A Probabilistic Model of Early Argument Structure Acquisition

by

Afra Alishahi

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

Copyright © 2008 by Afra Alishahi

Abstract

A Probabilistic Model

of Early Argument Structure Acquisition

Afra Alishahi

Doctor of Philosophy

Graduate Department of Computer Science

University of Toronto

2008

Developing computational algorithms that capture the complex structure of natural language is an open problem. In particular, learning the abstract properties of language only from usage data remains a challenge. In this dissertation, we present a probabilistic usage-based model of verb argument structure acquisition that can successfully learn abstract knowledge of language from instances of verb usage, and use this knowledge in various language tasks. The model demonstrates the feasibility of a usage-based account of language learning, and provides concrete explanation for the observed patterns in child language acquisition.

We propose a novel representation for the general constructions of language as probabilistic associations between syntactic and semantic features of a verb usage; these associations generalize over the syntactic patterns and the fine-grained semantics of both the verb and its arguments. The probabilistic nature of argument structure constructions in the model enables it to capture both statistical effects in language learning, and adaptability in language use. The acquisition of constructions is modeled as detecting similar usages and grouping them together. We use a probabilistic measure of similarity between verb usages, and a Bayesian framework for clustering them. Language use, on the other hand, is modeled as a prediction problem: each language task is viewed as finding the best value for a missing feature in a usage, based on the available features in that same

usage and the acquired knowledge of language so far. In formulating prediction, we use the same Bayesian framework as used for learning, a formulation which takes into account both the general knowledge of language (i.e., constructions) and the specific behaviour of each verb. We show through computational simulation that the behaviour of the model mirrors that of young children in some relevant aspects. The model goes through the same learning stages as children do: the conservative use of the more frequent usages for each individual verb at the beginning, followed by a phase when general patterns are grasped and applied overtly, which leads to occasional overgeneralization errors. Such errors cease to be made over time as the model processes more input.

We also investigate the learnability of verb semantic roles, a critical aspect of linking the syntax and semantics of verbs. In contrary to many existing linguistic theories and computational models which assume that semantic roles are innate and fixed, we show that general conceptions of semantic roles can be learned from the semantic properties of the verb arguments in the input usages. We represent each role as a semantic profile for an argument position in a general construction, where a profile is a probability distribution over a set of semantic properties that verb arguments can take. We extend this view to model the learning and use of verb selectional preferences, a phenomenon usually viewed as separate from verb semantic roles. Our experimental results show that the model learns intuitive profiles for both semantic roles and selectional preferences. Moreover, the learned profiles are shown to be useful in various language tasks as observed in reported experimental data on human subjects, such as resolving ambiguity in language comprehension and simulating human plausibility judgements.

Dedication

To my parents,

who taught me to believe in myself.

Acknowledgements

I am forever grateful to my supervisor, Suzanne Stevenson, for believing in me and guiding me through the maze of the academic world. She has been an inspiring teacher and a wonderful friend. I wish to thank my advisory committee, Graeme Hirst, Gerald Penn, and Richard Zemel for their stimulating questions, valuable feedback and general friendliness, which made our annual meetings something to look forward to. I am also thankful to my external examiner, Joshua Tenenbaum, for his encouraging comments and for taking the time to fly to Toronto and attend my defense.

My PhD studies turned out to be an unforgettable experience, mostly thanks to the support from my colleagues and friends. I was lucky to be part of a great group of students in the computational linguistics group: Vivian Tsang, Saif Mohammad, Diana Inkpen, Faye Baron, Yun Niu, Jane Li, Robert Swier, Christopher Collins, Paul Cook, Michael Demko, Ulrich Germann, and Christopher Parisien. Their kindness helped me feel at home, and their enthusiasm was contagious. I would like to especially thank my dear friend Afsaneh Fazly, who volunteered to be my mentor even before I arrived in Toronto, and has been there for me ever since.

Going through this process would not have been possible without the emotional support from my friends outside the university. Golnaz Tajeddin has a special place in my heart, and has been by my side in happy and sad times. Golnaz, Afsaneh, Iman Hajizadeh, Reza Azimi, Solmaz Kolahi and Vahid Hashemian have been wonderful companions for things to see and do, both in and out of town. I would like to thank Kaave Lajevardi, who has always been a great friend, and without whose encouragement I would not have started this path. My special thanks to Marge Coahran, Amber Wilcox-O'Hearn, Nina Theissen, Jane Yu and Mahsa Moallem for our joyous coffee klatches, and to Jane Morris, Vivian Tsang and Afsaneh Fazly for our girls nights out. Most especially, my heartfelt gratitude goes to my aunt and uncle, Nahid Keyvani and Asad Nakhaei, for providing me with a much needed piece of family and home.

A note to my family: maman, baba, Yasi and Kasra, at times I missed you so much that it was hard to stay motivated and keep going with my studies. But your love has been my main source of energy through the years. This is for you.

Contents

1 Child Argument Structure Acquisition **1**

1.1 Overview 1

1.2 Challenges of Early Verb Learning 2

1.3 Acquisition of Verb Argument Structure 5

1.4 Semantic Roles and Selectional Preferences 7

1.5 Focus of the Study 9

2 A Bayesian Model of Early Verb Learning **12**

2.1 Overview 12

2.2 The Representation of Probabilistic Constructions 14

2.2.1 The Representation of Meaning 15

2.2.2 Input Pairs and Extracted Frames 16

2.2.3 Constructions as Groups of Frames 18

2.3 Acquisition of Constructions 20

2.3.1 Overview of the Acquisition Process 20

2.3.2 Details of the Probability Model for Learning 23

2.4 Language Use as Prediction 25

2.5 The Expected Behaviour of the Model 27

2.6 Summary 28

3 Acquisition of Verb Argument Structure **30**

3.1 Overview . . . 30

3.2 Theories of Argument Structure Acquisition . . . 32

3.3 Related Computational Models . . . 35

3.4 A Computational Model of Verb Argument Structure Acquisition . . . 39

3.4.1 Basic Properties of the Input . . . 40

3.4.2 Argument Structure Frames . . . 42

3.5 Experimental Set-Up . . . 43

3.5.1 The Input-generation Lexicon . . . 43

3.5.2 The Input Corpora . . . 46

3.5.3 Setting the Parameters of the Model . . . 48

3.6 Experimental Results . . . 50

3.6.1 Construction Formation . . . 51

3.6.2 Learning Curves . . . 56

3.6.3 Generalization . . . 61

3.6.4 Overgeneralization and Recovery . . . 67

3.6.5 Productive Generalization . . . 69

3.7 Discussion of the Results . . . 74

3.7.1 The Emergence of Constructions . . . 74

3.7.2 Generalization in Comprehension and Production . . . 76

3.7.3 Mechanisms for Recovery from Overgeneralization Errors . . . 77

3.7.4 Comparing Recovery and Productive Generalization . . . 80

3.8 Summary of Contributions . . . 81

4 Learning and Use of Verb Semantic Roles **84**

4.1 Overview . . . 84

4.2 Theories of Semantic Role Learning . . . 85

4.3 Related Computational Models . . . 88

4.4 A Computational Model for Learning Verb Semantic Roles 90
4.4.1 The Input and the Argument Structure Frames 91
4.4.2 Probabilities of Semantic Properties 93
4.4.3 Representation of Semantic Profiles 95
4.5 Experimental Set-Up 96
4.5.1 The Input Corpora 96
4.5.2 Setting the Parameters of the Model 97
4.6 Experimental Results 97
4.6.1 Formation of Semantic Profiles for Roles 98
4.6.2 Multiple Possible Roles for a Position 102
4.6.3 Verb-Based vs. General Semantic Profiles 104
4.6.4 Using Semantic Profiles in Comprehension 106
4.7 Discussion of the Results 109
4.7.1 Where Do Abstract Roles Come From? 109
4.7.2 Age-Dependence of Semantic Roles 110
4.8 Summary of Contributions 113

5 Learning Verb Selectional Preferences 115

5.1 Overview 115
5.2 Verb Selectional Preferences 117
5.3 Related Computational Models 119
5.4 A Computational Model of Learning Verb Selectional Preferences 121
5.4.1 The Input and the Argument Structure Frames 122
5.4.2 Predicting Semantic Profiles for Verbs 123
5.4.3 Verb-Argument Compatibility 126
5.5 Experimental Set-up 127
5.5.1 The Input Corpora 127
5.5.2 Setting the Parameters of the Model 127

5.6 Experimental Results . . . 128
5.6.1 Formation of Semantic Profiles for Verbs . . . 128
5.6.2 Verb-Argument Plausibility Judgments . . . 133
5.6.3 Implicit Object Alternations . . . 134
5.7 General Discussion . . . 136
5.7.1 The Effect of Processing More Input Data . . . 138
5.7.2 Noun Selectional Preferences . . . 139
5.8 Summary of Contributions . . . 139

6 Summary and Future Directions **141**
6.1 Summary of Contributions . . . 141
6.1.1 Representation . . . 142
6.1.2 Underlying mechanisms . . . 142
6.1.3 Properties of the model . . . 143
6.2 Short-term Extensions . . . 145
6.2.1 Using Child Data as Input . . . 145
6.2.2 Compressed Representation of Constructions . . . 146
6.2.3 Sensitivity to the Order of Input . . . 147
6.2.4 The Generalization Factor . . . 148
6.2.5 Argument Indexing in Syntactic Patterns and Semantic Representations . . . 149
6.3 Future Directions . . . 150
6.3.1 Learning Word-Meaning Mapping . . . 150
6.3.2 Language Processing . . . 152
6.3.3 Second Language Acquisition . . . 153
6.3.4 Language Evolution . . . 154

Appendices **155**

A Constructions in an Argument Structure Acquisition Simulation **155**

B Constructions in a Semantic Role Learning Simulation **159**

Bibliography **164**

List of Figures

2.1 An input pair and extracted frame for a verb argument structure. 17

2.2 A portion of the lexicon showing an intransitive construction, with a dominant syntactic pattern Subj(ect) V(erb), and a transitive construction, with a dominant syntactic pattern Subj(ect) V(erb) Obj(ect). Numbers on links represent the observed frequency of the verb in a frame that participates in that construction. For example, *eat* has been seen once in an intransitive construction and 11 times in a transitive construction. 19

2.3 The legend for symbols representing verbs and constructions in the diagrams. 19

2.4 Processing a new frame: (a) A snapshot of the lexicon before the new frame is processed. (b) Addition of the new frame to the lexical entry for its main verb, in this case v_n. All existing constructions and a potential new construction are considered as a match for the new frame, as indicated by the dashed lines. Two alternatives are possible, as shown in panels c1 and c2. (c1) The new construction is the best match, hence it is added to the lexicon (and becomes k_{m+1}), and the new frame is linked to it; OR (c2) One of the existing constructions (in this case k_1) is the best match, hence the new frame is linked to it, and no new construction is created. . 22

3.1 A sample input pair 39

3.2 The semantic hierarchy 40

3.3 A sample lexeme 47

3.4 The probability distribution of semantic primitives in the transitive construction at two different points in learning. Darker squares correspond to higher probabilities. 51
3.5 The probability distribution of semantic primitives in the intransitive construction at two different points in learning. Darker squares correspond to higher probabilities. 52
3.6 The probability distribution of semantic primitives in the ditransitive construction at two different points in learning. Darker squares correspond to higher probabilities. 52
3.7 Transitive construction for the extended lexicon. Darker squares correspond to higher probabilities. 56
3.8 Intransitive construction for the extended lexicon. Darker squares correspond to higher probabilities. 57
3.9 Cumulative accuracy of the model for predicting syntactic pattern. Solid lines show performance with noiseless input data, and dashed lines show performance with noisy input data. 59
3.10 Sample learning curves for different verbs, showing cumulative accuracy over time. 60
3.11 Comparing the first eight uses of *fall* for Adam with the first eight sentences generated by our model. 68
3.12 The lexical entry for *dance* in the input-generation lexicon. 70
3.13 The lexical entry for *put* in the input-generation lexicon 71

4.1 Semantic properties for *cake* from WordNet 92
4.2 An input pair and its corresponding frame. 93
4.3 Semantic profiles of argument positions *Agent* and *Theme*. 99
4.4 Semantic profiles for the semantic roles *Destination* and *State*. 100
4.5 Semantic profile of the subject position in an intransitive construction. . 101

4.6 The predicted semantic profile for the subject position of an intransitive usage. The left profile is predicted for a usage with semantic primitives similar to verb *go*, whereas the right profile is for a usage with semantic primitives similar to verb *fall*. 103
4.7 The evolution of the transitive object role. 105
4.8 Learning curves for semantic profiles. The x-axis is time (#inputs), and the y-axis is divergence from the profile that the model eventually converges to. 106
4.9 Semantic profiles for Subject position of a transitive usage for a novel verb *gorp*, and an intransitive verb *dance* 111
4.10 Semantic profiles for Object position of a transitive usage for a novel verb *gorp*, and an intransitive verb *dance* 112

5.1 A verb usage and its corresponding frame. 123
5.2 Semantic properties for *dinner* from WordNet 124
5.3 Semantic profiles of *write* and *sing* for the direct object position. 130
5.4 Semantic profiles of *want* and *put* for the direct object position. 131
5.5 Semantic profile of a novel verb for the direct object position. 132
5.6 Compatibility scores for plausible vs. implausible verb-noun pairs. The correctly detected plausible pairs are shown in boldface. 135
5.7 Similarity with the base profile for Alternating and Non-alternating verbs. 137

List of Tables

3.1 Top 20 frequent verbs for three children in Browns corpus, CHILDES database 45

3.2 Predicate terms in the lexicon. 46

3.3 The four test input pairs used in the experiment of Section 3.6.3. 64

3.4 $\log P(k|F)$ values for matched and unmatched scene/utterance pairs. . . 64

3.5 Average probability of the predicted patterns. 66

3.6 Probabilities of semantic primitives associated with a typical and "unusual" usage of the verb *dance*. 73

4.1 log(**match_score**) for Recipient-first and Theme-first frames after processing 20 and 200 input pairs. The score of the best match (highest score) is shown in bold. 108

Chapter 1

Child Argument Structure Acquisition

1.1 Overview

The nature and amount of information needed for learning a natural language is the subject of much debate: is it possible to learn a language from usage data only, or some sort of innate knowledge is needed to boost the process? This is a topic of interest to (psycho)linguists who study human language acquisition, as well as computational linguists who develop the knowledge sources necessary for large-scale natural language processing systems. Children are a source of inspiration for any such study of language learnability. They learn language with ease, and their acquired knowledge of language is flexible and robust.

Much of the research on child language learning has been focused on the acquisition of verbs. Verbs are the main predicate terms in a language, and one of the most complex building blocks of both semantic representations and syntactic forms. Many linguistic theories have tried to explain learning the meaning of verbs, that is, the mapping between a word form and the appropriate elements of meaning. Even more extensively studied are

the rules that govern verb argument structure, or the relationship between the semantics of verbs and their syntactic form (i.e., the surface realization of the verb arguments in a sentence). The acquisition and use of form-meaning relations have caused one of the most controversial debates in the domain of language, with many researchers challenging the learnability of these associations solely from language input.

The focus of this dissertation is on verbs and their arguments. We propose a probabilistic model of the representation, acquisition and use of verb argument structure. The main goal of this model is to show that the association between verb syntax and semantics can be learned from verb usage data only, in a way that accounts for general verb regularities as well as the idiosyncratic properties of verbs. Moreover, we show through computational simulation that our proposed model accounts for the general patterns observed in child experimental data.

1.2 Challenges of Early Verb Learning

In order to achieve mastery of their native language, children must acquire knowledge of the regularities that govern the usage of various linguistic forms, as well as the exceptions to them. Such patterns are ubiquitous across all areas of language. In morphology, for example, there are word-formation rules such as the regular past tense in English (*move*/*moved*), which has various exceptions (*go*/*went*, *break*/*broke*). In syntax, there are word-order rules such as the head-first noun-adjective order in French (*fille intelligente*), also with numerous lexical exceptions (*belle fille*). It continues to be a subject of much debate how children manage to discover and use the general patterns, while keeping track of the exceptions—especially in the absence of explicit instruction, and in the face of incomplete and noisy input.

Verb usage provides a rich testbed for exploring such issues—specifically, what syntactic constructions a verb can occur in. Appropriate usage of verbs is guided by the

knowledge of verb argument structure: the number and type of the arguments that a verb takes (e.g., *The man built the house* but not *The man built* or *The house built the man*), the semantic roles[1] that the arguments receive in the event described by the verb (e.g., *She*$_{\langle Agent \rangle}$ *broke the window*$_{\langle Theme \rangle}$ and *The window*$_{\langle Theme \rangle}$ *broke*, but not *She*$_{\langle Agent \rangle}$ *broke*), and the syntactic realization of the verb and its arguments in a sentence (e.g., *I filled the glass with water* but not *I filled water into the glass*). Verbs pose particular challenges to children in this regard, due to the complexity of their possible usages and the interacting semantic and syntactic factors that determine both the general patterns and the exceptions. Children seem to possess some knowledge about the general regularities in the relationship between semantic roles such as Agent and Theme and syntactic functions such as Subject and Direct Object as early as age three (MacWhinney, 1995; Demuth et al., 2002). They use this knowledge to produce utterances they have never heard before, and to generalize the behaviour of verbs they have already learned to new ones. This ability sometimes leads to overgeneralization, in which a verb is used in a frequent construction that is not applicable to that particular verb, as in **I said her no* or **you can drink me the milk* (Bowerman, 1982, 1996). These kinds of errors gradually cease as children get older, and by teenage years they acquire almost adult-like linguistic competence (Demuth et al., 2002; Bowerman, 1996).

