Correspondences between Syntactic Form and Meaning: From Anarchy to Hierarchy

Justin Peterson
College of Computing
Georgia Institute of Technology
Atlanta, Georgia 30332
justin@cc.gatech.edu

Dorrit Billman
School of Psychology
Georgia Institute of Technology
Atlanta, Georgia 30332
billman@cc.gatech.edu

Abstract

If we are to develop language processing systems that model human capabilities and performance, we must identify correspondences between the grammatical features and meaning of language and employ them in our computational models of sentence interpretation. In this paper, we present a computational model of sentence interpretation and a theory of compositional semantics. Our model provides a method for addressing a range of lexical novelty (e.g., novel verbs, novel uses of known verbs), relying on a semantic representation that maintains principled correspondences with syntactic form. In our approach, syntactic structure preserves critical information about the hierarchical structure of semantic interpretations. This property of the semantic representation along with restrictions on semantic interpretations enable the model to infer the semantics of novel verbs, disambiguate the semantics of known verbs, and determine the contributions that verb arguments make to sentence interpretation in a constrained and principled manner. This research offers a fruitful approach for using linguistic analysis to address the recovery of meaning in natural language processing systems.

Introduction

Words have meaning as do longer stretches of language. On this we all agree. But the meaning of utterances is not exhausted by the meaning of their constituent words. Identifying this other, non-lexical, source of meaning is an important project for cognitive science and its constituent disciplines. For psychology, it provides hypotheses about how people derive meaning from language. The most relevant case is the interpretation of utterances with novel verbs; here, an appeal to lexical meaning is clearly an insufficient explanation of what people can do. Correspondences between syntax and meaning also figure into an explanation of how people come to acquire their word meanings in the first place. For artificial intelligence, correspondences between syntax and meaning enable constrained, principled, and efficient natural language processing. They provide information about meaning which is reliable, local, and consistent. They can also identify what aspects of meaning **must** depend on appeal to world knowledge. For linguistics, an analysis of the relations between form and meaning is critical to finding regularities within the lexicon. And it is a necessary step in understanding the generativity of semantics as well as syntax. Finally, the identification of correspondences between syntax and meaning is a prerequisite for developing language processing systems that model human capabilities and performance.

In this paper, we present a computational model that uses correspondences between syntax and verb semantics to interpret sentences. In doing so, we demonstrate that syntax is a rich source of evidence for sentence interpretation. The theory of syntactic-semantic correspondences that underlies our model departs from past proposals (Rappaport & Levin, 1988; Pinker, 1989; Jackendoff, 1990) because it holds that the structure of syntax preserves critical aspects of the hierarchical structure of semantic interpretations. We articulate this preservation of hierarchical information in terms of a set of structural correspondences that operate over the mapping of syntactic form onto semantic interpretations. By combining these structural correspondences with a compositional semantic representation, the model can infer the semantics of novel verbs, disambiguate the semantics of known verbs, and determine the contributions that verb arguments make to sentence interpretation in a constrained and principled manner.

We begin with a discussion of the problems of verb semantics and the past proposals that have sought to resolve them. So that the our work can be clearly distinguished from that of others, we focus on the critical role that syntactic-semantic correspondences play in theories of verb semantics. Second, we present the Prominence Correspondence rule and describe its implications for syntactic-semantic correspondence. Prominence Correspondence is central to the preservation of the hierarchical structure of semantic interpretations in syntax. Third, we present our computational model and demonstrate how it infers the semantics of novel verbs, disambiguates the semantics of known verbs, and determines the contributions that verb arguments make to sentence interpretation. We conclude with a discussion of the specific benefits our work provides to the constituent disciplines of Cognitive Science.

Background

In verb semantics, researchers investigate what types of representation are appropriate for articulating the interpretation of a verb and its arguments and how these representations relate to syntactic form. Typically, verbs are taken to denote events and states in people's representations of the world, and their arguments are taken to play the roles of causal agents, objects undergoing change, destinations, and locations in these events and states. The goal of any theory of verb semantics is to provide 1) a representation that can articulate these aspects of the sentence interpretation and 2) specify the relation between this representation and the syntactic representation of the verb and its arguments. Theories of verb semantics ask what the appearance of a subject, object or particular preposition tells us about the states and events denoted by the verb,

and what the appearance of a noun phrase as a subject, object, or prepositional object tell us about the role its referent plays in these event and states.

