three anonymous reviewers for their insightful comments. Special thanks also go to the management at Statistics Canada and my home institute at Carleton University for jointly funding my research at Berkeley.

References

Baker, C. F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (ACL-COLING'98)*, Montreal, Canada, pp. 86–90.

Baker, Collin F., and Josef Ruppenhofer. 2002. FrameNet's frames vs. Levin's verb classes. In J. Larson and M. Paster (eds.), *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*, pp. 27–38. Berkeley: Berkeley Linguistics Society.

Boas, Hans C. 2005. Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography* **18**(4): 445–78.

Boers, F., and Demecheleer, M. 1998. A cognitive semantic approach to teaching prepositions. *ELT Journal* **52**: 197–204.

Brent, Michael R. 1993. From grammar to lexicon: unsupervised learning of lexical syntax. *Computational Linguistics* **19**(2): 243–62.

Celce-Murcia, M. and Larsen-Freeman, D. 1999. *The Grammar Book*, 2nd ed., Boston, MA: Heinle and Heinle Publishers.

Cortes, C., and V. Vapnik. 1995. Support-vector networks. *Machine Learning* **20**(3): 273–97.

Cuyckens, H. and Radden, G. (eds.) 2002. *Perspectives on Prepositions*, Tübingen, Germany: Niemeyer.

Dorr, Bonnie J. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation* **12**(4): 271–322.

Dorr, Bonnie J., and Doug Jones. 1996. Role of word sense disambiguation in lexical acquisition: predicting semantics from syntactic cues. In *Proceedings of the 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, pp. 322–7.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Fillmore, Charles J. 1966. A proposal concerning English prepositions. In Francis P. Dinneen (ed.). *Report of the 17th Annual Round Table Meeting on Linguistics and Language Studies*, pp. 19–34, Washington, DC: Georgetown University Press.

Fillmore, Charles J. 1968. The case for case. In E. Bach and R. Harms (eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart and Winston.

Fillmore Charles J. 1971. Some problems for case grammar. In R. J. O'Brien (ed.), *Twenty-Second Annual Roundtable Monograph*, pp. 19–33, Washington, DC: Georgetown University Press.

Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, pp. 111–37. Seoul, South Korea: Hanshin.

Fillmore, Charles J., and Atkins, B. T. S. 1998. FrameNet and lexicographic relevance. In *Proceedings of the First International Conference on Language Resources and Evaluation, 28–30 May 1998*, Granada, Spain.

Fillmore, Charles J., Christopher R. Johnson, and Miriam R. L. Petruck. 2003. Background to Framenet. *International Journal of Lexicography* **16**(3): 235–50.

Fillmore, Charles J., Josef Ruppenhofer, and Collin F. Baker. 2004. FrameNet and representing the link between semantic and syntactic relations. In Huang, Chu-Ren and Winfried Lenders (eds.), *Computational Linguistics and Beyond*, pp. 19–59. Institute of Linguistics, Academia Sinica.

Gildea, Daniel and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics* **28**(3): 245–88.

Goldberg, Adele. 1995. *Constructions*. Chicago, IL: Chicago University Press.

Gropen, Jess, Steven Pinker, Michelle Hollander, Richard Goldberg, and Ronald Wilson. 1989. The learnability and acquisition of the dative alternation. *Language* **65**(2): 203–57.

Herskovits, Anette. 1986. *Language and Spatial Cognition, An Interdisciplinary Study of Prepositions in English*. Cambridge, UK: Cambridge University Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Kingsbury, P., and Palmer, M. 2002. From TreeBank to PropBank. In *Proceedings of Third International Conference on Language Resources and Evaluation, LREC 2002*, Las Palmas, Canary Islands, Spain.

Kipper, Karin. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD Thesis, University of Pennsylvania, Philadelphia, PA.

Kipper, Karin, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *AAAI-2000 Seventeenth National Conference on Artificial Intelligence*, Austin, TX, pp. 691–96.

Kipper, Karin, Benjamin Snyder, and Martha Palmer. 2004. Extending a verb-lexicon using a semantically annotated corpus. In *Fourth International Conference on Language Resources and Evaluation (LREC 2004),* Lisbon, Portugal.

Klavans, Judith and Min-Yen Kan. 1998. Role of verbs in document analysis. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montreal, Quebec, Canada, pp. 680–8.

Korhonen, Anna. 2002. Semantically motivated subcategorization acquisition. In *Proceedings of the ACL Workshop on Unsupervised Lexical Acquisition*, Philadelphia, PA, pp. 51–8.

Lakoff, George. 1987. *Women, fire and dangerous things: what categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Langacker, Ronald. 1987. *Foundations of Cognitive Grammar*, vol. I. Stanford, CA: Stanford University Press.

Langendonck, W. Van. 1974. Internally referring prepositions and the subcategorization of space and time denominations in a netherlandic case grammar. *Leuvense Bijdragen* 63, pp. 1–41.

Levin, Beth. 1993. *English Verb Classes and Alternations*. Chicago, IL: University of Chicago Press.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Levin, Beth and Malka Rappaport Hovav. 2005. *Argument Realization (Research Surveys in Linguistics)*. Cambridge, UK: Cambridge University Press.