Experimental data on language learning demonstrates consistent patterns among children: for a given syntactic construction, there are few overgeneralization errors made at the very early steps of learning; the number of errors increases considerably as the general pattern starts to emerge, but after a while errors decrease again (Marcus et al., 1992). Studies on children's use of past-tense verbs in English (Marcus et al., 1992) and verb argument structure (Bowerman, 1982, 1990) both show the same pattern: overall, for all the constructions in a language, overgeneralization errors are relatively rare and occur

[1]Such elements are also termed participant, thematic, or case roles, and more or less fine-grained semantic distinctions are attributed to them. We use the widely accepted labels such as Agent and Theme for ease of exposition.

at a roughly constant low rate from the age of two into the school-age years. The over-generalization pattern can be considered an important clue to the internal mechanisms of language learning. As for whether corrective feedback from caretakers, or *negative evidence* (i.e., information about which strings of words do not belong to a language) plays a role in unlearning overgeneral rules, there is not a consensus among researchers. Some studies suggest that there is no reliable negative evidence available to children, and even if such feedback is provided, children ignore it (McNeill, 1966; Brown and Hanlon, 1970; Marcus, 1993). Others suggest that substantial amount of negative evidence can be found in child-directed speech, and this data may play an important role in the learning process (Saxton, 2000; Chouinard and Clark, 2003). However, it is safe to assume that not every child is consistently corrected for every error that they make, and unlearning overgeneralization errors must be possible only by receiving and processing more input from the environment.

In order to give an explicit account of the different aspects of language learning, many computational models have been developed in recent years. One group of models has been proposed for learning the idiosyncratic meaning of single words (e.g., Siskind, 1996; Fleischman and Roy, 2005; Yu et al., 2005), a fundamental step in the acquisition of the relationship between the syntactic and semantic structures. Moreover, the learning of the word-to-meaning mapping is most probably interleaved with the acquisition of verb argument structure and syntactic constructions. However, all through this dissertation, we assume that the non-trivial task of learning the correct mapping between a word and the semantic elements that represent its idiosyncratic meaning has been accomplished. We instead focus on the acquisition of the relation between the syntactic form, and the semantics of a verb and its arguments.

1.3 Acquisition of Verb Argument Structure

Learning a language has been suggested to rely on rich innate knowledge of verb regularities. The nativist view of language acquisition, advanced by Chomsky (1965, 1980), assumes that human beings are equipped with an innate set of syntactic categories and phrase structure rules that should be adjusted to the language to be learned. The primary argument for the innateness hypothesis, *Argument from the Poverty of Stimulus* (Chomsky, 1980), states that the input available to the language learner is too impoverished to account for the richness and complexity of natural languages. More specifically on argument structure acquisition, Pinker (1984, 1989) suggests that children have access to a set of semantic verb classes and a set of thematic linking rules that map the verb semantic roles to sentence structure and constrain the syntactic properties of verbs—a hypothesis known as *semantic bootstrapping*.

However, a number of usage-based proposals have argued that children learn language regularities from the input alone, without guidance in the form of innate principles (e.g., Bowerman, 1982; Akhtar, 1999; Tomasello, 2000a; Ninio, 2006). The usage-based theories of language acquisition are motivated by the experimental studies on language comprehension and generation in young children, which show that children build their linguistic knowledge around individual items (Akhtar, 1999; Tomasello, 2000a; Bowerman, 1982). Tomasello (2000a,b) suggests that young children initially learn verbs and their arguments as lexical constructions, on an item-by-item basis, and only later begin to generalize the patterns they have learned for one verb to another.

A family of usage-based theories that has grown in prominence in the study of verb argument structure acquisition is *construction grammar*. Construction grammars posit that, in addition to the idiosyncratic meanings associated with individual words or morphemes, meaning may also be directly associated with syntactic forms (e.g., Lakoff, 1987; Fillmore et al., 1988; Langacker, 1999). In particular, an argument structure construction is a mapping between underlying verb-argument relations and the syntax used to express

them, as in (1) and (2) below, taken from Goldberg (1995).

(1) Subj V Obj Obj_2 $\Leftrightarrow$ X CAUSE Y RECEIVE Z Ex: *Pat faxed Bill the letter.*
(2) Subj V Oblique $\Leftrightarrow$ X MOVE Y Ex: *The fly buzzed into the room.*

Associations between argument configurations (such as the transitive) and semantic features (such as causation) capture important linguistic regularities, and appear to play a role in both child language acquisition (Gleitman, 1990; Naigles, 1990; Fisher, 2002) and adult sentence interpretation (Bencini and Goldberg, 2000). However, the details of the acquisition of these associations and the constraints that govern their use is not clearly depicted. Explicit models must be explored, both of the underlying mechanisms of learning these regularities, and of the use of the acquired knowledge. In particular, computational experiments must show the feasibility of learning general patterns of the language from individual input without built-in knowledge of argument structure regularities, and without relying on negative evidence.

With respect to the claims of insufficiency of input data for language acquisition, connectionist models have been popular, mainly because they reflect the emergence of abstract knowledge from exposure to a number of instances (Allen, 1997; Allen and Seidenberg, 1999; Jones et al., 2000; Desai, 2002). But they are usually simplistic and cannot be extended to deal with larger vocabulary and more complex grammar. Also, these models can use the acquired syntactic knowledge in limited comprehension tasks, and the study of the generalization patterns in their output is limited. Other models, such as those of Clark (2001b), Gobet et al. (2004), Solan et al. (2004) and Cartwright and Brent (1997), aim for learning syntactic knowledge from corpus data. These models ignore the role of semantics in learning syntax and the relationship between the two, and

most of them violate cognitive plausibility criteria such as incrementality. Others have explored the relationship between the external form (syntax) and the meaning (semantics) for the purpose of language acquisition. These models pursue a variety of goals, such as learning lexical constructions (Chang, 2004), learning verb meaning from image data (Dominey, 2003; Dominey and Inui, 2004), acquiring the syntactic and semantic features of verbs (Niyogi, 2002), and setting parameters of a universal grammar (Buttery, 2003, 2004). However, these models are limited in their ability to acquire knowledge of general verb argument structure patterns and to use that knowledge robustly when faced with new verbs or novel semantic combinations.

1.4 Semantic Roles and Selectional Preferences

One important aspect of linking the verb meaning to a valid syntactic form is the notion of verb semantic roles. Semantic roles, such as Agent, Theme, and Instrument, indicate the relations of the participants in an event to the main predicate. However, there is little agreement on what the nature of semantic roles is and how these roles are learned. Early theories suggested that roles are drawn from a pre-defined inventory of semantic symbols or relations (e.g., Chomsky, 1981; Bresnan, 2001), and that innate linking rules that map roles to sentence structure enable children to infer associations between role properties and syntactic positions (e.g., Pinker, 1989). However, numerous questions have been raised concerning the plausibility of innate linking rules for language acquisition (e.g., Fisher, 1996; Kako, 2006).

An alternative, usage-based view is that children acquire roles gradually from the input they receive, by generalizing over individually learned verb usages (e.g., Lieven et al., 1997; Tomasello, 2000a). In this view, there are no general labels such as Agent and Theme, but rather generalized conceptions of verb-specific roles such as 'kicker' and 'the thing being kicked.' Recent experimental evidence confirms that access to general

notions like Agent and Theme is age-dependent (Shayan and Gershkoff-Stowe, 2007). It remains unexplained, though, precisely how the transition from verb-specific roles to general semantic roles takes place. Moreover, experiments with children have revealed the use of verb-specific biases in argument interpretation (Nation et al., 2003), as well as of strong associations between general roles and syntactic positions (e.g., Fisher, 1996).

The usage-based view of verb semantic roles, which sees them as particular properties that each argument structure construction imposes on its arguments, is very similar to the notion of the selectional preferences associated with each verb. Verbs have preferences for the semantic properties of the arguments filling a particular role. For example, the verb *eat* expects that the object receiving its Theme will have the property of being edible, among others. Learning verb selectional preferences is also an important aspect of human language acquisition, and the acquired preferences have been shown to guide children's expectations about missing or upcoming arguments in language comprehension (Nation et al., 2003). However, it is not clear how the semantic properties that verbs demand from their arguments, or their selectional preferences, are represented and acquired from usage data.

In the context of learning verb semantic roles from input data, the appropriate computational models are lacking. Many computational systems model human learning of the assignment of general roles to sentence constituents, using a multi-feature representation of the semantic properties of arguments (e.g., McClelland and Kawamoto, 1986; Allen, 1997). However, these models require explicit labelling of the arguments that receive the same role in order to learn the association of the roles to semantic properties and/or syntactic positions. Automatic induction of verb selectional preferences from usage data has received more attention, especially in the context of computational linguistic applications. A variety of computational models have been proposed, most of which first initialize the concepts in a hierarchy with their frequency of use as the particular argument of a verb, and then find the appropriate level in the hierarchy for

capturing the verb's restrictions on that argument (e.g., Resnik, 1996; Clark and Weir, 2002). However, existing selectional preference models use the training data in "batch mode", since they use statistical measures that rely on total counts from a corpus. This property makes the existing models cognitively implausible, and it is not clear how the representation of selectional preferences could be updated incrementally in these models as the person receives more data.

1.5 Focus of the Study

Despite extensive research on child language acquisition, and the development of numerous computational models for simulating various aspects of language learning, a comprehensive model that uses a unified framework to explain different aspects of child verb learning is lacking. In this dissertation, we present a Bayesian model of the early stages of verb argument structure learning in children. We take a statistical usage-based approach, in which the model learns word-order regularities in the expression of arguments solely from examples of utterances paired with appropriate semantic descriptions. Generalization in our model is achieved through an unsupervised Bayesian algorithm that groups similar argument structure usages. The same Bayesian process underlies language use—comprehension and production—in our model, giving a unified mechanism for language acquisition and processing.

In Chapter 2, we propose a new view on the nature of constructions as probabilistic associations between syntactic forms and properties of meaning. Constructions emerge through a process of detection and clustering of similar verb usages. The probabilistic nature of argument structure constructions in the model enables it to capture both statistical effects in language learning, and adaptability in language use. Language use is formulated as a Bayesian prediction problem: in production, the best syntactic form is predicted to express an intended semantics, while in comprehension, the missing se-

mantic elements of a perceived utterance are predicted. Prediction is modeled as an interaction between verb-based and construction-based knowledge. While knowledge of established verb usages accounts for the evidence on children's verb-specific behaviour, the use of constructions leads to appropriate generalization of observed patterns to new or low-frequency situations.

In Chapter 3, we present an implemented version of our model for the early stages of argument structure acquisition in children. We describe the model more fully in the context of argument structure acquisition, present experiments that simulate studies on children, and discuss the results and their import for accounting for child data. We show that our system mimics children's behaviour in forming word-order generalizations, and in recovering from overgeneralizations without the use of negative evidence. Our experiments demonstrate that the model learns basic argument structure constructions such as the transitive and intransitive, and exhibits patterns of error and recovery in their use, similar to those observed in children.

In the version of the model presented in Chapter 3, we assume that the child can detect the correct semantic role for an argument upon watching a scene. As a result, the input to the model contains semantic roles as explicit labels such as Agent and Theme. We extend our model in Chapter 4 and focus on the learning and use of these roles from input alone. We represent each role as a *semantic profile* for an argument position in a general construction, where a profile is a probability distribution over a set of semantic properties that verb arguments can take. It is shown that, at the beginning, the semantic profiles reflect verb-specific properties of the role associated with an argument position. However, as the model is exposed to more input, the semantic profiles for each role demonstrate the general properties that are common among the arguments for the observed verb usages.

In Chapter 5, we propose further extensions to our model that enable it, in addition to learning general semantic roles for constructions, to use its verb-specific knowledge to predict intuitive selectional preferences for each verb argument position. We introduce

a new notion, a *verb semantic profile*, as a probability distribution over the semantic properties of an argument for each verb. A verb semantic profile is predicted from both the verb-based and the construction-based knowledge that the model has learned through clustering, and reflects the properties of the arguments that are observed for that verb. Our proposed prediction model makes appropriate generalizations over the observed properties, and captures expectations about previously unseen arguments.

Chapter 6 summarizes our contributions and presents possible directions for future research.

Chapter 2

A Bayesian Model of Early Verb Learning

2.1 Overview

We present a probabilistic model of the early stages of verb learning in children; in particular, we focus on the acquisition of verb argument structure. Our goal is to provide an answer to the following questions:

- How does general knowledge emerge from knowledge of specific verb usages, and what form does this knowledge of argument structure regularities take?
- How do general and specific knowledge interact over the course of acquisition, leading to the observed developmental stages in argument structure and use?

In response to the first theoretical question above, concerning the form of general knowledge and how it emerges, we propose a variation on the construction grammar approach. Our model induces probabilistic associations between syntactic and semantic features of a verb usage; these associations generalize over the syntactic patterns and the fine-grained semantics of both the verb and its arguments. The resulting associations have a similar

function to the constructions of Goldberg (1995), and indeed we refer to the associations formed by our model as constructions. However, our representation enables a new view on the nature of constructions: each construction is not simply a form/meaning pair, but rather a grouping of verb usages which share similar syntactic and semantic features, inducing a probability distribution over the values for these features. The probabilistic nature of argument structure constructions in the model enables it to capture both statistical effects in language learning, and adaptability in language use.

In addressing the second question above, a key property of our computational model is the interaction between the specific knowledge of observed verb usages, and the constructions that generalize over such observations. This interaction between verb-based and construction-based knowledge plays a crucial role in language use in the model, which is formulated as a Bayesian prediction problem. For example, in production, the model predicts the syntax to express an intended semantics, while in comprehension, it predicts (some of) the semantics of a perceived utterance. The emergent constructions in the model enable it to generalize observed patterns of association to new or low-frequency situations, while knowledge of specific verb usages is used when it is more established. This interaction leads to behaviour reminiscent of the general patterns of language use observed in children.

In this chapter, we present the model and its modules in detail. In Section 2.2 we look at the representation of the probabilistic constructions in our model, and explain the basic assumptions regarding the input to our model. We describe the acquisition of constructions in Section 2.3, and the formulation of language use as a prediction problem in Section 2.4. In Section 2.5, we provide a brief analysis of the expected behaviour of the model, as a result of the interaction between the factors which constitute the prediction model.

We will present adaptations of this model for various verb learning tasks in the following chapters: we study the acquisition of verb argument structure in Chapter 3, the

acquisition of verb semantic roles in Chapter 4, and the acquisition of verb selectional preferences in Chapter 5, all of which use the Bayesian model presented in this chapter as the core. For each variation of the model, we present and discuss computational experiments that demonstrate the match between the behaviour of the model and that of children in a particular aspect of verb learning.

2.2 The Representation of Probabilistic Constructions

The input to our model consists of natural language utterances, each paired with a logical representation of the scene that the utterance describes. By providing such input pairs to our model, we are assuming that the child can pick up the correct meaning of a sentence upon watching the corresponding event. Moreover, we assume that the model already knows the meaning of each individual word, i.e., the correct mapping between a word form and a meaning symbol.

We view the input pairs as a simulation of the data that children use in language learning. However, the model internally represents this data as argument structure frames, which it extracts from the input items. Therefore, all through this dissertation, we use the scene/utterance pairs as the original form of the input to our model, and the argument structure frames (whose feature structure might differ for different tasks) as the internal blocks of information, each corresponding to a usage. The argument structure constructions are represented as a group of such frames which share similar features.

The basic assumptions regarding the representation of meaning in our model are discussed in Section 2.2.1. The details of the input pairs and the extracted frames are explained in Section 2.2.2. In Section 2.2.3, the probabilistic representation of a construction in our model is described.

2.2.1 The Representation of Meaning

Following Jackendoff (1990), we assume that, upon observing a simple event, the child can construct a conceptual structure to represent that event. Moreover, if the child hears a sentence while watching the scene, she can establish a link between the linguistic description of the event and its conceptual representation. However, detecting the correct meaning of an utterance is not a trivial task. One well-known problem in language acquisition is that of *referential indeterminacy* (also termed *referential uncertainty*), in which the child may perceive many aspects of the scene that are unrelated to the perceived utterance (e.g., Quine, 1960; Gleitman, 1990). For example, when the speaker utters *Mom put toys in boxes*, the child may also form a mental representation of the description of various toys in the scene, of the events of Mom picking up and moving toys, and even of unrelated entities and events, such as a kitten playing with a string. A number of computational models have dealt with the problem of picking the right meaning for an utterance, as well as learning the meaning of individual words (e.g., as in Siskind, 1996; Fleischman and Roy, 2005; Yu et al., 2005). In our work, we assume that the (non-trivial) task of picking out the appropriate semantics for the utterance from the full representation of the perceived aspects of the scene has been performed.

We also assume that for each word, the child knows the corresponding meaning (i.e., its semantic symbol). Learning the correct meaning for both verbs and nouns has been suggested to be based on cross-situational observation (Pinker, 1989; Fisher et al., 1994): the meaning of a word is consistent across multiple usages, and can be learned by detecting the set of meaning elements that are common across all situations in which a word occurs. It has been argued that, for certain verbs or situations, using cross-situational learning is not enough. For example, Gleitman (1990) argues that paired verbs such as *chase* and *flee* can always be used to describe the same event. Therefore, just by watching a chasing scene, it is impossible for the child to decide whether the verb describing the scene means "chase" or "flee". However, for the purpose of learning the simple verbs

discussed in this dissertation, we make the simplifying assumption that the child can learn the idiosyncratic meaning of simple verbs (e.g., to *dance* means to move in a rhythmic manner) before learning their full argument structure, although the event-specific meaning of the verb (e.g., that causation is part of the meaning of *break* in *the girl broke the mirror*, but not in *the mirror broke*), which we call the *semantic primitives* of the verb, are learned as part of the process we model.