In the past, two approaches to verb semantics have been proposed: Thematic Role theories and Compositional Semantics theories. Thematic Role theories (Bruce, 1975) represent verb semantics in terms of a set of labels (i.e., thematic roles) and specify the relation between semantic interpretations and syntax in terms of mappings between role labels and verb argument positions. Unfortunately, these theories are inadequate for representing certain aspects of meaning that are signaled by the syntax of verbs and their arguments (Rappaport & Levin, 1988).

Compositional Semantics theories (Pinker, 1989), on the other hand, represent verb semantics in terms of semantic tree structures and specify the relation between semantic interpretations and syntax in terms of mappings between verb argument positions and positions in semantic trees (i.e., semantic arguments). In these theories, the set of possible semantic trees is defined in terms of the allowable combinations of semantic elements. Although Compositional Semantics theories are sufficiently expressive to represent the aspects of meaning that are signaled by grammatical features, this increased expressiveness has made it hard to give a principled account of which mappings between syntax and semantics are allowed and which are excluded.

We offer a theory of compositional semantics that provides a principled account of the mapping between syntax and semantics. The mapping projects combinations of words and phrases in syntax onto combinations of semantics elements and expressions in semantics. Central to the principled nature of this mapping are *structural correspondences*. Structural correspondences map the structural relations of syntax onto the structural relations of semantics, deriving much of the hierarchical structure of semantic trees from syntactic structure. In the next section, we identify the general characteristics of this semantic representation and discuss the rules of structural correspondence. The majority of the section is devoted to introducing Prominence Correspondence, the dominant structural correspondence rule.

Correspondences between Syntax and Semantics

We propose a compositional semantic representation that expresses a subset of the conceptual content denoted by the combination of a verb and its arguments. Semantic interpretations are expressed in terms of strictly hierarchical semantics trees that maintain *lexical* and *structural* correspondences with the syntactic representations of the sentences that invoke them. *Lexical* correspondences identify relations between the syntactic categories of words and phrases and the semantic categories of elements and expressions in the semantic representation. *Structural* correspondences specify how these semantic elements and expressions combine. These properties of the representation along with restrictions upon the types of combinations that semantic elements and expressions can form provide the basis for a model of sentence interpretation as well as explain a number of linguistic phenomena.

Four different types of combinations of words and phrases occur within the syntactic structure of a verb and its arguments: 1) noun phrases are the arguments of the verb, 2) prepositions are complements of the verb, 3) noun phrases are objects of prepositions, and 4) noun phrases maintain remote relations with each other. In our approach, each of these syntactic combinations conveys distinct information about the combinations of semantic elements in the semantic interpretation, evoking four different rules of structural correspondence. Three of these structural correspondence rules claim that syntactic arguments of prepositions and verbs designate corresponding semantic arguments: roughly, if x is an argument of y in syntax, then the semantic element designated by x is an argument of the semantic element designated by y in semantics. Although these structural correspondences specify important constraints on the local combinations of semantic elements and their arguments, they tell us little about how these local combinations combine with each other. This information is provided by the remote relations between noun phrase arguments.

Remote combinations of noun phrase arguments correspond to remote combinations of their mappings in the semantic interpretation. We describe this property of syntactic-semantic correspondence in terms of syntactic and semantic *prominence relations*[1]. Prominence relations are hierarchical relations that are defined in terms of nodes and dominance relations, as follows

Prominence Relation

A node x is more prominent than another node y if and only if the immediate dominator of x dominates the immediate dominator of y.

Syntactic prominence relations between noun phrase arguments are defined by the syntactic structure of sentences, and semantic prominence relations are defined by the semantic tree structures these sentence invoke. Syntactic prominence relations map into semantic prominence relations by Prominence Correspondence:

Prominence Correspondence Rule

Given a set of noun phrase arguments of a verb (i.e., subject, object, prepositional object) and a set of their semantic denotations, a noun phrase argument x is syntactically more prominent than another y if and only if the semantic denotation of x is semantically more prominent the semantic denotation of y.