Levow, Gina-Anne, Bonnie Dorr, and Dekang Lin. 2000. Construction of Chinese–English semantic hierarchy for information retrieval. Technical Report LAMP-TR-043/UMIACS-TR-2000–36/CS-TR-4145, University of Maryland, College Park.

Litkowski, Kenneth C. 2002. Digraph analysis of dictionary preposition definitions. In *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions, Association for Computational Linguistics*, Philadelphia, PA, pp. 9–16.

Litkowski, Kenneth C. 2007. CLR: integration of FrameNet in a text representation system. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic, Association for Computational Linguistics, pp. 113–6.

Litkowski, Kenneth C. and Orin Hargraves. 2007. SemEval-2007 task 06: word-sense disambiguation of prepositions. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Association for Computational Linguistics, Prague, Czech Republic, pp. 24–9.

Manning, Christopher D. 1993, Automatic acquisition of a large subcategorization dictionary from corpora. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics,* Columbus, OH, pp. 235–42.

Nübel, Rita. 1996. Knowledge sources for the disambiguation of prepositions in machine translation. In *Proceedings of the Workshop on Future Issues for Multilingual Text Processing, Pacific Rim International Conference on Artificial Intelligence*, Cairns, Australia.

O'Hara, Tom and Janyce Wiebe. 2003. Preposition semantic classification via Treebank and FrameNet. In *Proceedings of Conference on Natural Language Learning (CoNLL-03), 30–31 May, 2003*, Edmonton.

Olteanu, M. and D. Moldovan. 2005. PP-attachment disambiguation using large context. In *Proceedings of HLT/EMNLP 2005*, Vancouver, Canada.

Palmer, M., Kingsbury P. and Dan Gildea. 2005. The Proposition bank: an annotated corpus of semantic roles. *Computational Linguistics* **31**(1): 71–105.

Palmer, Martha and Zhibiao Wu. 1995. Verb semantics for English-Chinese translation. *Machine Translation* **10**(1–2): 59–92.

Petruck, Miriam R. L. 1996. Frame Semantics. In Jef Verschueren, Jan-Ola stman, Jan Blommaert and Chris Bulcaen (eds.), *Handbook of Pragmatics*. Philadelphia: John Benjamins.

Pütz, M. and Dirven, R. (ed.) 1996. *The Construal of Space in Language and Thought*. Berlin and New York: Mouton de Gruyter.

Radden, G. 1989. Figurative use of prepositions. In R. Dirven (ed.), *A User's Grammar of English: Word, Sentence, Text, Interaction*, pp. 551–76. Frankfurt, Germany: Lang.

Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, and Chris Johnson. 2006. *FrameNet: Theory and Practice*. E-book available at http://framenet.icsi.berkeley.edu/book/book.html.

Saint-Dizier, Patrick. 2005. PrepNet: a framework for describing prepositions: preliminary investigation results. In *International Workshop on Computational Semantics*, Tilburg, The Netherlands.

Stockwell, Robert P., Paul Schachter, and Barbara H. Partee. 1973. *The Major Syntactic Structures of English*. New York: Holt, Rinehart and Winston.

Sumita, Eiichiro, Osamu Furuse and Hitoshi Iida. 1993. An example-based disambiguation of prepositional phrase attachment. In *TMI-93: The Fifth International Conference on Theoretical and Methodological Issues in Machine Translation, July 14–16, 1993*, Kyoto, Japan, pp. 80–91.

Talmy, L. 1985. Lexicalisation patterns: semantic structure in lexical forms. In T. Shopen (ed.), *Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon*. Cambridge, UK: Cambridge University Press.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press

Trujillo, Arturo. 1992. Locations in the machine translation of prepositional phrases. In *Fourth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92), Empiricist vs. rationalist methods in MT, June 25–27, 1992*, Montreal, CCRIT-CWARC, pp. 13–20.

Tyler, Andrea and Vyvyan Evans. 2001. Reconsidering prepositional polysemy networks: the case of over. *Language* **77**(4):724–765.

Tyler, Andrea and Vyvyan Evans. 2003. *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. Cambridge, UK: Cambridge University Press.

Ushioda, Akira, David A. Evans, Ted Gibson, and Alex Waibel. 1993. The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora. In B. Boguraev and James Pustejovsky (eds.), *Proceedings of the Workshop on Acquisition of Lexical Knowledge from Text*, Columbus, Ohio, pp. 95–106.

Vandeloise, C. 1994. Methodology and Analyses of the Preposition. *Cognitive Linguistics* **5**(2): 157–84.

Wesche, Birgit. 1986. At ease with "At". *Journal of Semantics* **5**(4): 385–99.

Ye, Patrick, and Timothy Baldwin. 2007. MELB-YB: preposition sense disambiguation using rich semantic features. In: *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, Association for Computational Linguistics, Prague, Czech Republic, pp. 241–4.

Zelinsky-Wibbett, Cornelia. 1993. *The Semantics of Prepositions: From Mental Processing To Natural Language Processing*. Berlin, Germany: Mouton de Gruyter.

https://doi.org/10.1017/S1351324908005044 Published online by Cambridge University Press