We recognize that resolving referential indeterminacy and learning word-to-meaning mappings are processes that are likely interleaved with argument structure acquisition. Future work will need to integrate these steps into the model, rather than assuming that the input to the model is the output of a prior processing step which extracts these aspects of the scene.

2.2.2 Input Pairs and Extracted Frames

An individual verb usage consists of a scene/utterance pair, which we also refer to as an input pair, since a sequence of such pairs comprises the input to the learner. The scene component of an input pair is a logical-form representation of the aspect of the observed scene that the utterance (the perceived linguistic datum) describes. For example, the top panel of Fig. 2.1 shows a sample input pair. In the scene representation, each predicate (shown in bold) is indexed by a (possibly empty) set of semantic primitives known to describe the action, state, or relation expressed by the predicate. Each entity (shown in all upper case) is indexed by the participant role which the entity plays with respect to the predicate. The corresponding utterance indicates the sequence of words used to express the semantics of the scene. For now, our utterance representation omits such aspects as tense marking on the verb and grammatical function words such as determiners. (Note however that prepositions are included as predicates, since they play an important role in argument structure.)

From each scene/utterance pair, our system extracts an argument structure frame.

Scene/Utterance Input Pair		
$\mathbf{Put}_{[cause,move]}(\mathsf{MOM}_{\langle agent\rangle}, \mathsf{TOYS}_{\langle theme\rangle}, \mathbf{In}_{[]}(\mathsf{BOXES}_{\langle destination\rangle})_{\langle destination\rangle})$ *Mom put toys in boxes.*		
Extracted Verb Argument Structure Frame		
head verb		*put*
verb semantic primitives		⟨*cause, move*⟩
number of arguments		3
arguments	roles	⟨*agent, theme, destination*⟩
	categories[a]	⟨*human, concrete, dest-pred*⟩
syntactic pattern		arg1 *verb* arg2 arg3

[a]Extracted from a representation of the child's ontology.

Figure 2.1: An input pair and extracted frame for a verb argument structure.

The bottom panel of Fig. 2.1 shows a sample frame corresponding to the main predicate, **Put**, of the input pair in the top panel of the figure. Since prepositions are predicates, an argument structure frame is also extracted for each use of a preposition in the utterance. Each frame records a number of syntactic and semantic features that represent an input pair. We will explain in detail the frame structure, as well as the type and range of the frame features, for each of the verb learning tasks in Chapters 3 to 5.

An extracted argument structure frame that has been previously unobserved for its head word[1] (verb or preposition) is stored in the lexical entry of the head word, and its frequency count is initialized to one. If the frame has been previously observed for the head word, then its already-stored frequency count is incremented. An extracted frame that has been unobserved previously is input to our Bayesian learner which forms

[1]A new frame is considered 'unobserved' for a head word if there is no frame previously extracted for that head word which has the exact same feature values as the new frame for all the features. However, in the implemented version presented in Chapter 3, we describe a merging mechanism for similar frames based on the semantic categories of their corresponding arguments (see Section 3.4.2).

groupings of frames into constructions.

2.2.3 Constructions as Groups of Frames

A construction in our model is a group of argument structure frames that are "similar enough" to be grouped together, according to the probabilities over their features. The notion of "similar enough" is described in detail in the next section; here we focus on the general properties of constructions, and their representation in the lexicon.

Due to the formulation of our learning model, frames with similar syntactic and semantic properties are usually grouped together in one construction. The values for some features, such as the syntactic pattern and the thematic role assignments, come from a limited set, due to the regular nature of natural languages. Therefore, there is often a dominant syntactic pattern and thematic role assignment among frames in each construction. However, due to the probabilistic nature of our model, a few "outliers" might be found in each construction as well. Features with more varied values (such as semantic categories of the arguments and semantic primitives of the verbs) are less uniform. However, it is often the case that frames in a construction share close values for such features (e.g., the semantic categories of the first argument in a group of frames that correspond to a transitive construction might be $\langle human \rangle$, $\langle animal \rangle$, $\langle living_thing \rangle$ and so on). Therefore, the primary property of constructions in our model is that they determine a probabilistic association between syntactic and semantic features. For example, usages such as *Jay got a tower* and *Kay made a tower* may yield frames that form a (transitive) construction. While the frames share the verb semantic primitive *act*, they differ in others (*possess* for the former, and *become* for the latter). If this observation holds across a number of usages that exhibit this form, then we would find a higher probability for the primitive *act* given this construction than for the other semantic primitives. In this way, constructions probabilistically generalize the properties of a set of frames.

Each verb with a frame that participates in a construction has a link to that con-

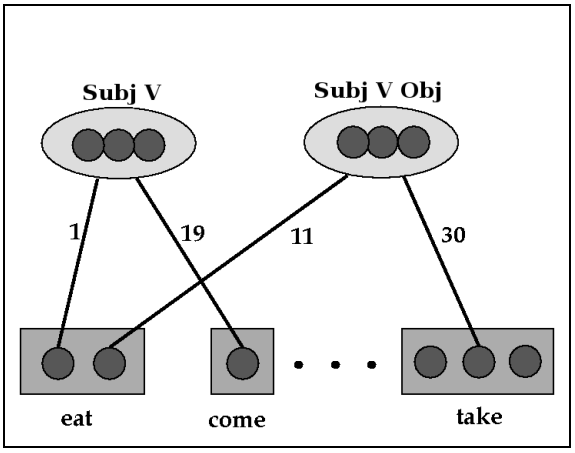


Figure 2.2: A portion of the lexicon showing an intransitive construction, with a dominant syntactic pattern Subj(ect) V(erb), and a transitive construction, with a dominant syntactic pattern Subj(ect) V(erb) Obj(ect). Numbers on links represent the observed frequency of the verb in a frame that participates in that construction. For example, *eat* has been seen once in an intransitive construction and 11 times in a transitive construction.

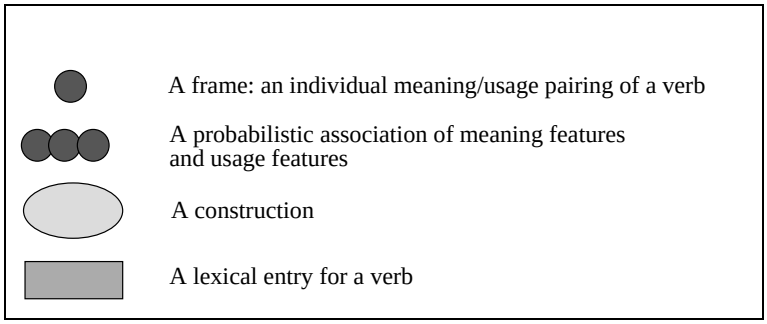


Figure 2.3: The legend for symbols representing verbs and constructions in the diagrams.

struction in the lexicon. The links are weighted by the frequency with which the verb has been seen in the participating frame, capturing the statistical usage pattern of the verb in various constructions. From the point of view of a construction, these frequencies are used in the calculation of the probabilities over the features occurring in its observed frames. Moreover, the sum of its incoming link frequencies contributes to the overall probability of a construction. Fig. 2.2 illustrates a simplified portion of the acquired

lexicon, showing the lexical entries of verbs containing frames, and their links to constructions.[2] (See Fig. 2.3 for the legend of symbols used in all depictions of the lexicon throughout the dissertation.) In the next section, we describe how the link between a frame and a construction is established by our Bayesian learning algorithm.

2.3 Acquisition of Constructions

2.3.1 Overview of the Acquisition Process

The argument structure acquisition mechanism that we propose incrementally learns from the processing of each scene/utterance pair that is input to the model. As noted above, the system first extracts an argument structure frame from the current input; this frame is then presented to an unsupervised Bayesian clustering process. This process groups the new frame together with an existing group of frames—a construction—that probabilistically has the most similar properties to the new frame. The Bayesian approach we use is an adaptation of a model of human categorization proposed by Anderson (1991), which incrementally groups perceived items (in our case, frames) into categories of items with similar features (in our case, constructions).[3] It is important to note that the categories (i.e., constructions) are not predefined, but rather are determined by the similarity patterns over observed frames. A new construction is created in response to a frame when none of the existing constructions are sufficiently similar to it—that is, when no existing construction has sufficiently high probability given the properties of the new frame. An overview of the process is shown in Fig. 2.4.

Grouping a new frame F with other frames participating in construction k is formalized as finding the k with the maximum probability given F:

[2] Appendices A and B represent snapshots of the lexicon and the acquired constructions for implemented versions of the model presented in Chapters 3 and 4, respectively.

[3] Anderson's (1991) model is shown to be directly equivalent to the *Dirichlet process mixture model* (Antoniak, 1974; Neal, 2000) if the parameters of the two models are properly set.

$$\textbf{BestConstruction}(F) = \underset{k}{\text{argmax}}\ P(k|F) \tag{2.1}$$

where k ranges over the indices of all constructions, including an index of 0 to represent a new construction. Using Bayes' rule, and dropping $P(F)$ which is constant for all k:

$$P(k|F) = \frac{P(k)P(F|k)}{P(F)} \ \propto\ P(k)P(F|k) \tag{2.2}$$

The prior probability of a construction, $P(k)$, is proportional to the total frequency of frames associated with k. The conditional probability of the new frame given a construction, $P(F|k)$, is determined by the proportion of both syntactic and semantic features it shares with the frames in the construction. Because this probability may be calculated over partial information, learning can proceed even in the face of an incomplete frame (i.e., when some of the frame features are missing).

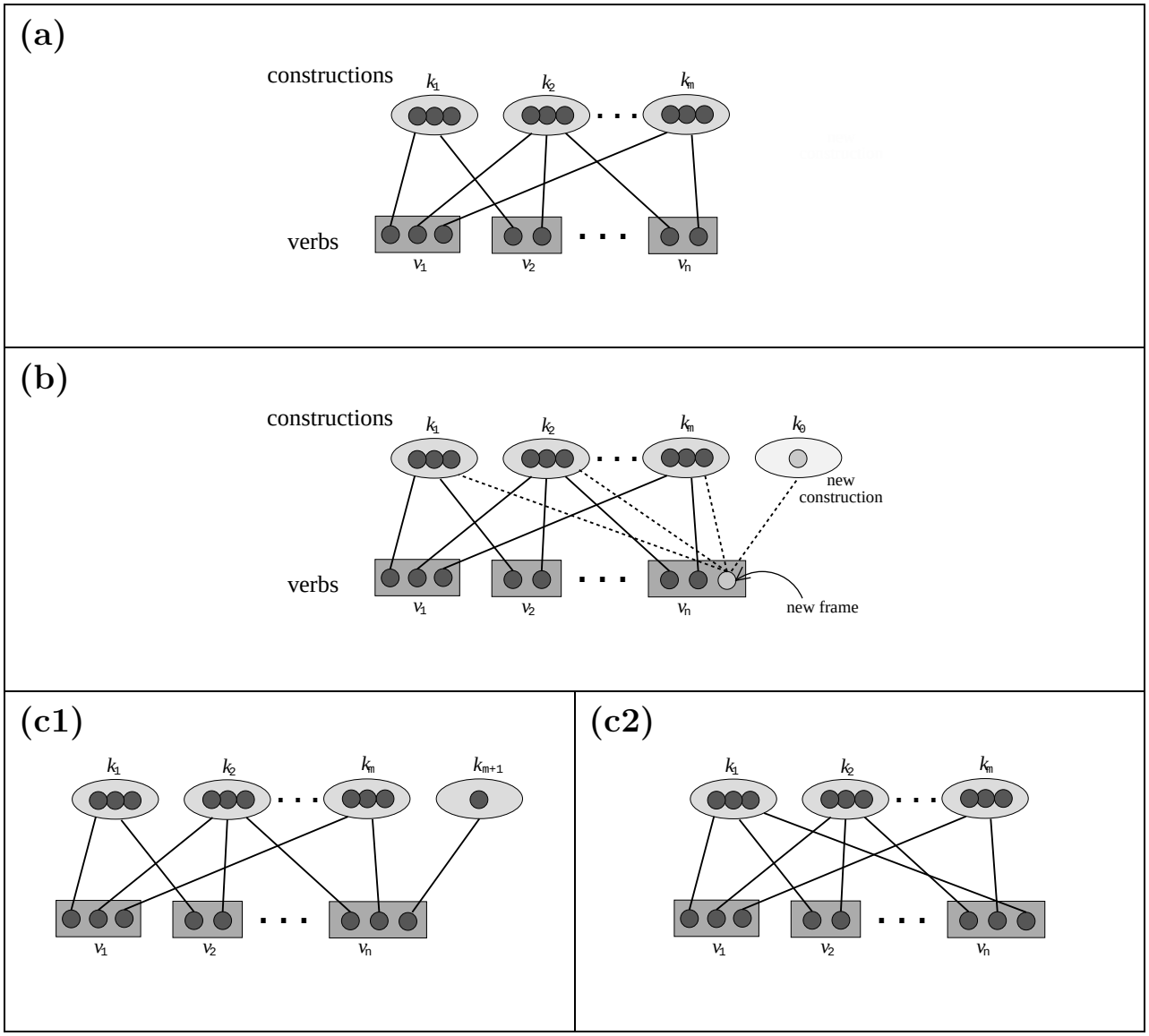


Figure 2.4: Processing a new frame: (a) A snapshot of the lexicon before the new frame is processed. (b) Addition of the new frame to the lexical entry for its main verb, in this case v_n. All existing constructions and a potential new construction are considered as a match for the new frame, as indicated by the dashed lines. Two alternatives are possible, as shown in panels c1 and c2. (c1) The new construction is the best match, hence it is added to the lexicon (and becomes k_{m+1}), and the new frame is linked to it; OR (c2) One of the existing constructions (in this case k_1) is the best match, hence the new frame is linked to it, and no new construction is created.

2.3.2 Details of the Probability Model for Learning

The following sections describe in detail the calculation in equation (2.2) of the prior and conditional probabilities, respectively, used in learning.

Prior probabilities

The higher the prior probability of a construction, the more likely for a new usage of a verb to appear in it. In a usage-based framework, the probability of a construction is assumed to change over time, as the child observes additional evidence of various constructions. In particular, as a construction is used more, it becomes more entrenched and is more likely to be used in the future. To model this intuition, we calculate the prior probability of a construction based on the frequency of its frames in the observed input, rather than assigning some pre-set prior distribution across constructions. In this way, the prior probability of each construction varies over the course of learning, reflecting how entrenched that construction has become.

The prior probability of construction k, for $k > 0$, is given by:

$$P(k) = \frac{n_k}{n+1} \tag{2.3}$$

where n is the total number of observed frames, and n_k is the number of frames participating in construction k. The normalizing factor is set to $n+1$ rather than n, to accommodate the probability mass of a potential new construction ($k = 0$), whose prior probability is given by:[4]

$$P(0) = \frac{1}{n+1} \tag{2.4}$$

[4]The formulas used in equation (2.3) and equation (2.4) are equivalent to those used by Anderson (1991), $P(k) = \frac{cn_k}{(1-c)+cn}$ and $P(0) = \frac{(1-c)}{(1-c)+cn}$, with his "coupling probability" c set to the mid value of .5. This value determines the degree to which items "prefer" to be grouped together. The mid value makes the fewest assumptions, by avoiding a strong inclination to either prefer or disprefer grouping an item into an existing category.

Thus, the prior probability for an existing construction is proportional to the frequency of its frames (equation (2.3)), following the intuition that it is more probable for a newly observed frame to come from a more entrenched construction. Moreover, the prior probability of a new construction is inversely proportional to the number of observed frames overall (equation (2.4)), capturing the intuition that the more exposure the child has to the language, the less likely it is that a brand new construction will be observed.

Conditional probabilities

The conditional probability of a frame F is expressed in terms of the individual probabilities of its features. To make the calculation feasible, we assume that these features are independent; thus, the conditional probability of a frame F is the product of the conditional probabilities of its component features:

$$P(F|k) = \prod_{i \in FrameFeatures} P_i(j|k) \tag{2.5}$$

where j is the value of the i^{th} feature of F, and $P_i(j|k)$ is the probability of displaying value j on feature i within construction k. This probability is estimated using a smoothed maximum likelihood formulation, reflecting the emphasis on usage statistics in child language acquisition. The conditional probability estimate for each of the features conforms to the following general pattern (Lidstone's law, Manning and Schütze, 2001):[5]

$$P_i(j|k) = \frac{\mathbf{count}_i^k(j) + \lambda}{n_k + \lambda\alpha_i} \tag{2.6}$$

where n_k is the number of frames participating in construction k, for $k \geq 0$ (note that $n_0 = 0$), and $\mathbf{count}_i^k(j)$ is the number of those with value j for feature i. The smoothing parameter α_i is set to (an estimate of) the number of possible values that feature i can

[5]Smoothing techniques, introduced in Good (1953), have been used extensively to enable statistical models of language to deal effectively with sparse data (Pereira, 2000). We make no claims regarding the psychological validity of this particular means for smoothing.

take on. Note that, for a new construction, because $\mathbf{count}_i^0(j)$ and n_0 both have the value 0, the estimated conditional probability is $1/\alpha_i$—that is, for a new construction, all feature values are equally likely. Parameter λ in equation (2.6) is set to a small constant so that the constructions which have no members with value j for feature i have a low (but non-zero) probability.[6]

Some specific properties are worth noting for the case in which k is a well-entrenched construction—that is, when n_k is large. In this case, the conditional probability of a particular feature value is close to the relative frequency of its observed occurrence in k, $\mathbf{count}_i^k(j)/n_k$ (because λ in the formula is small). If $\mathbf{count}_i^k(j)$ is zero, the conditional probability will also approach zero. Thus, even though a large n_k yields a large prior probability for the construction k (see equation (2.3)), if the feature value of i for frame F being processed has not been seen in construction k, the effect of the large prior will be greatly reduced. If this is true for all existing constructions ($k \geq 1$), the model will likely add a new construction to accommodate the previously unseen feature value.