Simply put, Prominence Correspondence states that the hierarchical relations that arise between verb arguments in syntax preserve aspects of the hierarchical structure of semantic interpretations.

Consider the semantic interpretations of the following sentences:

1 a. The silver dollar is in the Potomac.
  b. The silver dollar went into/went in/entered the Potomac.

The corresponding syntactic and semantic trees for these sentences appear in Figure 1. In the syntactic structure for (1a), the noun phrase *the silver dollar* is more prominent than the noun phrase *the Potomac*; **S**, the immediate dominator of *the silver dollar*, dominates **PP**, the immediate dominator

[1] Although *relations of prominence* appear in (Grimshaw, 1991), our use of the term is particular to our own work.

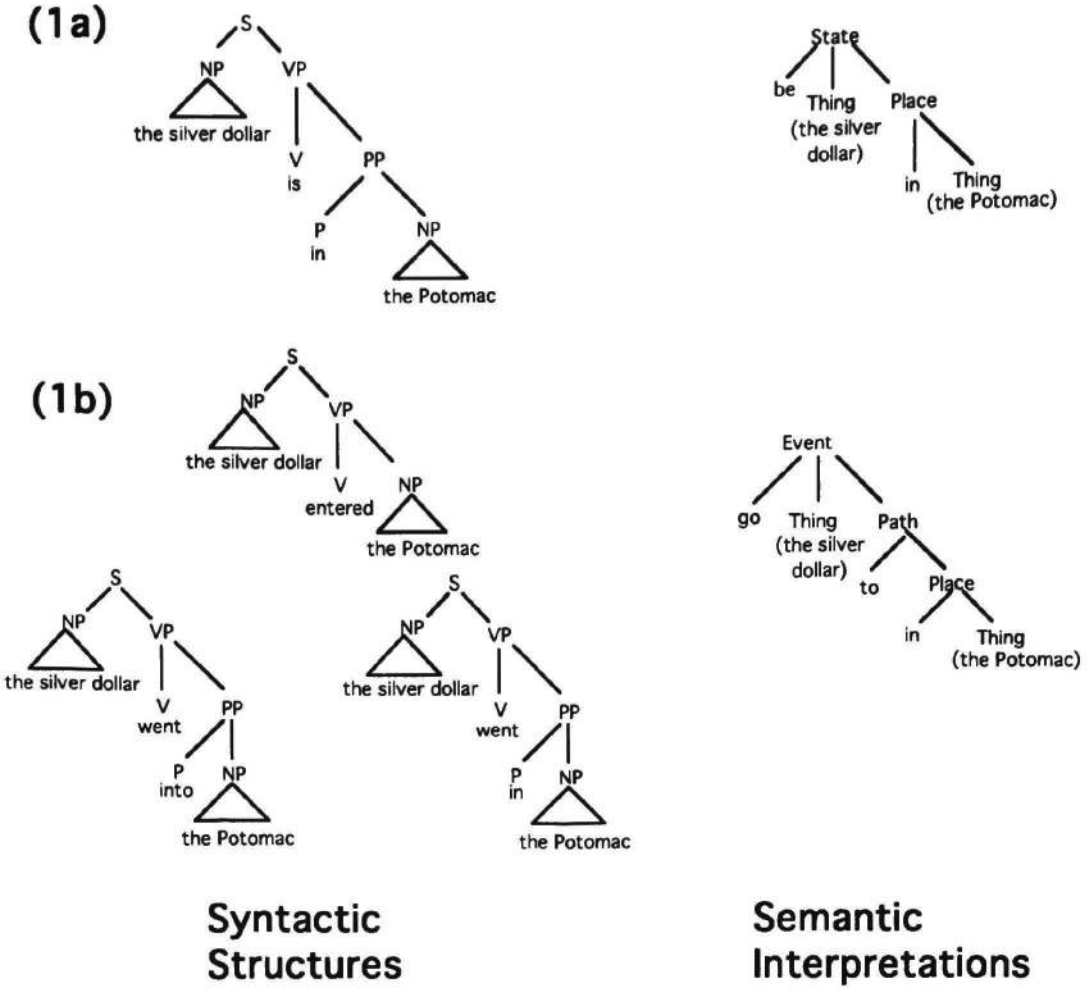

Figure 1: Examples of Prominence Correspondence

of *the Potomac*. Consistent with Prominence Correspondence, **(Thing the silver dollar)**, the semantic mapping of the noun phrase *the silver dollar*, is semantically more prominent than **(Thing the Potomac)**, the semantic mapping of the noun phrase *the Potomac*; **State**, the immediate dominator of **(Thing the silver dollar)**, dominates **Place**, the immediate dominator of **(Thing the Potomac)**.