2.4 Language Use as Prediction

We formulate language use (production and comprehension) as a prediction process, in which missing features in a frame are set to the most probable values given the available features. Thus, we model sentence production as predicting the most probable syntactic pattern, given the semantic features of a scene representation. Comprehension is conversely the prediction of (some aspects of) the intended semantics from a given utterance. For example, to produce a sentence to express the following scene semantics:

$$\mathbf{Kick}_{[cause,contact]}(\mathsf{TIM}_{\langle agent\rangle}, \mathsf{BALL}_{\langle theme\rangle})$$

we ask the model to predict the "missing" syntactic pattern, based on the feature values

[6]The setting of these parameters in the implemented model is described in the experimental set-up sections of the following chapters (see Sections 3.5.3, 4.5.2 and 5.5.2).

for the head word (**Kick**), the number of arguments (2), the semantic roles of the arguments (⟨*agent,theme*⟩), the semantic categories of the arguments (⟨*human,concrete*⟩), and the semantic primitives of the verb ([*cause,contact*]), recorded in the extracted frame from the corresponding input.

This prediction process integrates both sources of acquired knowledge—the frames stored with each verb, along with their organization into constructions. If a usage of a verb is sufficiently complete and frequent, then it will have a strong influence on the prediction of unknown features. On the other hand, if relevant verb-based information is incomplete (or missing), constructions play an important role. Constructions are the source of generalization over observed knowledge: a prediction based on a construction generalizes over all its frames, and allows the model to extrapolate usages associated with some verbs to new usages of others. This enables the model to produce or understand a verb in a novel (for that verb) frame, as long as semantically similar verb usages have been observed.

The value of an unobserved feature is predicted based on the match between the given partial set of observed features and the learned constructions:

$$\begin{aligned} \textbf{BestValue}_i(F) &= \operatorname*{argmax}_j P(i,j|F) \\ &= \operatorname*{argmax}_j \sum_k P_i(j|k)P(k|F) \end{aligned} \tag{2.7}$$

Here, F is a partial frame, i is a feature missing from that frame, j ranges over the possible values of feature i, and k ranges over all constructions. Intuitively, the model generalizes over the items within a construction to predict the most probable value for the missing feature across those items ($P_i(j|k)$); this is then weighted by the probability of the construction given the partial frame ($P(k|F)$). The first component of the sum, $P_i(j|k)$, is calculated as in the learning phase of the model, using equation (2.6). To calculate the factor $P(k|F)$, we use Bayes' rule and drop the constant $P(F)$ term (cf. equation (2.2)

on page 21), to yield:

$$P(k|F) \propto P(k)P(F|k) \tag{2.8}$$

Here the prior probability, $P(k)$, and the conditional probability, $P(F|k)$, are simply determined as in our learning module (using equation (2.3) on page 23, and equation (2.5) on page 24, respectively).

2.5 The Expected Behaviour of the Model

Recall that one of the features in a frame is the head verb of the usage. Due to the formulation of the conditional probability, a construction which has no frames with the same head verb as the new frame F will have a very low conditional probability for the head word feature, $P_{verb}(j|k)$, compared to constructions which include that verb. This property reflects the influence of the observed usages of the verb, which gives a verb-based bias to our prediction model. On the other hand, the conditional probability of each construction is modulated by its prior probability, $P(k)$, which reflects the entrenchment of that construction. The outcome of the prediction model is mainly determined by these two important factors, which interact over the course of learning to lead to three phases of acquisition with respect to verb argument structure usage. First, we observe a period of imitation. Early on in learning, constructions generally are not highly entrenched, so the constructions that contain specific observed usages of the verb in the frame usually outweigh other constructions (even if they are not the best intuitive match for the new frame). Over time, however, constructions become more entrenched as more examples of them are observed. During this period, the knowledge of general constructions may be stronger than the knowledge of a certain usage of a particular verb. This leads to a second phase of learning, that of generalization. In some cases, this is valid generalization of the argument structure properties of the language, while in others,

this leads to overgeneralization—incorrect usage of a verb according to a general pattern of the language that does not apply to that verb. With additional exposure to the language, however, the model recovers from such errors. As the frequency of the verb in its allowable usages increases, the knowledge of these usages specific to the verb again overrides the weight of a construction unseen with that verb. However, it is important to note that such recovery from overgeneralization does not prevent the model (like children) from making productive generalizations, a point we return to in our experimental results.

2.6 Summary

This chapter presents a general usage-based model for early verb learning. A Bayesian approach is proposed to capture the general knowledge of verb regularities from individual verb usages. Each verb usage is represented as a collection of syntactic and semantic features, and linguistic regularities are represented in the form of general constructions, which are a grouping of similar frames that share certain properties. From such a grouping of frames, the model induces a probabilistic association between syntactic and semantic features. The same Bayesian process underlies language use in our model, giving a unified mechanism for language acquisition and processing: a novel formulation of language use is proposed as a prediction problem, where the best value for a missing feature is predicted based on other available features, using both the general knowledge represented by constructions and the more specific knowledge acquired for each individual verb.

The following three chapters present the application of our model to learning three different aspects of verb argument structure constructions: the basic argument structure itself, semantic roles, and selectional preferences. In each chapter, the input to the model is explained, as well as the frame structure used for modeling the task in hand. Moreover, each chapter presents computational experiments that demonstrate the match between

the behaviour of the model and that of children in learning linguistic regularities and in using that knowledge in different language tasks.

Chapter 3

Acquisition of Verb Argument Structure

3.1 Overview

Verb argument structure details the semantic relations of a verb to its arguments—i.e., the participants in the event the verb describes—and how those arguments are mapped into valid syntactic expressions of the language. This complex aspect of language exhibits both general patterns across semantically similar verbs, as well as more idiosyncratic mappings of verbal arguments to syntactic forms. For example, as shown in (1)–(3), many verbs of movement allow both an intransitive form (as in the (a) examples) and a related transitive form in which the agent causing the movement is expressed as the subject (as in (1b,2b)). However, for the verb *fall*, the transitive form is not acceptable, as shown in (3b).

1a) The chain dropped into the box. **1b)** The girl dropped the chain.

2a) The ball rolled into the room. **2b)** The child rolled the ball.

3a) The book fell into the fire. **3b)** *The boy fell the book.

Thus, children must learn both the general argument structure patterns exemplified by the examples in (1) and (2), and the finer-grained semantic restrictions that govern the exceptions, as in (3). Given the number of verbs in the language, and the wide range of syntactic structures that can be used to express verbal arguments, the acquisition of such knowledge is quite complex, and numerous questions remain concerning how children attain an adult level of competence.

It appears that children are sensitive to such regularities from an early age, producing novel utterances that obey the mapping of arguments to syntactic positions in their language (Bowerman, 1982; Pinker, 1989; MacWhinney, 1995). Moreover, children may even *overgeneralize* the observed patterns, thereby producing incorrect forms such as *Don't you fall me down* (Bowerman, 1982). Eventually, such erroneous verb usages discontinue. In this way, most children exhibit a "U-shaped learning curve" seen in other areas of language learning (Marcus et al., 1992). A U-shaped learning curve indicates an early phase of correct imitation of observed forms, such as *Ball roll*, *Adam roll ball*, and *Ball fall* (the left tip of the U). This is followed by a period of generalization of observed patterns in which children may incorrectly apply the regularities they have perceived to inappropriate forms, such as *Adam fall ball*, lowering their proportion of correct forms (the bottom of the U). Finally, children recover from such overgeneralizations, leading once more to a higher rate of correct forms (the right tip of the U). As is generally the case in language learning (Marcus, 1993), explicit negative evidence in the form of corrective input may not always be available, and thus cannot be relied on in children's cessation of incorrect forms.

In this chapter, we present an implemented version of our computational model for the early stages of argument structure acquisition in children. The model learns word-order regularities in the expression of arguments solely given examples of utterances paired with appropriate semantic descriptions. Our system is shown to mimic children's behaviour in forming word-order generalizations, and in recovering from overgeneralizations without

the use of negative evidence.

Section 3.2 discusses the linguistic theories of argument structure acquisition, and lays out the major questions concerning the acquisition of argument structure in children. The existing computational models of verb learning are reviewed in Section 3.3. In Section 3.4 we put forward an adaptation of our Bayesian model for argument structure acquisition and use. We describe our experimental set-up in Section 3.5, and present the experimental results in Section 3.6. Section 3.7 presents a general discussion about the experimental results, and relates them to the important questions on the learnability of argument structure constructions, and the various stages of learning. Finally, we summarize our contributions in Section 3.8.

3.2 Theories of Argument Structure Acquisition

Various proposals have been put forward to account for the acquisition of verb argument structure. The nativist theories of language acquisition assume that children are equipped with innate knowledge of language (e.g., Chomsky, 1980; Pinker, 1989; Grimshaw, 1990). The extent of this knowledge, however, is not agreed upon among all the researchers in this group. The innateness theory, first proposed by Chomsky (1965), states that children do not have to learn or construct abstract syntactic structures, but rather they already possess them as part of their innate language faculty. The general principles and elements common to all human languages are encoded in human brain in the form of a *Universal Grammar* (UG). The principal argument for the innateness of UG is *the Argument from the Poverty of Stimulus* (APS), which states that there are principles of grammar that cannot be learned on the basis of positive input alone, however complete and grammatical that evidence is.

The Semantic Bootstrapping Hypothesis, proposed by Pinker (1984, 1989), is a nativist account specialized for verb argument structure acquisition. According to this

hypothesis, the semantic roles of a verb are positions in decompositional representations of verbs' meanings (e.g., Agent is the first argument of CAUSE, Patient is the second argument of CAUSE, and Theme is the first argument of GO and BE). Moreover, each semantic role is associated with its own linking rule (e.g., Theme is linked to Subject if that syntactic function has not already been assigned, otherwise to Object). Since this model specifies that the way a verb's arguments are linked follows directly from its semantic representation, differences in the way closely related verbs (or alternative constructions involving the same verb) map their arguments must reflect differences in their meaning. Thus, since a verb's semantic structure and the mapping of its arguments are in perfect correspondence, a child with innate linking rules can predict correct mappings once he knows what a verb means.

In the nativist view, learning a language is reduced to adjusting the abstract and innate rules and structures to the input data. However, a number of recent psycholinguistic experiments question this view. Experimental studies on language comprehension and generation in young children (up to age three) show that they build their linguistic knowledge around individual items, rather than adjusting some general constructions they already possess (e.g., Bowerman, 1982; MacWhinney, 1982; Akhtar, 1999; Tomasello, 2000a). Two-year old children show little tendency to apply syntactic structures they have already learned to new verbs, but rather conservatively use each verb in structures they have heard it in before. Also, young language learners quite readily substitute nominals for one another and begin to form an abstract category, "nominal", from very early in development, but are not productive with their language in some other ways. The tendency to generalize familiar constructions to new forms increases with the age of the subjects, showing that some general structures are learned over time.

With regard to language use, of particular interest has been how general and specific knowledge of verb argument structure interact to yield a U-shaped learning curve. The mechanisms that children use to recover from making overgeneralization errors has

especially been the subject of much debate (Pinker, 1984, 1989; Bowerman, 1988, 1996; MacWhinney, 2004). This problem is of special interest since, unlike additional positive evidence of correct usages, the availability of negative evidence (corrective feedback) cannot be relied on in the process of "unlearning" overgeneralized rules (Brown and Hanlon, 1970; Marcus, 1993). However, novel forms for a verb cannot be ruled out simply on the grounds that the form has not been observed before, since adult-level competence includes an ability to comprehend the meaning of unusual utterances, or to generate novel utterances for unusual situations. Accounting both for this aspect of language, known as *productive generalization*, and for recovery from overgeneralization, necessitates a detailed usage-based theory of the mechanisms at play in argument structure acquisition.

These findings have led to the development of a number of usage-based theories of argument structure acquisition, which argue that children learn such regularities from the input alone, without guidance in the form of innate principles (e.g., MacWhinney, 1987; Tomasello, 2003). One important such theory is Tomasello's (2003) *Verb Island Hypothesis*, which states that children have an early period in which each of their verbs forms its own "island" consisting of verb-specific constructions with open nominal slots. Children use cognitive and socio-cognitive processes such as imitation (reproducing the language adults produce for the same communicative function), analogy, and structure mapping (detecting both structural and functional similarities in utterances independent of the specific words involved) to gradually categorize the relational-syntactic structure of their various item-based constructions, and therefore become productive with their language in more adult-like ways.

Another group of theories called *Construction Grammar* proposes a different approach to the representation and acquisition of argument structure, by positing abstract constructions that pair a syntactic form with meaning, independently of the verbs that appear in them (e.g., Lakoff, 1987; Fillmore et al., 1988; Langacker, 1999). Goldberg (1995)

offers an account of how the meaning associated with argument structure constructions is acquired: argument structure patterns are initially acquired on a verb-by-verb basis, and constructions associated with the common syntactic patterns are learned through a process of categorization and generalization over the input. The generalization of constructional meaning is based largely on the meanings of highly frequent light verbs.[1] Bencini and Goldberg (2000) compare the role of the main verb and the argument structure construction in sentence meaning by asking a number of participants to sort sentences according to their meaning. The results suggest that adults probably see both verbs and constructions as relevant to determining meaning. An established form-meaning mapping may even impose an "unusual" meaning when a verb is used in a manner that is not typical for it. For example, in *the fly buzzed into the room*, using the verb *buzzed* in a construction with a path argument induces a semantics of movement as well as the standard sound emission sense for that verb (Goldberg, 1995). Unlike overgeneralization errors, the ability of humans to generate novel utterances for unusual situations, or to comprehend the meaning of the unusual utterances (i.e., productive generalization) does not fade over time by processing more input.

Usage-based theories in general, and those founded on argument structure constructions in particular, have recently received much attention in the language acquisition community. These proposals raise the question of whether or not a psychologically plausible usage-based learning strategy can be successfully exploited in a computational model.

3.3 Related Computational Models

Learning the meanings and appropriate usages of verbs is a challenge for children, as well as for computational approaches trying to model this aspect of human language

[1]Light verbs are highly frequent verbs with very general meanings, such as *go, do, make, give, put*. Goldberg (1995) shows that the meanings of some of the light verbs correspond closely to the meanings associated with argument structure constructions. For example, the intransitive syntactic pattern "Subj V Obl" paired with the meaning"X moves Y" corresponds to the meaning of the light verb *go*.

acquisition. As in learning nouns (the words for objects), the learner must determine the meaning primitives that are associated with the verb in the face of referential uncertainty. Siskind (1994, 1996) and Thompson and Mooney (1999) propose algorithms for learning the association between words and the meaning primitives in the face of ambiguity and referential uncertainty. Yu et al. (2005) use a probabilistic framework to learn the degree of association between each word and possible meaning elements. Fern et al. (2002) address the issue of acquiring the set of logical formulas that describe an event expressed by a verb, using a set of annotated videos of simple actions in a supervised learning setting. Fleischman and Roy (2005) instead focus on determining which aspect of an event is being described by an utterance, in forming the word-to-meaning mappings in an expectation-maximization framework. For example, their system determines whether an utterance is describing an action such as "get axe", or a subcomponent of that action, such as "find axe." None of these models deal with the syntactic structure of the input sentences, and the impact of syntax on learning the meaning of words. However, Yu et al. (2005) show that integrating feedback from a syntactic category learning module into the word learning model can improve the performance.

Verbs pose the added complexity of having to determine the participants in the event described by the verb, how the participants relate to the event and to each other, and how they can be expressed in a syntactic form. Computational models of verb acquisition have tackled various aspects of this difficult learning problem. A number of connectionist models have been proposed in recent years that attempt to learn syntactic structure from sequences of words. The most influential approach of this kind is due to Elman (1990, 1991), who trains a simple recurrent network to predict the next input word, for sentences generated by a small context-free grammar. The network learns an abstract representation of grammatical relations through learning lexical categories of the input words such as nouns and verbs (with further subcategorization of nouns as animate/inanimate, human/non-human, etc.), and the permitted combination of these

categories in a sentence. Elman (1991) extends the previous model by using stimuli in which there are underlying hierarchical and recursive relationships (e.g., subject nouns agree with their verbs; verbs have different argument structures; recursive relative clauses are permitted). Although Elman's model shows very well the possibility of categorizing and learning abstract structures from strings of words, the artificial language it learns is purely syntactic, whereas natural language learning is crucially an attempt to discover the relationship between meaning and linguistic form. In an attempt to fill the gap between the learning of syntax and semantics in connectionist models, McClelland and Kawamoto (1986) present a connectionist model for assigning roles to constituents of sentences. They train the model on two sets of input: the surface structure of the sentence, and the thematic roles assigned to each constituent. Their model can learn the linking between the syntactic constituents and the thematic roles that it has seen in the input, and can also generalize these linking patterns to unseen data. However, the learning process is fully supervised, and the only form of language use performed by this model is limited comprehension.

Here, we focus particularly on learning the argument structure of a verb—the roles assigned by a verb to its participants, and the word order patterns in which those roles are expressed. Our work thus follows that of other recent research which has paid close attention to the relation between expressed meaning and syntactic form, and the importance of the syntax-semantics mapping in verb learning. Niyogi's (2002) model, like ours, uses a Bayesian framework, and shows effects of both syntactic and semantic bootstrapping—i.e., the use of syntax to aid in inducing the semantics of a verb, and the use of semantics to narrow down possible syntactic forms in which a verb can be expressed. In contrast to our model, Niyogi's approach relies on extensive prior knowledge in the form of a Bayesian hypothesis space and the probabilities over it. Desai's (2002) work also demonstrates the role of syntactic and semantic bootstrapping, but does so using a very restricted set of semantic features. Moreover, the model shows only limited generalization abilities,

where a major focus of our work lies in showing how verb-specific knowledge can be generalized in a way that supports acquisition and exhibits child-like stages of behaviour. Allen (1997) proposes a connectionist system that focuses particularly on learning the mapping between roles and the entities that play those roles in an event. Due to the use of a distributed "microfeature" representation of the semantics, this model is able to make interesting generalizations over argument structure syntax and semantics. However, the acquired knowledge is only used for limited comprehension, and the potential of the model in performing more varied tasks of language use is not investigated, and a more explicit analysis of the formation of the argument structure constructions in the model is unavailable.

Other computational models are more similar to ours in relying on the form/meaning mapping that must be acquired for verbs, specifically looking at the acquisition of constructions. Chang (2004) presents a model for learning lexically-specific multi-word constructions from annotated child-directed transcript data. The goal of the model is to learn associations between form relations (typically word order) and meaning relations (typically participant role-filler bindings), and to use them in language comprehension. However, in contrast to our approach, this model can only acquire verb-specific constructions which cannot be generalized to other verbs. Another approach for learning grammatical constructions by pairing form and meaning is presented by Dominey (2003) and Dominey and Inui (2004), in a system that learns from video. While the model has some ability to generalize its knowledge to new verbs, its learning is highly dependent on the unrealistic assumption of having each syntactic form uniquely identify the associated meaning (i.e., forms and meanings are in a one-to-one mapping). Both of these models are thus limited in their ability to acquire knowledge of general verb argument structure patterns and to use that knowledge robustly when faced with new verbs or novel semantic combinations.

Many computational systems have instead addressed the problem of the acquisition

U(tterance):	*Mary put ball on table.*
S(cene):	$\mathbf{Put}_{[cause,move]}(\mathsf{MARY}_{<agent>}, \mathsf{BALL}_{<theme>}, \mathbf{On}_{[]}(\mathsf{TABLE}_{<destination>})_{<destination>})$

Figure 3.1: A sample input pair

of general grammatical knowledge (e.g., Jones et al., 2000; Clark, 2001a,b, 2006; Pereira, 2000; Gobet et al., 2004; Solan et al., 2004; Lappin and Shieber, 2007). These approaches extract general syntactic rules from sentential data, in contrast to our goal of learning form-meaning mappings, especially those that encapsulate verb argument structure. The latter relies crucially on the interaction between lexically-specific knowledge and general patterns over them. Interestingly, Onnis et al. (2002) study the problem of recovering from overgeneralization in grammar learning in the context of an artificial language. They show that a grammatical representation that incorporates both general rules and lexical exceptions provides a simpler encoding as the learner is exposed to larger amounts of data. While very different from our model in its intent, this work lends support to our assumption that it is important to balance verb-specific knowledge with that of general constructions for effective learning.

3.4 A Computational Model of Verb Argument Structure Acquisition

We use our Bayesian framework, presented in the previous chapter, for modeling verb argument structure acquisition in children. The basic assumptions and properties regarding the input to the model are explained in Section 3.4.1, and the frame structure and features used in this variation of the model are detailed in Section 3.4.2.

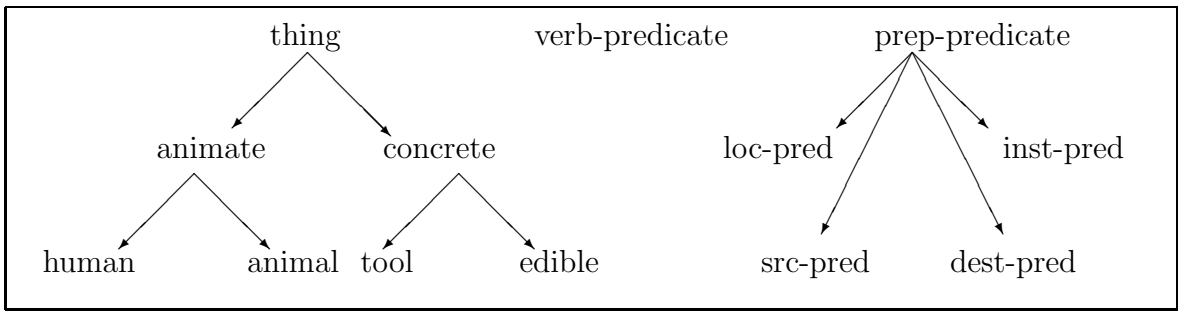


Figure 3.2: The semantic hierarchy

3.4.1 Basic Properties of the Input

We focus on learning the argument structure of a small group of verbs (and prepositions) whose semantic primitives are largely detectable by the child from the scene. As illustrated in Fig. 3.1, we use a simple logical form for representing the relevant semantics of an observed scene. Verbs and prepositions are represented as predicates (e.g., **Put**, **On**) that can take a number of arguments, each of which can be an entity (e.g., MARY, BALL, TABLE) or a predicate structure itself (e.g., **On**(TABLE)). As mentioned before in Section 2.2.1, we assume that the model already knows the correct mapping between each word and the corresponding meaning symbol.

We also assume that at the point of learning modeled by our system, the child is able to recognize linguistically relevant semantic properties such as the roles played by the arguments in the event. (For example, in an event described by *Mat gave Pat the hammer*, the child can detect Mat as the Agent, Pat as the Recipient, and the hammer as the Theme of the "giving" event.) Therefore, in the meaning representation of a scene, each argument is assigned a participant role in the event (e.g., *agent*, *theme*, *destination*). Also, each predicate is augmented by a set of semantic primitives which describes the characteristics of the event (e.g., [*cause*, *move*]). Currently, prepositions are associated with empty sets of primitives.

The representation of an utterance is simplified to include only words in a particular order, so that the model can extract a syntactic pattern that specifies the relative

ordering among a predicate and its arguments. In the current implementation, we do not address learning of morphology; all words appear only in their root forms. This simple formalism allows us to represent nested structures (when the argument of a predicate is another predicate). However, more sophisticated syntactic phenomena such as dependencies cannot be captured by this format.

The participant roles and semantic primitives used in the semantic representation are drawn from a finite, pre-determined set that is assumed to be known to the child at the stage of learning being modeled. The participant roles we use are a list of 10 commonly accepted thematic roles from the linguistic literature (e.g., Jackendoff, 1990): *agent*, *theme*, *destination*, *source*, *beneficiary*, *stimulus*, *state*, *experiencer*, *instrument*, and *co-agent*. We enumerate a set of 9 primitives for describing the coarse-level semantics of a verb in a usage: *act*, *cause*, *move*, *become*, *possess*, *change-of-state*, *perceive*, *contact*, and *manner*, drawing on linguistic proposals concerning fundamental event properties (e.g., Jackendoff, 1990; Rappaport Hovav and Levin, 1998). We augment these with features that capture the finer-grained meaning distinctions among our experimental predicates; e.g., *consume* for an eating event, *playfully* for a playing event, and *rest* for a sitting event. This allows us to compare the learning of general semantic properties of predicates to that of more specific ones.

In extracting the information for an argument structure frame from a scene representation, the model records the semantic category of all the arguments. These semantic categories are selected to reflect a simplified early ontology of a child, depicted in Fig. 3.2. (These categories may be used in merging highly similar argument structure frames, as will be described in Section 3.4.2.) We assume that at the stage of acquisition being modeled, these and other semantic properties can be largely determined by the child from the observed scene.

3.4.2 Argument Structure Frames

The model extracts an argument structure frame from each scene/utterance pair. Each frame records the head word (verb or preposition), the number of arguments, the semantic primitives that describe the predicate, the participant roles of the arguments with respect to the predicate, and the ontological category of each of the arguments. The frame also records the syntactic pattern of a predicate usage; currently, this syntactic pattern encodes the relative positions of the predicate word and each of its arguments. Since prepositions are predicates, an argument structure frame is also extracted for each use of a preposition in the utterance.

An extracted argument structure frame that has been previously unobserved for its head word (verb or preposition) is stored in the lexical entry of the head word, and its frequency count is initialized to one. If the frame has been previously observed for the head word, then its already stored frequency count is incremented. Highly similar frames may be merged into a single compatible frame in the lexical entry. For example, if two frames are identical except that the semantic categories of the arguments of one frame are more general than those of the other, the more specific frame is replaced with the more general one. An extracted frame that has been unobserved previously is input to our Bayesian learner which forms groupings of frames into constructions.

A property of our model is that, in the long run, it does not remember the verb usages that it has seen in the input. The model extracts those features mentioned above from each input pair in the form of an argument structure frame, and then the input pair is discarded. Among the extracted features are the semantic categories of the arguments that participate in a scene, but the actual words that represent those arguments are not kept. In a way, our model is "memoryless", i.e., immediately after extracting a frame from an input pair, it forgets the words that appear in each argument slot. While this enables the model to make appropriate generalizations, it limits the ability of the model to learn more fine-grained semantic properties of the arguments for each verb or construction, as

well as collocations (or word co-occurrences).

3.5 Experimental Set-Up

In the rest of this chapter, we report a number of computational experiments that demonstrate the feasibility of learning general constructions. We show that the behaviour of the model over the timecourse of acquisition mimics in relevant aspects the stages of learning exhibited by children.

Large-scale corpora of utterances paired with meaning representations, such as required by our model, do not currently exist, so we artificially generate such input based on the properties of existing child data. Artificially generating the input data enables us to simplify the syntactic and semantic properties of the input, in order to concentrate on those aspects of verb argument structure acquisition that are the focus of this chapter. Moreover, using automatically generated input allows us to test the model through a number of simulations, and evaluate it across minor variations in corpora.

To support our experimental simulations, we create an input-generation lexicon which is used to randomly generate corpora on which to test our model; this lexicon is described in Section 3.5.1. In Section 3.5.2, we describe how the input-generation lexicon is used to create the input corpora for each simulation in our set of experiments. Finally, in Section 3.5.3, we explain how the parameters of the model are set.

3.5.1 The Input-generation Lexicon

In order to provide the model with input data whose distributional characteristics are similar to those of the data children receive from their parents, we analyzed a portion of the CHILDES database (MacWhinney, 1995) as follows. We extracted from CHILDES all tenses of the 20 most frequent verbs in mother's speech to each of Adam (2;3–4;10), Eve (1;6–2;3), and Sarah (2;3–5;1). Table 3.1 shows the list of verbs and their frequencies for

each child. We selected 13 verbs from those in common across these three lists and added them to the input-generation lexicon. Each verb has associated with it in the lexicon the total frequency of its root form among the three children, and a unique symbol to represent its semantic predicate (e.g., the predicate symbol **Put** for the verb *put*). We also assigned each verb a set of possible argument structure frames and associated frequencies, which are manually compiled by examination of all uses of a verb in all conversations of the same three children. Prepositions used in these conversations were also added to the lexicon. The selected words are shown in Table 3.2, and a sample lexical entry is shown in Fig. 3.3.

For each semantic category from Fig. 3.2 that appeared as an argument in the manually compiled argument structure frames, a number of nominal terms were arbitrarily selected and added to the input-generation lexicon. For example, to satisfy the semantic category ⟨*edible*⟩ that appears in the argument structure frames for *eat*, we added the nouns *apple* and *cookie* to the lexicon. Since the focus of the model is on learning the argument structure of predicate terms, we do not expect the frequency of nouns to affect its behaviour. Therefore, instead of collecting the relative frequencies of the selected nouns from CHILDES, we assigned equal frequency to all the nouns in the lexicon (arbitrarily set to the value 1).

It is important to note that the input-generation lexicon is not used by the model in language learning and use, but only in producing the experimental corpora, as described next. The only lexical knowledge built into the model is the mapping between a word (e.g., *put*) and a semantic symbol (e.g., **Put**).

Adam		**Eve**		**Sarah**	
do	699	put	309	do	901
put	534	do	290	get	686
go	332	get	221	go	681
get	280	go	191	put	520
look	247	eat	128	come	404
take	212	take	117	take	296
play	191	sit	90	look	278
come	140	make	89	make	218
write	117	read	86	play	189
give	114	come	75	give	175
make	106	look	65	sit	144
sit	81	play	64	eat	127
eat	70	drink	61	write	78
find	66	give	59	find	67
read	59	write	50	fix	56
hold	55	find	43	drink	55
fall	46	hold	41	fall	50
draw	37	fix	41	hold	47
drink	24	draw	13	dance	37
fit	19	fall	7	read	29

Table 3.1: Top 20 frequent verbs for three children in Browns corpus, CHILDES database

Verbs	*go, put, get, make, look, take, play, come, eat, fall, sit, see, give*
Prepositions	*to, in, on, onto, under, from, for, at, with*

Table 3.2: Predicate terms in the lexicon.

3.5.2 The Input Corpora

For each simulation in our set of experiments, an input corpus of scene/utterance pairs is automatically created from the input-generation lexicon. The input pairs are randomly generated, using the frequencies in the input-generation lexicon to determine the probabilities of selecting a particular verb and argument structure for each input. Arguments of verbs are also probabilistically selected, constrained to conform to the indicated semantic category of the argument. Arguments which are predicates (such as prepositional phrases) are constructed recursively. The syntactic pattern associated with the selected predicate is used to produce the utterance that accompanies the created scene representation.

We assume that, at the point of learning modeled by our system, the child is able to recognize the syntactic pattern of the utterance and the semantic properties of the event (see Section 2.2.1). However, in reality, the input to children is often noisy or incomplete. For example, the child might mishear the utterance, or might not be able to extract the correct syntactic pattern from it. Similarly for the semantic information, the child might not be able to recognize the correct roles that the participants receive in a particular event, or the semantic categories that these participants belong to, or the semantic properties of the event itself. More problematic for the process of learning, children might sometimes misinterpret the perceived scene or utterance by filling the gaps based on their own (imperfect) linguistic knowledge, which leads to noisy data. We simulate two types of noise in our input corpora: incomplete data (where some pieces of information are missing) and misinterpreted data (where the child replaces the missing

WORD:	come
SYMBOL:	COME
CATEGORY:	[predicate]
FREQUENCY:	98
FRAME:	COME($1)
ARG-NO:	1
SPRIMITIVES:	act,move
ARGS:	$1 is [agent]
CONDITION:	$1 is [animate]
PATTERN:	$1 [word]
CONST-ID:	1
FREQUENCY:	35
FRAME:	COME($1,$2)
ARG-NO:	2
SPRIMITIVES:	act,move
ARGS:	$1 is [agent]
ARGS:	$2 is [destination]
CONDITION:	$1 is [animate]
CONDITION:	$2 is [dest-pred]
PATTERN:	$1 [word] $2
CONST-ID:	5
FREQUENCY:	58
FRAME:	COME($1,$2,$3)
ARG-NO:	3
SPRIMITIVES:	move
ARGS:	$1 is [agent]
ARGS:	$2 is [theme]
ARGS:	$3 is [destination]
CONDITION:	$1 is [animate]
CONDITION:	$2 is [concrete]
CONDITION:	$3 is [loc-pred]
PATTERN:	$2 [word] $3
CONST-ID:	8
FREQUENCY:	5

Figure 3.3: A sample lexeme

data with her own inference). Other types of noise, such as ungrammatical or incomplete sentences, are not currently modeled in the input.

During the input-generation process, two generated input pairs out of every five have one of their features randomly removed. The removed feature can be one of the following: the syntactic pattern, the semantic category of one argument, the participant role of one argument, and all the semantic primitives of the main predicate. One of these modified input pairs is used to simulate incomplete data. This modified input pair is left as is—i.e., with one feature missing—in the generated corpus. The other modified input pair is used to simulate noise. During a simulation, the missing feature of this input pair is replaced with the most probable value predicted for it at that point in learning; the completed input pair is then used in the learning process. This corresponds to a child using her own inferred knowledge to fill in information missing from an observed scene/utterance pair. The resulting input pair is noisy, especially in the initial stages of learning.[2]

To investigate the effect of this partial and noisy data on the performance of the model, we run a suite of experiments both on a set of original generated input corpora (the "original" corpora), and on their corresponding versions modified as described just above (the "noisified" corpora). With this exception (which is noted explicitly in the results below), all experiments are over noisified corpora only.

3.5.3 Setting the Parameters of the Model

For the purpose of implementing the model and running simulations, we need to assign values to parameter α_i for each feature i in equation (2.6). In our model, each frame has six different features: the head word, the semantic primitives of the head word, number of arguments, participant roles of the arguments, semantic categories of the arguments,

[2]In predicting a syntactic pattern for an incomplete frame, it is possible that the pattern will have place-holders for more arguments than are present in the scene representation. In cases such as these, when creating the corresponding utterance, the excess argument slots in the predicted pattern are simply left blank.

and syntactic pattern.

The parameter α_i is an estimate of the number of possible values that feature i can take on. We assign $\alpha_{head} = 25$ and $\alpha_{patt} = 15$ as an upper bound on the number of the predicates and the distinct syntactic patterns that can be observed in the input. We assume that each predicate structure can have 0 to 5 arguments, therefore $\alpha_{arg_num} = 6$. The estimated α_i value for other features is determined by the number of semantic primitives, thematic roles, and semantic categories, respectively, as defined in Section 3.4.1: $\alpha_{prim} = 12$, $\alpha_{role} = 10$, and $\alpha_{arg_cat} = 13$.