Prominence Correspondence is undeterred by semantic incorporation[2]. The sentences in (1b) are sentences with equivalent semantic interpretations that demonstrate various levels of semantic incorporation: (i) *enter* incorporates the semantics of *into*, (ii) *went* in its *went in* usage incorporates the semantics of *to*, and (iii) *went* in its *went into* usage does not incorporate any prepositional semantics at all. Although they differ with respect to semantic incorporation, their syntactic and semantic prominence relations are the same: the noun phrase *the silver dollar* is syntactically more prominent than the noun phrase *the Potomac* in all three sentences, and thus **(Thing the silver dollar)** is semantically more prominent than **(Thing the Potomac)** in all three sentences.

The hierarchical relations that arise between verb arguments in syntax preserve information about the hierarchical structure of semantic interpretations. This information is critical to inferring the semantics of novel verbs and disambiguating the semantics of known verbs.

The Computational Model

The purpose of the computational model is to produce semantic interpretations for natural language sentences. The model uses correspondences between syntax and semantics to generate semantic interpretations when the verb is novel, and to identify the semantic composition of a verb and its arguments for both novel and known verbs.

The model breaks sentence processing down into syntactic parsing and semantic interpretation. It generates a syntactic representation given a natural language sentence, and in concert, it constructs a semantic interpretation combining partial information provided by the lexical correspondences between syntax and semantics and the semantic mappings of known words. The construction of these semantic interpretations is

[2]Semantic incorporation occurs when a verb specifies semantic elements that are commonly denoted by prepositions; semantic incorporation verbs *incorporate* the semantics of prepositions (e.g., *enter* incorporates *into*). For a discussion of the difficulties that semantic incorporation creates for theories of syntactic-semantic correspondence, see (Jackendoff, 1987).