We also need to set the parameter λ in equation (2.6). For all features i in frame F, we need λ to be small enough that the resulting $P_i(j|k)$, where j differs from the value in k, penalizes construction k in the competition in equation (2.1). If we wanted each construction k to contain only frames that have the same values on all of the features, we would want the posterior probability of k to be less than that of a new construction, or $P(k|F) < P(0|F)$. The extreme case would be when only one construction exists which contains all n observed frames, every one of which has the same values as F for all features, except for feature i. By simplifying the inequality $P(k|F) < P(0|F)$ for this case, we obtain $\lambda < \prod_l \frac{1}{\alpha_l}$. This gives us an lower bound of 10^{-7} for the value of λ. We set λ to a less strict value, 10^{-4}, to allow for constructions which have frames with similar but not identical values for some features.

In future, we need to investigate in more detail the effect of parameter setting in the output of our model. Whereas the value for each parameter α_i is determined by the number of the possible values that feature i can take, setting λ is more arbitrary. Although we set this parameter by making minimal assumptions it is worth examining the changes in the behaviour of the model due to a different setting for this parameter. We expect a smaller value for λ to result in the formation of more fine-grained constructions (i.e., those which contain frames with more similar feature values for all the features), which might limit the generalization capabilities of the model at the earlier stages of

learning. However, such effects might diminish over time as more input is processed by the model.

3.6 Experimental Results

We report computational experiments that demonstrate the ability of our model to learn constructions representing the knowledge of argument structure regularities. We also show that the model is able to generalize this knowledge to novel situations in language use. Language use is formulated as a set of prediction tasks corresponding to comprehension (prediction of semantic properties from a partial input) and production (prediction of a syntactic form—i.e., the word order of a verb and its arguments—from the semantics to be expressed). The simulations we report all use the same parameter settings; only the generated input corpora differ.

In the experimental results reported here, we first show how general constructions emerge as the model is exposed to a range of verb usages over time. We then describe the learning curves displayed by our model for some example verbs; these show interesting variations on the well-known U-shaped learning curve, in which use of (correct) imitative forms is followed by a period of (possibly incorrect) generalizations, which is eventually followed by recovery from overgeneralization errors. In successive subsections, we examine in more detail some interesting stages in the generally observed U-shaped learning curve: the process of generalization that follows use of imitative forms; the possible production of overgeneralization errors, and recovery from them; and the stable process of productive generalization. Appendix A shows the constructions learned by the model in one of the simulations, after processing 200 input pairs.

Due to the limited range of verbs in the lexicon, as well as other missing perceptual and linguistic input of importance in language development, it is not possible to directly compare the behaviour of the model to that of children from different age groups. How-

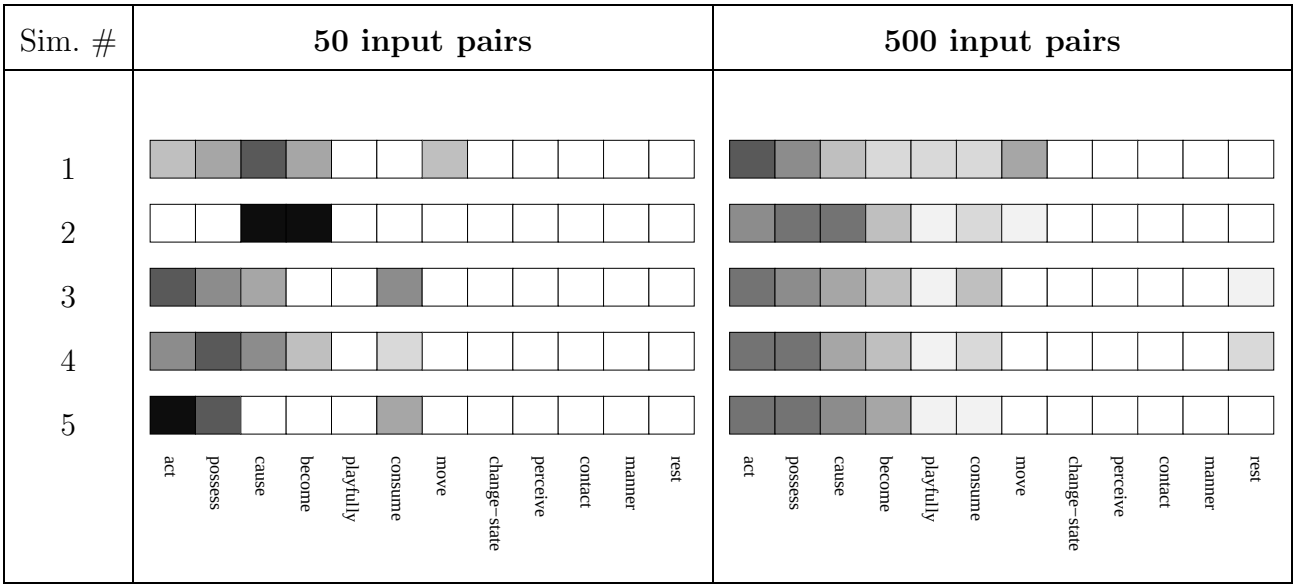


Figure 3.4: The probability distribution of semantic primitives in the transitive construction at two different points in learning. Darker squares correspond to higher probabilities.

ever, in the experiments reported below, we change the relative amount of input to which we expose our model to show that, over time, the model goes through different stages of learning similar to those observed for children.

3.6.1 Construction Formation

The formation of argument structure constructions is the primary means of organizing lexico-syntactic knowledge about predicates, since such constructions encode the mapping of predicate argument relations to their syntactic expression. Here we focus on verbs, since (in contrast to prepositions, the other type of predicate in our model) they exhibit a range of argument structure variations, with the potential for acquisition of interesting form-meaning mappings. In particular, we are interested in the probabilistic relations that develop in our model between verb semantic primitives (such as *act* and *cause*) and a formal pattern (the argument roles and their expression in word order).

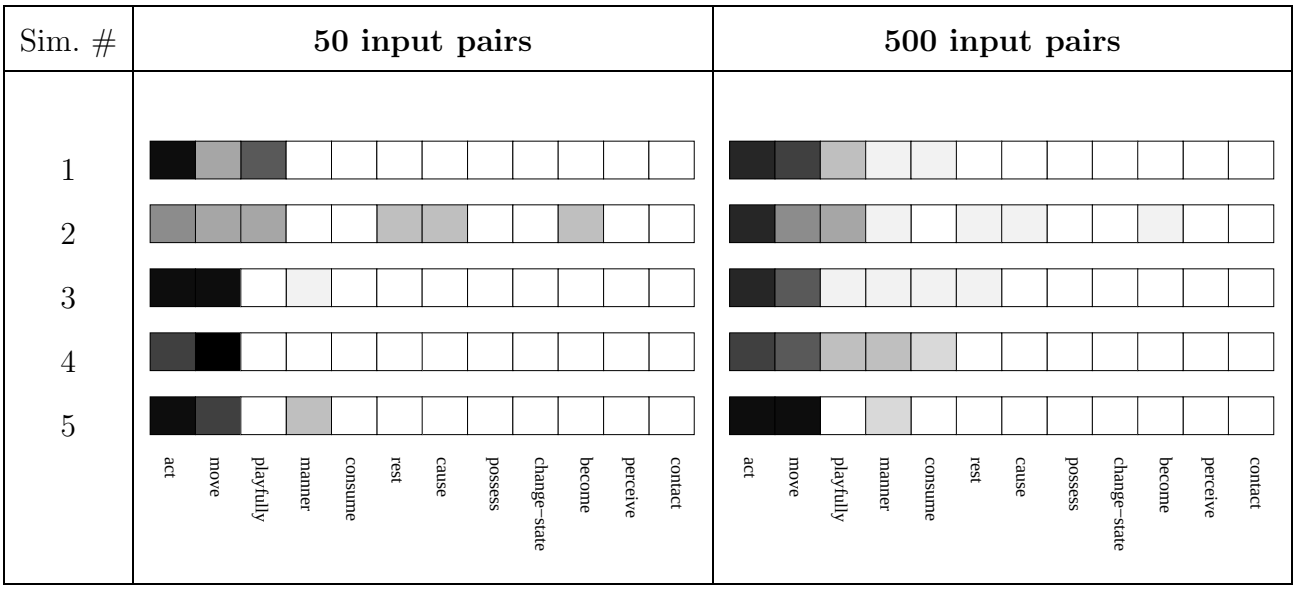


Figure 3.5: The probability distribution of semantic primitives in the intransitive construction at two different points in learning. Darker squares correspond to higher probabilities.

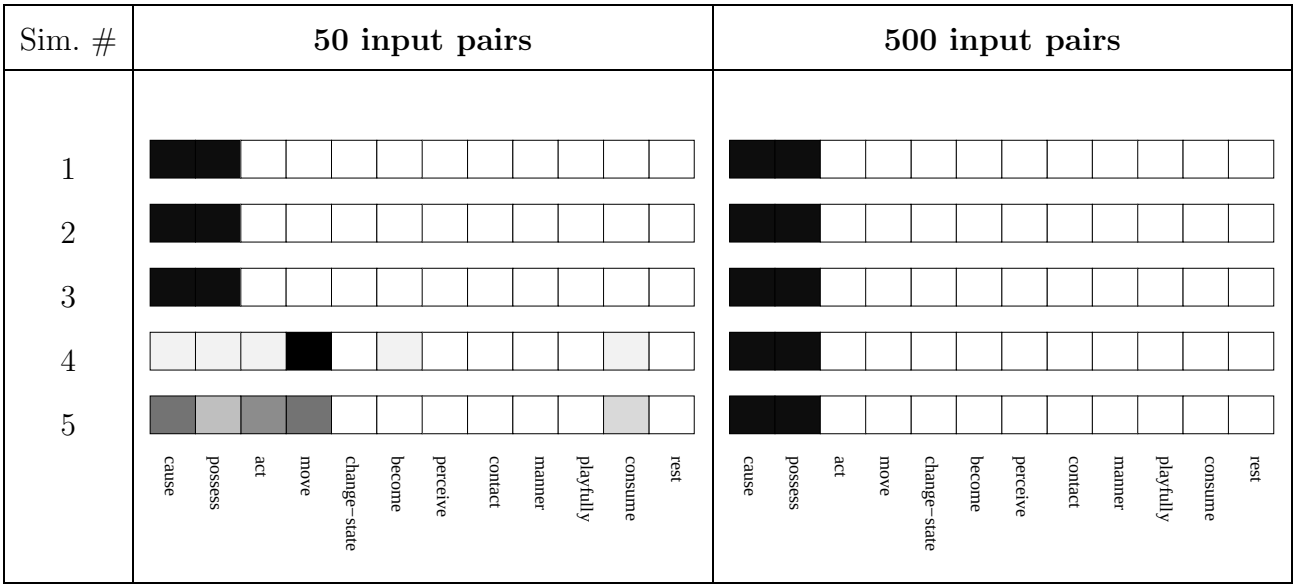


Figure 3.6: The probability distribution of semantic primitives in the ditransitive construction at two different points in learning. Darker squares correspond to higher probabilities.

Basic Construction Learning

We collect evidence for our view of construction learning by tracking the formation of several different constructions in five different simulations of our model (i.e., after processing five different input corpora). We examine the probabilities of the set of verb semantic primitives associated with transitive, intransitive, and ditransitive constructions, as follows. In each simulation, we train the model on either 50 input pairs (representing the early stage of learning) or 500 input pairs (a more advanced stage of learning). We then present the model with a typical usage of a novel verb in each construction, with the verb semantic primitives omitted. Given this partial input, the model must predict the missing semantic primitives.

Fig. 3.4, Fig. 3.5 and Fig. 3.6 show the probabilities of primitives associated with the transitive, intransitive, and ditransitive constructions, respectively, across the five simulations. (Note that the order of the semantic primitives is different in for each construction, because they are sorted in a way that represents the general ordering of the probability values in that construction.) Both the transitive usage (as in *Mary made cake*), and the intransitive usage (as in *John came*), occur frequently for a number of verbs in our lexicon. The formation of the two corresponding constructions shows a similar pattern; see Fig. 3.4 and Fig. 3.5. After only 50 input pairs, the probabilistic association is unevenly distributed among a small set of primitives, and differs for each simulation. However, after processing 500 input pairs, all the simulations display similar distributions across the full set of primitives, with higher probabilities assigned to the primitives more frequently associated with the construction.

For example, a closer look at the intransitive construction in Fig. 3.5 shows that, after processing 50 input pairs, the probabilistic distribution of semantic primitives differs across the five simulations. Moreover, each simulation shows a bias towards one or two of the verbs in the lexicon that appears in the intransitive construction: *act* and *playfully* in simulation 1 correspond to the verb *play*; *act* and *move* in simulations 3, 4 and 5 to the

verbs *go* or *come*, with *manner* corresponding to *dance* in simulation 5. In simulation 2, we see a more even distribution over *act*, *move*, *playfully*, *rest*, *cause* and *become*, due to receiving more intransitive usages than other simulations in the first 50 input pairs. However, after 500 input pairs, all the simulations display a similar distribution across the primitives *act*, *move*, *playfully*, *manner*, *consume* and *rest*, with a higher probability for *act* (because it occurs with all intransitive verb usages in our lexicon). Most other primitives have a probability near zero.

An interesting trend can be observed for the ditransitive construction (as in *John gave Mary the book*). In our lexicon, the only verb that appears in this construction is *give*. An examination of Fig. 3.6 reveals a very different pattern of development from that of the more common transitive and intransitive constructions. The primitives *cause* and *possess*, which are associated with the verb *give* in the input-generation lexicon, have a very high probability assigned to them from early on, and the probabilistic distribution for the construction does not change much after 500 input pairs. (In simulations 4 and 5, after receiving 50 input pairs, the ditransitive construction is not formed yet.) While this particular pattern of learning occurs for the ditransitive because of our limited lexicon, the emergence of highly specific constructions in the language more generally, such as the *way* construction (as in *She typed her way to a promotion*), would be expected to unfold similarly.

Using an Extended Lexicon

The experiments reported above are compatible with our view of the formation of constructions as a probabilistic association between a form (a syntactic pattern) and elements of meaning (the semantic primitives of predicates). However, the small number of verbs in the lexicon, and the limited set of semantic primitives associated with them (especially the more specific ones), raise the concern that the observed patterns may not be valid for a larger and more varied lexicon.

To address this issue, we artificially expanded our input-generation lexicon by adding new verbs that differ from existing verbs only in their fine-grained semantic primitives. We chose three verbs in the original lexicon, *sit*, *eat*, and *play*, as the starting point for each of three sets of six semantically and syntactically related verbs. For each of the three selected verbs, we added five verbs to the lexicon with the same argument structure frames and frequencies as the original verb. (Note that the verb *sit* only has intransitive frames, while the verbs *eat* and *play* have both intransitive and transitive frames.) The five added verbs also had the same general semantic primitive(s) as the original verb, but all differed in their fine-grained primitives. That is, for each verb v in the set {*sit,eat,play*}, we added to the lexicon five verbs, v_1 to v_5, each of which had the original general semantic primitives (e.g., *act*) associated with it. But in place of the fine-grained primitive *prim* that occurs with v (i.e., *rest* for *sit*, *consume* for *eat*, and *playfully* for *play*), each verb v_i is associated with the primitive $prim_i$. Thus each such group represents a class of verbs which have similar meaning and appear in the same argument structure constructions, and can in this way be thought of as a semantic verb class similar to those of Levin (1993).

Fig. 3.7 and Fig. 3.8 show snapshots of the transitive and intransitive constructions, respectively, after 50 and 500 input pairs, for one sample simulation using the extended lexicon. The same effect can be observed for both constructions. After 50 input pairs, the probabilities over the semantic primitives are sporadically distributed, with a few verb-specific primitives having much higher probability than others. However, after 500 input pairs, the probability mass is more evenly distributed among primitives associated with all verbs that appear in that construction, with more general primitives (such as *act*) having a higher probability. Most of those primitives associated with verbs from the same semantic "class" (e.g., all the verbs sit_i) have very close probabilities after 500 input pairs (but not after 50 pairs). These results confirm that the general properties of a frequent construction can be the result of probabilistic association between the semantic

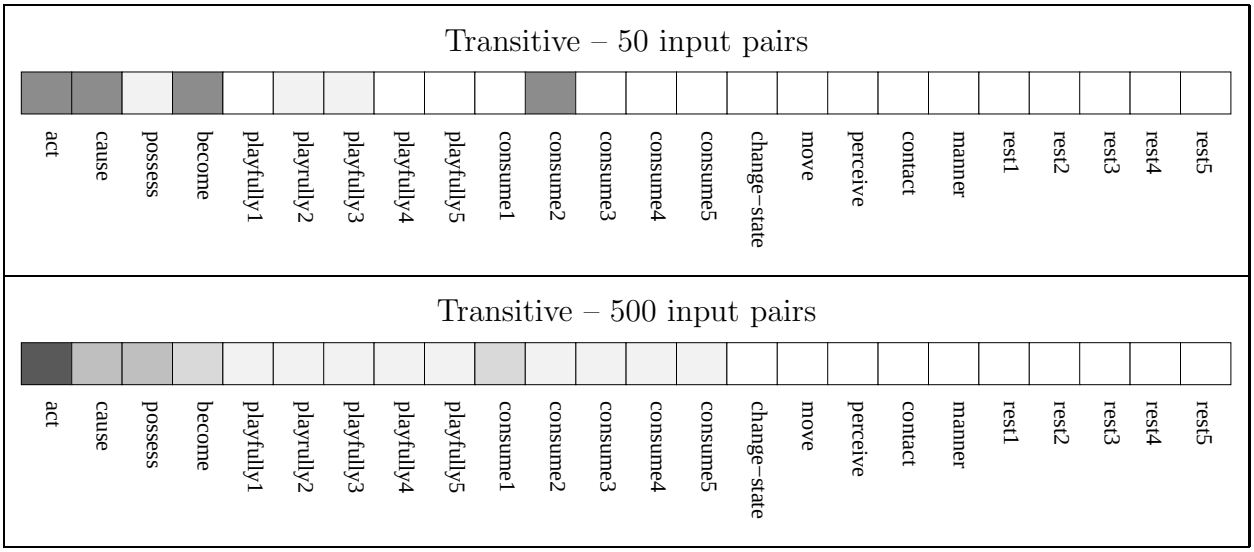


Figure 3.7: Transitive construction for the extended lexicon. Darker squares correspond to higher probabilities.

primitives and the form. In the long run, this is determined by the frequencies of all the verbs that appear in that construction.