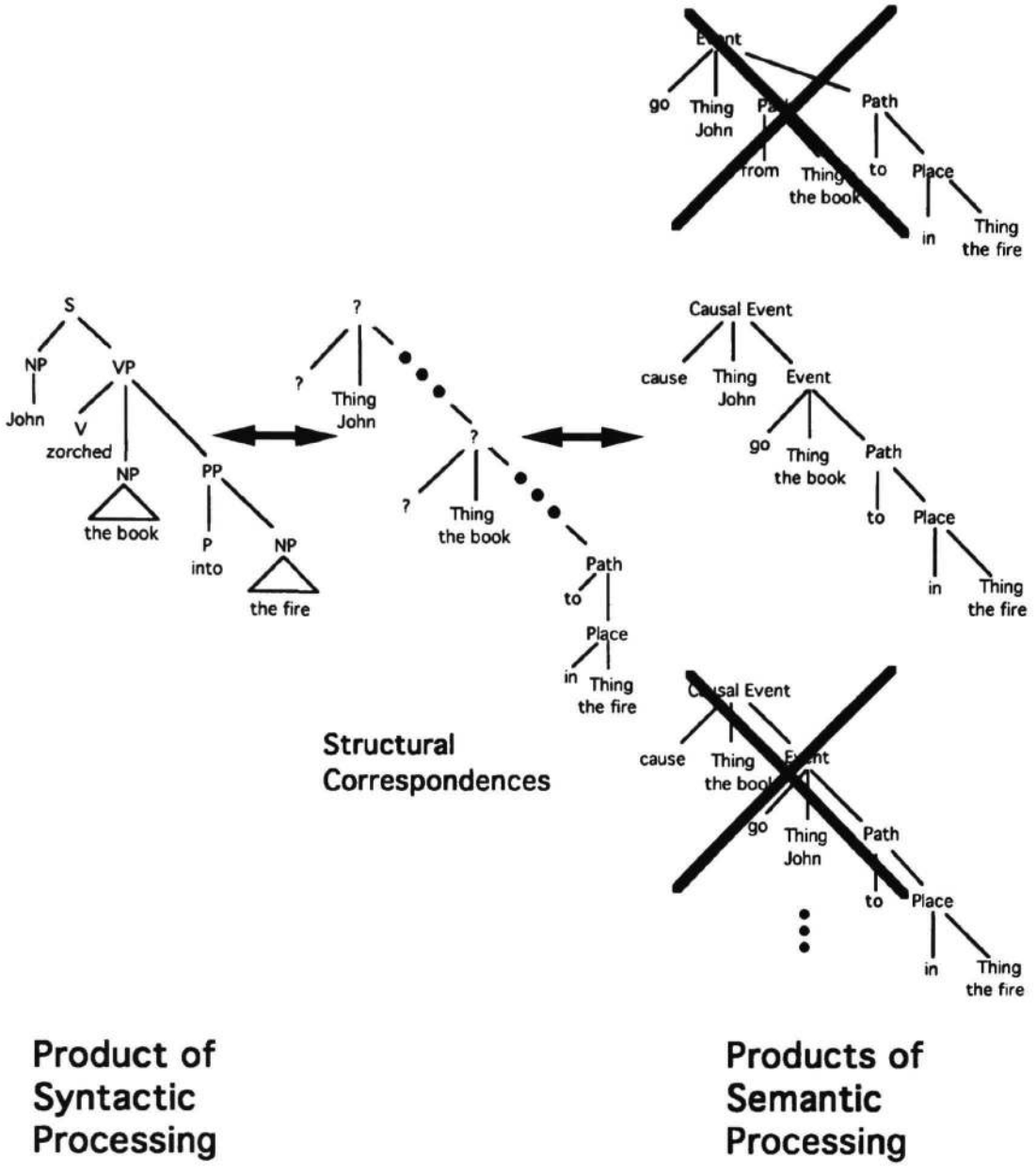

Figure 2: Interpreting *John zorched the book into the fire*

constrained by the structural correspondences between syntax and semantics and the restrictions upon the types of combinations that semantic elements can form.

Novel Verbs

As an illustration of how our system produces interpretations for novel verbs, consider the following nonsense verb example:

2. John zorched the book into the fire.

Given the surface form in (2), the system begins engaging in three activities 1) generating a syntactic parse tree, 2) using the structural and lexical correspondences to map aspects of syntactic structure into semantic structure and 3) constructing a semantic interpretation. All of these activities occur concurrently and interactively.

For the sentence in (2), the syntactic parse tree in Figure 2 is generated. As this parse tree is being constructed, the system produces the mappings into semantic representation using both the lexical correspondences and a semantic lexicon. Once these mappings are made available, they are combined to form semantic interpretations in accordance with the restrictions on semantic trees. Since these mappings are quite numerous, only a few are provided in Figure 2. As these semantic interpretations are being constructed, the structural correspondences between syntax and semantics are employed. These correspondences derive aspects of the structure of the semantic tree from the syntactic parse tree. For example, given that the noun phrase *John* is more prominent than the noun phrase *the book* in the syntactic tree, the system uses Prominence Correspondence to determine that **(Thing John)** is more prominent than **(Thing the book)** in the semantic

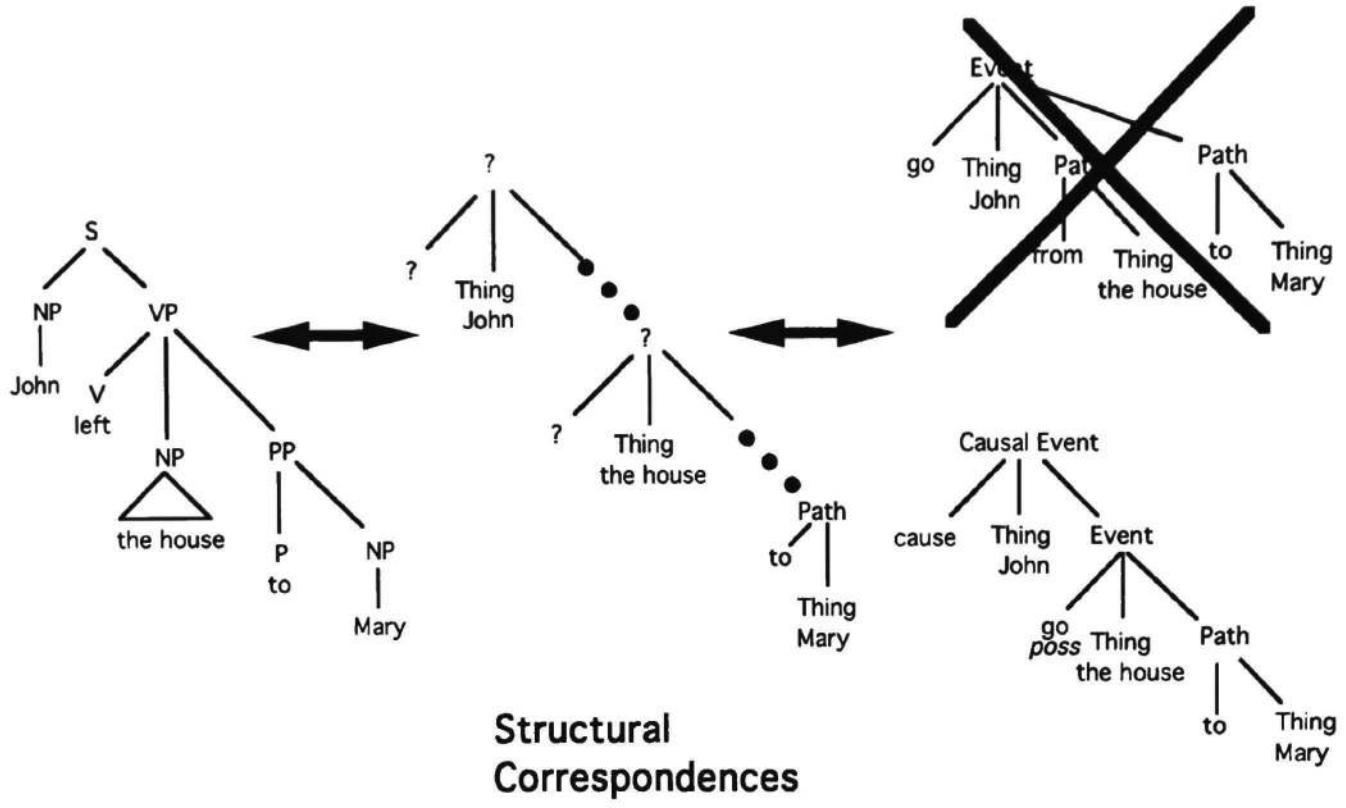

Figure 3: Interpreting *John left the house to Mary*

tree. Structural correspondences specify the locations of the semantic arguments (e.g. **(Thing John)**) with respect to each other and the structure of the semantic interpretation of the verb. As illustrated in Figure 2, the structural correspondences constrain but do not completely specify the structure of the semantic interpretation.

Finally, the system combines the structural constraints with the semantic interpretations, ruling out those interpretations that are inconsistent. As illustrated in Figure 2, a consistent semantic interpretation is then produced as output. This semantic interpretation states that the sentence in (2) describes a causal event in which 1) the book was not in the fire at the onset, 2) John caused something to happen, 3) what happened was that the book went along a path leading to the fire, and 4) the result is that the book is now in the fire.

Disambiguating Known Verbs

As an illustration of how the system disambiguates known verbs, consider

3. John left the house to Mary.

This sentence is particularly interesting because its interpretation involves two implications of the Prominence Correspondence rule: 1) semantic incorporation verbs may exhibit the same grammatical features as non-semantic incorporation verbs, and 2) verb arguments that designate equally prominent semantic elements must be equally prominent in the syntactic parse tree. Consider the verb *left*. It may denote either the causal transfer of possession as in (3) or a departure as in *John left the house and went to the store*. In its departure usage, *left* incorporates the semantics of the preposition *from*. In its causal transfer of possession usage, *left* does not incorporate any prepositional semantics. Both of these uses of *left* are grammatically indistinguishable in the sentence fragment *John left the house*. However, when combined with the prepositional phrase *to Mary*, the departure usage of *left* can be ruled out by Prominence Correspondence.