3.6.2 Learning Curves

The simulations reported above present one view of learning in the model over time, focusing on the changing probabilistic associations within the constructions that are gradually learned. Here we turn to a different view of development in the model, focusing on the learning curve that reflects the accuracy of the model over time in using this acquired knowledge, and the effect of noise on the performance of the model. To investigate this, we look at the output for sentence production, i.e., how the syntactic pattern that the model predicts based on available semantic information varies over time as the model receives more input data.

In each simulation below, we ran our model over 600 input pairs divided into 100 sequences. Each sequence had 5 training input pairs, followed by a 6^{th} test input in which the utterance was removed. Our prediction model was used to find the syntactic

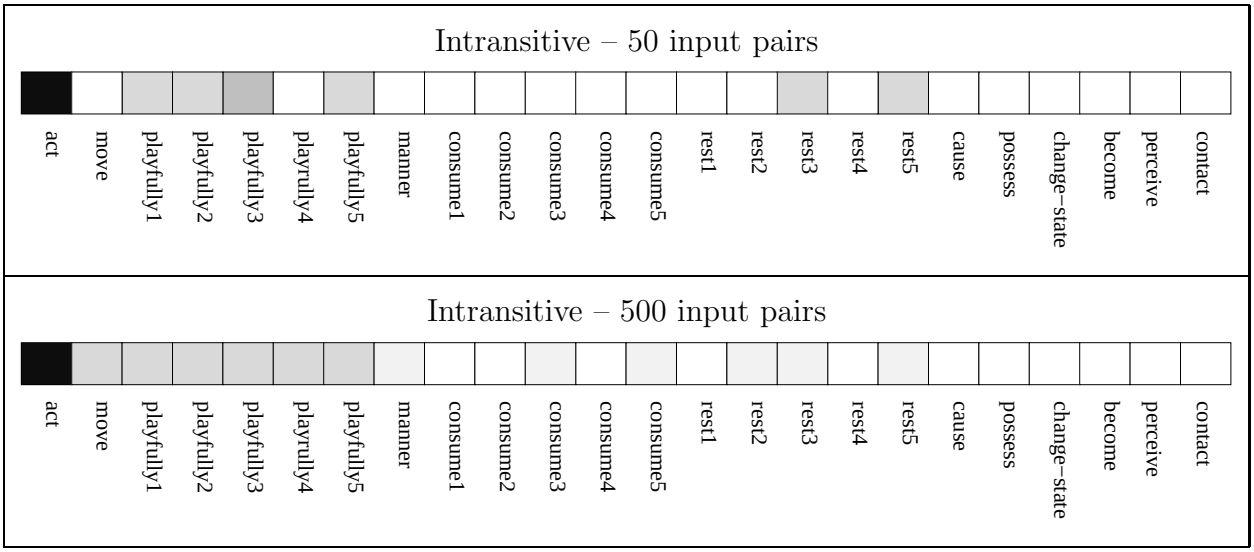


Figure 3.8: Intransitive construction for the extended lexicon. Darker squares correspond to higher probabilities.

pattern with the highest probability for the 6^{th} (partial) test frame of each sequence. (The model does not use the test input in learning.) After each test input, we measured the cumulative accuracy of the model by counting the total number of times, up through that sequence, that the predicted pattern for a test input was exactly the same as that used in the removed utterance. (Note that this is a very conservative estimate of correctness, since another syntactic pattern may express the given semantics equally acceptably.)

We first performed a set of 10 such simulations to examine the learning curve for the model over all verbs. Fig. 3.9 shows the learning curves obtained from four simulations on different randomly generated input corpora. Each panel of the figure includes the learning curves for both the original input corpus and its "noisified" version (see Section 3.5.2); this allows us to compare the effect of noisy and partial data on the performance and convergence of the model. Since the input corpora are randomly generated, the performance of the model varies across the simulations. However, we often observe early periods of fluctuation in the accuracy, followed by a more stable and accurate performance. It is evident from the graphs that the presence of noise mainly degrades performance of the

model at the early stages of learning. As more input is received and processed, the model typically recovers from early mistakes, and its accuracy in the presence of noise often approaches its accuracy in the absence of noise. However, there are some cases where the order of inputs and early prediction mistakes make it impossible to completely recover from noise (see the top right panel of Fig. 3.9). In an actual child language situation with much more exposure to language, we expect that such scenarios would be avoided.

The variations of the learning pattern of the model across different simulations resembles the differences between individual children: not all the children make the same mistakes, and their pattern of learning and production depends on the input they receive. Even siblings of the same gender have been shown to make different types of overgeneralization mistakes (Bowerman, 1982). Moreover, for a single child, the onset of overgeneralization errors is not the same for different verbs or constructions. We expect an overall U-shaped learning curve from our system, in which early periods of more accurate performance are followed by less accurate performance, before more consistent performance is eventually achieved. However, because the graphs above report accuracies over a range of verbs and argument structures, the hypothesized U-shaped curve is only roughly evident. We also examine the individual learning curves of every verb to see whether particular verbs exhibit a clear U-shaped pattern. For each verb, we ran eight separate simulations using the same methodology as above, but in this case, we used only "noisified" corpora and the test input always contained the same target verb. Fig. 3.10 shows four sample learning curves for each of the verbs *go* and *fall*. With testing of individual verbs, the variability in the performance of the model early on is more pronounced, due to variations in exposure to that verb in the random corpora. For frequent verbs with various argument structures, such as *go*, a U-shaped curve is often observed. The learning curve for *fall*, which is less frequent, shows a delay compared to more frequent verbs such as *go*, but the U-shaped curve can still be observed in most simulations.

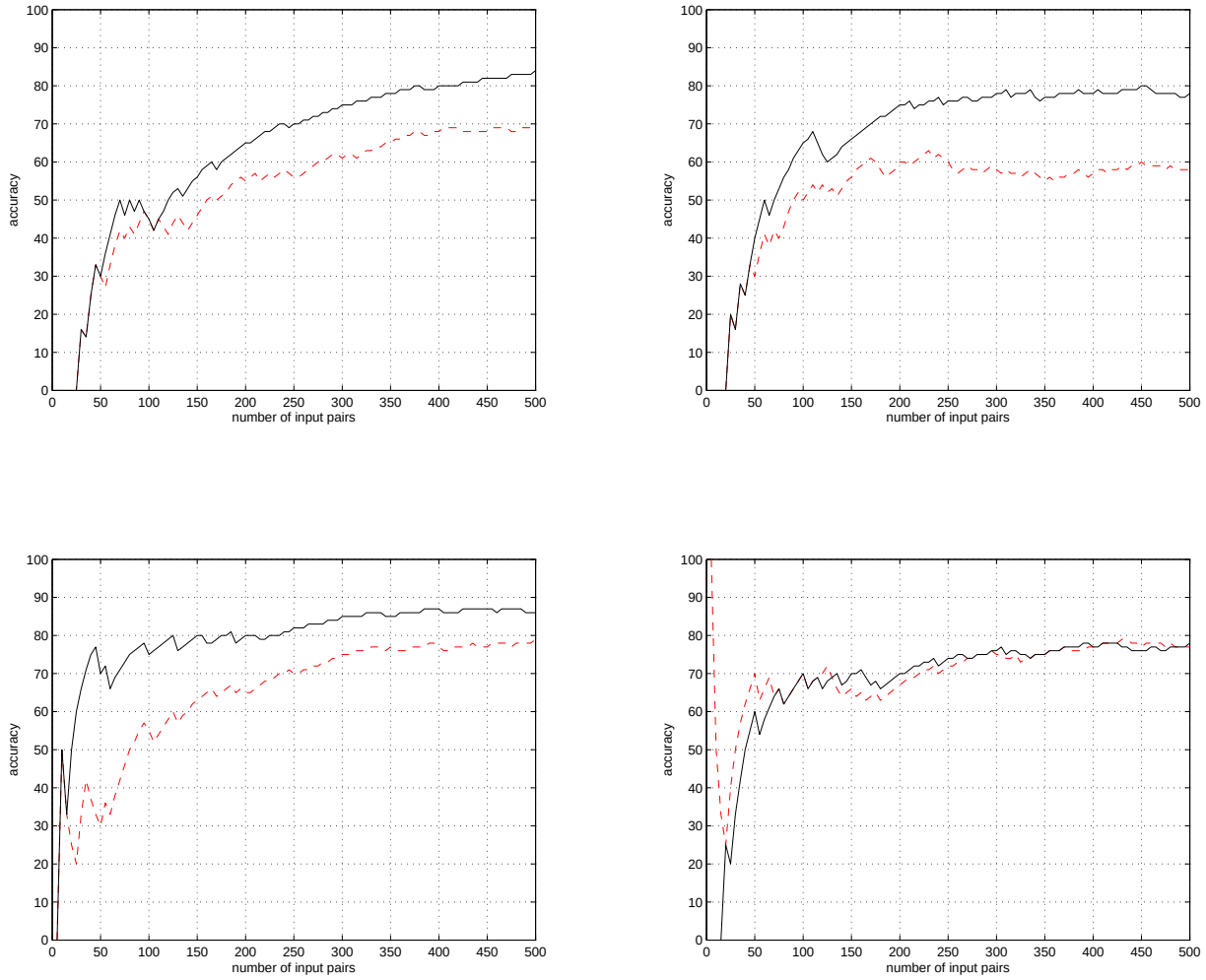
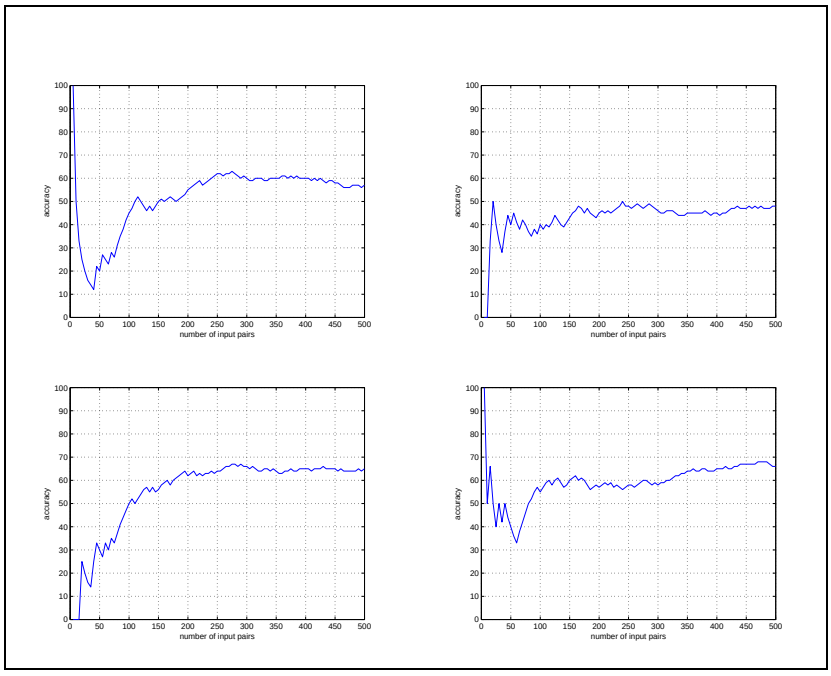


Figure 3.9: Cumulative accuracy of the model for predicting syntactic pattern. Solid lines show performance with noiseless input data, and dashed lines show performance with noisy input data.

go



fall

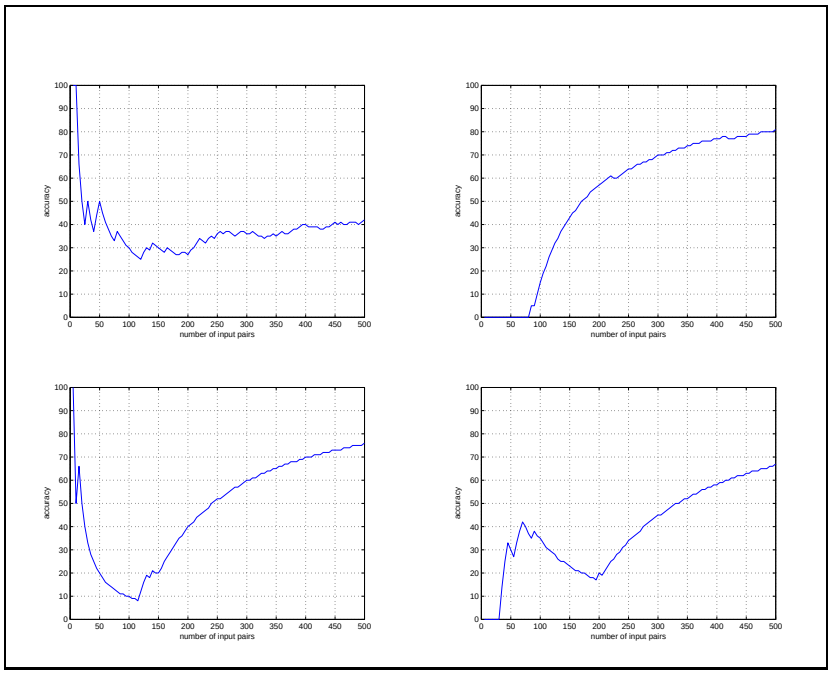


Figure 3.10: Sample learning curves for different verbs, showing cumulative accuracy over time.

Fig. 3.10 shows that the model generally exhibits the characteristics of a U-shaped learning curve observed for children. In the following subsections, we turn to a more detailed description of the behaviour of the model as it uses the knowledge it has acquired at various stages of learning along the "U": specifically, during generalization (including an initial stage of imitation), possible overgeneralization and recovery, and productive generalization.

3.6.3 Generalization

Here we investigate the ability of the model to generalize its item-specific knowledge in different tasks, drawing on the constructions it has acquired. First, we present the process of generalization in comprehension, showing that the model can use its learned constructions to recognize appropriate form-meaning pairings in the input. Next, we look at generalization in production, where the model uses the properties of a scene description to choose the best syntactic pattern with which to express the semantics. In both cases, we discuss the behaviour of the model in the context of observations of actual child acquisition. We use the same corpora for both sets of simulations in order to compare the onset of generalization in production and comprehension, which is discussed in Section 3.7.2.

Form-Meaning Mappings in Comprehension

A number of experiments have indicated that children use the evidence of syntactic form to infer general semantic properties of a novel verb (e.g., Naigles, 1990; Fisher, 2002), a phenomenon sometimes referred to as syntactic bootstrapping (Gleitman, 1990). For example, in Naigles's (1990) preferential-looking experiment, children of age 2 to 3 were shown two pictures of actions, presented with an utterance with a novel (nonce) verb, and told to "find" the novel action in the pair of scenes. Naigles (1990) found that children who heard an intransitive utterance (*The bunny and duck are blicking*) were more likely

to look at a picture of two characters independently performing an action (the bunny and duck each twirling their arms, for example), while those who heard a similar transitive form (*The bunny is blicking the duck*) were more likely to look at a picture of one character (the bunny) performing an action on the other (the duck). This result demonstrates that the children have learned a reliable association between a syntactic form (such as the transitive) and a coarse semantics for the expressed event (i.e., one participant causally affecting another), and are able to determine the scene that is more compatible with an utterance according to this acquired knowledge.

In this section, we use our computational model to simulate Naigles's (1990) preferential looking experiment. As described in Section 2.4, language use is modeled as choosing the best value j for a missing feature i, based on other observed features, using equation (2.7). We use this approach to model sentence production in the next section. However, the same approach cannot be used to directly mimic the conditions of a preferential-looking experiment of the type used in Naigles's study, because here the problem is to choose from two possible pairings of scenes with the utterance (similar to children choosing one of the two videos being shown upon hearing the utterance). However, we can demonstrate the ability to create an association between form and meaning by showing a preference in the model for the combination of an appropriate scene paired with an utterance, compared to an inappropriate scene paired with that utterance.

First, we consider transitive and intransitive utterances, U_T and U_I respectively, analogous to those used in Naigles's (1990) experiment:

U_T : *Bunny blick duck*

U_I : *Bunny and duck blick*

Next, we create corresponding scene representations analogous to the pictures shown to the children, in which one entity is causally affecting another, S_{CA}, or the two entities are performing independent actions, S_A:

S_{CA} : $\mathbf{Blick}_{[cause,act]}(\mathsf{BUNNY}_{\langle agent\rangle}, \mathsf{DUCK}_{\langle theme\rangle})$

S_{A} : $\mathbf{Blick}_{[act]}(\mathbf{And}_{[]}(\mathsf{BUNNY}, \mathsf{DUCK})_{\langle agent\rangle})$

These utterance and scene representations form the basis for input pairs used to test the model's acquired knowledge of the association between syntactic forms and semantic properties of a scene.