As illustrated in Figure 3, only the semantic prominence relations in the semantic tree corresponding to the causal transfer

of possession usage of *left* (the lower semantic tree in Figure 3) are consistent with the syntactic prominence relations. In the semantic tree corresponding to the departure usage of *left* (upper semantic tree in Figure 3), the arguments of the semantic elements **from** and **to**, **(Thing the house)** and **(Thing Mary)**, are equally prominent. To be in accord with Prominence Correspondence, these two semantic arguments must correspond to equally prominent noun phrase arguments in the syntactic tree. However, as Figure 3 illustrates, this is not the case; the noun phrase *the house* is syntactically more prominent than the noun phrase *Mary*.

Consider how the system operates on the sentence in (3). At the onset, both uses of *left* are retrieved and processed until *to Mary* is encountered. At this point, Prominence Correspondence excludes the departure use of *left*, leaving a semantic interpretation of the sentence that describes a causal event in which *John caused the house to come into Mary's possession.*

Conclusion

The theory of semantics and syntactic-semantic correspondence presented above offers a number of benefits to the various Cognitive Science disciplines. Though it is not a theory of language development or performance, it has important implications for word meaning acquisition and human sentence processing behavior. First, it articulates a body of knowledge that competent speakers could employ in acquiring the meanings of new words given syntactic evidence. As discussed in (Landau & Gleitman, 1985), such knowledge is critical to language learners given the limits of acquiring word meanings, especially verb meanings, from simply observing the environment. More importantly, language learners (unfamiliar with the lexical semantics of natural language) could use the formal correspondences between syntax and semantics to acquire information about semantics in general as well as a particular word's meanings. The theory also provides for the development of performance models such as that envisioned in (Carlson & Tanenhaus, 1988). It offers principles of interaction between syntactic structure and semantic ambiguity resolution. And since intermediate semantic decisions in the theory are related to intermediate syntactic decisions, it enables a fine-grained model of interaction in which intermediate semantic results are fed back to syntactic processing.

The theory also benefits linguistics. The explanatory power of the semantic representation and its formal correspondences with syntax is greater than that of previous approaches (Rappaport & Levin, 1988; Pinker, 1989; Jackendoff, 1990). In particular, it explains the mappings of semantic arguments (e.g., causal agents) into syntactic positions (e.g., subject) rather than simply stipulating the mappings as has been done previously. In our account, the mapping of semantic arguments into syntax is directly associated with their position in the semantic interpretation of the sentence.

More important are the benefits to computing, specifically to the development of intelligent systems that communicate in natural language. Our model offers a robust, extendible, portable method for performing semantic interpretation. Robustness is achieved by reducing the dependence upon the semantic lexicon and conceptual and world knowledge. The system is easily extended since adding elements to the semantic representation requires no change to the correspondences rules. Finally, semantic processing as well as syntactic processing can be transferred to new problems without modification.

Our research offers a fruitful approach for using linguistic analysis to address recovery of meaning in natural language processing systems. We have provided a model of sentence interpretation that addresses a range of lexical novelty (e.g., novel verbs, novel uses of known verbs) using parsimonious, broad semantic knowledge and furnishes a principled account of the correspondences between the form and meaning of language.

Acknowledgements

During the course of this work, the first author has been supported in part by a research gift from Northern Telecom. We would also like to thank Kurt Eiselt, Ashwin Ram, and Richard Billington for their helpful feedback and criticism.

References

Bruce, B. (1975) Case Systems for Natural Language. *Artificial Intelligence* 6, 327–360.

Carlson, G. & Tanenhaus, M. (1988). Thematic Roles and Language Comprehension. In W. Wilkins (Ed.), *Syntax and Semantics, Volume 21: Thematic Relations* (pp. 263–288). New York,NY: Academic Press.

Grimshaw, J. (1991). *Argument Structure*. Cambridge,MA: MIT Press.

Jackendoff, R. (1987) The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry* 18, 369–411.

Jackendoff, R. (1990). *Semantic Structures*. Cambridge,MA: MIT Press.

Landau, B. & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge,MA: Harvard University Press.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge,MA: MIT Press.

Rappaport, M. & Levin, B. (1988). What to do with Theta-roles. In W. Wilkins (Ed.), *Syntax and Semantics, Volume 21: Thematic Relations* (pp. 7–36). New York,NY: Academic Press.