In one experimental condition, analogous to the child hearing the transitive form, each of the above scenes is combined with the transitive utterance, U_T, to form two input pairs, S_{CA}-U_T and S_A-U_T. The input pair S_{CA}-U_T corresponds to the appropriate selection of the causal action scene to go along with the perceived transitive utterance; we call this the "matched" pair. The input pair S_A-U_T corresponds to the inappropriate selection of the independent action scene for the utterance; we call this the "unmatched" pair. The other experimental condition uses analogous pairings with the intransitive form, U_I. The four resulting input pairs are shown in Table 3.3.

We then (separately) present each of these four input pairs to our model. The model will extract a frame F from the presented input pair, and will determine the best construction k for that frame F. The model records the value of $P(k|F)$ (equation (2.1) on page 21), for the best construction k given F, as its response to each input. If the input pair yields a frame that is compatible with an existing construction, this indicates that the scene/utterance combination corresponds to a form-meaning association in the model, represented as an existing construction. In this case the value of $P(k|F)$ will be relatively high. If the extracted frame is not compatible with an existing construction, then the best construction for the input pair is a new one, which will have low prior probability. Thus, when comparing the values recorded in response to the matched and unmatched pairing for each utterance, a higher value of $P(k|F)$ corresponds to the child "recognizing" the appropriate scene for the utterance.

Table 3.4 shows the value of $\log P(k|F)$ across the conditions, after varying amounts of learning (10, 50, 100, 150, and 500 training input pairs; averaged over 10 simulations).

U_T-S_{CA}	*Bunny blick duck.* $\mathbf{Blick}_{[cause,act]}(\mathsf{BUNNY}_{\langle agent\rangle}, \mathsf{DUCK}_{\langle theme\rangle})$	Matched
U_T-S_A	*Bunny blick duck.* $\mathbf{Blick}_{[act]}(\mathbf{And}_{[]}(\mathsf{BUNNY}, \mathsf{DUCK})_{\langle agent\rangle})$	Unmatched
U_I-S_{CA}	*Bunny and duck blick.* $\mathbf{Blick}_{[cause,act]}(\mathsf{BUNNY}_{\langle agent\rangle}, \mathsf{DUCK}_{\langle theme\rangle})$	Unmatched
U_I-S_A	*Bunny and duck blick.* $\mathbf{Blick}_{[act]}(\mathbf{And}_{[]}(\mathsf{BUNNY}, \mathsf{DUCK})_{\langle agent\rangle})$	Matched

Table 3.3: The four test input pairs used in the experiment of Section 3.6.3.

	# Training Input Pairs				
Test Input Pair	10	50	100	150	500
U_T-S_{CA} (matched)	-7	-6	-6	-6	-6
U_T-S_A (unmatched)	-7	-7	-7	-8	-8
U_I-S_{CA} (unmatched)	-10	-10	-10	-10	-11
U_I-S_A (matched)	-4	-4	-3	-3	-3

Table 3.4: $\log P(k|F)$ values for matched and unmatched scene/utterance pairs.

(We report the log probabilities because the actual probability values are very small.) The sizable difference between the two (matched and unmatched) pairs for each utterance type mimics the child's ability to pick out the appropriate scene for an utterance based on learned argument structure regularities. Even when the log values are the same (i.e., after 10 input pairs in the transitive condition), the actual score value for the matched pair (7.5×10^{-7}) is almost four times higher than the one for the unmatched pair (1.9×10^{-7}). Although this does not directly show that the model can *infer* aspects of the meaning of a verb from its syntactic usage, as "syntactic bootstrapping" proponents have suggested, we do demonstrate that ability in the later experiments on productive generalizations

(Section 3.6.5).

Generalization in Sentence Production

The above simulations show that the model can use its knowledge of constructions in comprehension. In this section, we turn to the use of constructions in production. Specifically, we examine the behaviour of the model when presented with a novel verb for which it must produce a sentence. Children can, after a certain age, use a novel verb in a construction that has not been observed for that verb, when the construction is one to which there has been frequent exposure (Maratsos et al., 1987; Pinker et al., 1987; Tomasello et al., 1997). In some experiments, a novel verb is modeled in some construction and then children are expected to produce it in a related construction (Maratsos et al., 1987; Pinker et al., 1987). In others, the verb is explicitly introduced through minimal syntactic contexts, such as *Look! Gorping!*, so that only the relation between the verb form and the semantics of the action is observed by the child (Tomasello et al., 1997).

We show a similar ability in our model through the following simulation. After training the model, we present it with the following input pair, in which the scene representation is as given and the utterance is left empty:

S_{G1}: $\mathbf{Gorp}_{[cause,act]}(\mathsf{KITTY}_{\langle agent \rangle}, \mathsf{DOGGY}_{\langle theme \rangle})$

U_{G1}: ?

The main predicate, **Gorp**, corresponds to a verb, *gorp*, which has not been previously seen by the model in an utterance. We build in the knowledge that the predicate **Gorp** is the meaning of *gorp*, analogous to children being told that "this [demonstrated action] is gorping", as in the Tomasello et al. (1997) experiments. Given the above utterance-less input, the model will extract a frame that has no syntactic (word order) pattern. We then monitor the syntactic pattern that the model predicts in response. Since the scene representation resembles the meaning of a typical transitive usage, we expect the model to predict the transitive syntactic pattern, "arg1 *verb* arg2", despite not having seen

		# Input Pairs				
		10	50	100	150	500
Generated Pattern	observed (SOV)	0.90	0.70	0.40	0.00	0.00
	"corrected" (SVO)	0.10	0.30	0.60	1.00	1.00

Table 3.5: Average probability of the predicted patterns.

usages of this particular verb. To observe the pattern of responses over the course of acquisition, we test the model after varying amounts of training data. Averaging over 10 simulations on different input corpora, the model predicts the expected transitive word order pattern with 100% probability after processing 50 input pairs. These results show that our model, like children, can start generalizing ubiquitous constructions to new verbs from early on.

More interesting is the varying influence of specific (verb-based) and general (construction-based) knowledge in production over the course of acquisition, which is a key feature of our model. To explore this, we mimic the conditions of an experiment by Akhtar (1999), in which English-speaking children aged 2 to 4 years old were taught novel verbs used in non-standard English word orders. (These were Subject-Object-Verb [SOV] or Verb-Subject-Object [VSO], rather than the standard Subject-Verb-Object [SVO]). In both spontaneous and elicited productions of these verbs, 2- and 3-year-olds imitated the observed patterns roughly half the time and "corrected" the order to the English SVO pattern roughly half the time. On the other hand, the 4-year-olds rarely imitated the observed order, almost always correcting to SVO order. Thus, the children show an early stage of imitation, followed by a later stage in which they generalize the pervasive word order pattern of their linguistic experience to a new verb, even when that verb has been observed in a different pattern.

We trained our model using different numbers of input pairs to simulate differing amounts of exposure to the language, roughly corresponding to different "ages". We then

provided the model the following training input pair with a novel (previously unseen) verb used in an utterance with a non-standard word order:

S_{G2}: $\mathbf{Gorp}_{[cause,act]}(\mathsf{KITTY}_{\langle agent \rangle}, \mathsf{DOGGY}_{\langle theme \rangle})$

U_{G2}: *kitty doggy gorp*

Note that the scene representation, S_{G2}, resembles observed scenes with transitive (SVO) utterances, but U_{G2} has a different (SOV) word order. Next, we presented the same input pair to the model, but without the utterance, as in S_{G1}-U_{G1} earlier, and recorded the syntactic pattern predicted by the model.

We performed 10 simulations for each number of training input pairs, and averaged the probability of predicting each of the SOV (observed) and SVO ("corrected") patterns. As can be seen in Table 3.5, when 100 input pairs have been processed, the model predicts the observed pattern a little more than half the time, and the corrected pattern a little less than half the time. This resembles the 2- and 3-year-old age group, who imitate and correct about equally often. However, when the number of input pairs is small, the model is more likely to imitate in production the single usage of the new verb that it has seen. When the number of input pairs goes up to 150 or more, the model predicts the corrected pattern almost all the time, which is similar to what the 4-year-old children do. These results show that the model, like children, shows a shift with more exposure to the language, from imitation, in which even a form that is incompatible with the ambient language is repeated, to generalization, where the ubiquitous patterns of the language predominate in production.

3.6.4 Overgeneralization and Recovery

In addition to enabling effective generalization of acquired knowledge, as shown above, the use of general knowledge sometimes leads children to overgeneralize. However, children eventually recover with only additional positive evidence. A typical overgeneralization in

Adam	Our model
go fall!	*John fall ball*
no no fall no!	*toy fall*
no fall!	*Mary fall book*
oh Adam fall.	*toy fall*
Adam fall toy.	*cookie fall*
Adam fall toy.	*kitty fall*
oh fall.	*spoon fall*
I not fall.	*ball and toy fall*

Figure 3.11: Comparing the first eight uses of *fall* for Adam with the first eight sentences generated by our model.

the area of argument structure is when a non-causative (intransitive-only) verb is used causatively (in a transitive structure). This occurs in an example such as *Don't you fall me down* (Bowerman, 1982), in which the intransitive verb *fall* is used transitively, on analogy to a verb like *drop*, that can be used in both forms. In the simulations reported here, we tracked the usage of *fall* by our model to see if we can detect a pattern of overgeneralization and recovery similar to that in children.

The entry for *fall* in the input-generation lexicon allows only an intransitive syntactic pattern (as in *The blocks fell*). However, the scene representation for a use of *fall* may include the agent who caused the falling (e.g., if Adam pushed the blocks over), since this may be part of the observation of the child. Each use of *fall* in these simulations has an intransitive syntactic pattern with 1.00 probability, but includes a causal agent in the scene description with a .50 probability. We therefore expect the semantic similarity of the scene to that of a transitive construction to sometimes lead to overgeneralization—i.e., the prediction of a transitive syntactic pattern with *fall* in a scene with a causal agent, even though this pattern has never been observed with *fall*.

In these simulations, we test the behaviour of the model in producing a syntactic pattern for *fall* over the course of acquisition. Every 5 training input pairs was followed by a test pair in which the utterance was left empty, and the scene representation had **Fall** as its main predicate. We then record the prediction of the model for the missing syntactic pattern. (Note that the model does not learn from this test pair.) Over 10 such simulations, the average probability of receiving a training pair containing *fall* was .01, and the average probability of receiving an instance of the transitive construction was .16. In all 10 simulations, the model showed a pattern of overgeneralization and recovery—using the transitive syntax for *fall* at some point, and eventually producing only correct intransitive forms as it received more examples of *fall* in the training input.

Fig. 3.11 shows the first 8 uses of *fall* for Adam in CHILDES (at 27 months), together with the first 8 sentences generated by our model in one of the simulations. The output of the model illustrates the mix of the two (transitive and intransitive) syntactic patterns at this stage. (Note that a number of the patterns shown in Adam's speech are unknown to our model, such as negatives and imperatives.) By the age of 31 months, the causative (transitive) uses of *fall* gradually disappear from Adam's conversations. Over the 10 simulations in our model showing the pattern of overgeneralization and recovery, causative sentences were no longer output after processing an average of 125 training inputs.

3.6.5 Productive Generalization

We noted that children sometimes mistakenly overgeneralize, but eventually recover from these errors only by receiving additional positive evidence. However, this ability to converge on appropriate argument structures for each verb does not prevent a speaker from making productive generalizations in which a verb may be used in a construction that is "unusual" for it in order to convey particular semantic properties (Goldberg, 1995). For instance, in *The fly buzzed into the room*, the sound emission verb *buzz*

WORD:	dance
SYMBOL:	DANCE
CATEGORY:	[predicate]
FRAME:	DANCE($1)
ARG-NO:	1
SPRIMITIVES:	act,manner
ARGS:	$1 is [agent]
CONDITION:	$1 is [animate]
PATTERN:	$1 [word]

Figure 3.12: The lexical entry for *dance* in the input-generation lexicon.

adopts a manner of motion interpretation due to the directional movement construction it occurs in. We showed above that phases of overgeneralization and recovery, similar to those of children, can be observed in the output of our model. In this section, we demonstrate that the model can handle instances of productive generalization as well. We test our model with a verb appearing in an unusual (for that verb) construction, and show that the model can determine appropriate semantic properties of the usage based on what it has learned for that construction when used with other verbs. We also show that the model can produce an appropriate syntactic form for a predicate used with semantic properties that are unusual for it.

We add a new verb *dance* to the input generation lexicon, with a single intransitive frame, as shown in Fig. 3.12. We train the model on 500 training input pairs, an amount corresponding to an advanced stage of learning. In all training pairs in which *dance* appears, it is used intransitively. We then present the model with the following input

WORD:	put
SYMBOL:	PUT
CATEGORY:	[predicate]
FRAME:	PUT($1,$2,$3)
ARG-NO:	3
SPRIMITIVES:	cause, move
ARGS:	$1 is [agent]
ARGS:	$2 is [theme]
ARGS:	$3 is [destination]
CONDITION:	$1 is [animate]
CONDITION:	$2 is [concrete]
CONDITION:	$3 is [dest-pred]
PATTERN:	$1 [word] $2 $3

Figure 3.13: The lexical entry for *put* in the input-generation lexicon

pair, in which *dance* is used in a different argument structure, and the semantic primitives of the predicate and the semantic roles of the arguments are omitted:

S_{P1}: $\mathbf{Dance}_{[???]}(\mathsf{KITTY}_{\langle ? \rangle}, \mathsf{DOGGY}_{\langle ? \rangle}, \mathbf{Under}_{[]}(\mathsf{TABLE}_{\langle ? \rangle})_{\langle ? \rangle})$

U_{P1}: *kitty dance doggy under table*

The partial scene representation corresponds to a child hearing the verb *dance* in this novel syntactic form, and being able to determine the mapping of syntactic arguments to semantic argument positions. (That is, we have built in the output of that process into the scene representation, since our current implementation cannot support the prediction of an entire scene representation from an utterance alone.) Given this partial input, the

model must predict the multiple missing semantic features, based on the utterance.

Averaged across 10 simulations, the model predicts novel semantic primitives for the predicate **Dance** in this usage with a probability of .46 for *cause*, .33 for *move*, .19 for *possess*, and negligible probabilities for other primitives. The roles predicted for the three arguments, with associated probabilities, are *agent* (.99), *theme* (.99), and *destination* (.60), respectively. (The probability of the predicted role for the last argument is lower than the other two. Because other prepositional phrases may appear as the third argument in this syntactic form, different role assignments are possible for that argument.) Thus the model assigns semantic properties that capture the causative movement intended by this novel usage of *dance*. Like humans, the model can successfully comprehend the meaning of a situation described by a sentence with a verb used in an unusual construction. It accomplishes this by generalizing the feature values of a construction associated with other verbs, in this particular case, one corresponding to the usage of a verb such as *put*, shown in Fig. 3.13.

We also test the ability of the model to use its learned constructions to correctly produce a syntactic form for a predicate used with an "unusual" semantics. Here, we present the model with an input pair having the completed scene representation corresponding to S_{P1} and the utterance missing, as in:

S_{P2}: $\mathbf{Dance}_{[cause,act]}(\mathsf{KITTY}_{\langle agent\rangle}, \mathsf{DOGGY}_{\langle theme\rangle}, \mathbf{Under}_{[]}(\mathsf{TABLE})_{\langle destination\rangle})_{\langle destination\rangle})$

U_{P2}: ?

We then record the model's prediction of the most probable syntactic pattern. The appropriate sentence, *kitty dance doggy under table*, is produced as expected.

Table 3.6 shows the probabilistic association of verb semantic primitives with the verb *dance* in a typical usage (e.g., *kitty dance*), and the "unusual" one (as in U_{P1}), after processing 50 and 500 training input pairs. It can be seen that the semantic primitives predicted for the unusual usage are as strong as those in a typical usage. This shows that the model is as powerful in making productive generalizations as in applying the

	# Input Pairs	
Primitives	50	500
act	.44	.41
move	.34	.31
playfully	.18	.14
manner	.04	.09
consume	10^{-4}	.02
rest	10^{-4}	.01
cause	10^{-4}	10^{-3}
change-state	10^{-4}	10^{-5}
possess	10^{-4}	10^{-5}
become	10^{-4}	10^{-5}
perceive	10^{-4}	10^{-5}
contact	10^{-4}	10^{-5}

Typical usage

	# Input Pairs	
Primitives	50	500
cause	.41	.46
move	.39	.33
possess	.04	.19
act	.02	10^{-2}
become	10^{-3}	10^{-3}
change-state	.09	10^{-4}
playfully	10^{-3}	10^{-5}
consume	10^{-3}	10^{-5}
rest	10^{-3}	10^{-6}
perceive	10^{-3}	10^{-6}
manner	10^{-3}	10^{-6}
contact	10^{-3}	10^{-6}

"Unusual" usage

Table 3.6: Probabilities of semantic primitives associated with a typical and "unusual" usage of the verb *dance*.

item-based information in a typical usage of a verb.

3.7 Discussion of the Results

The onset of abstract knowledge in children has been extensively studied, with the goal of determining when children internalize and use general patterns in their language. This is an important step in language acquisition, as it allows the learner to creatively and productively use their language. The issues surrounding this ability are particularly interesting in the domain of argument structure knowledge, which is a very complex aspect of language for children to master. The computational experiments presented here show that knowledge of argument structure constructions can be represented as probabilistic form-meaning associations that reflect generalizations over specific verb usages. Such constructions emerge in the model over time, reflecting stages that match well with those of children: early imitation of observed forms; later generalization of such observed knowledge to new situations; mistakes reflecting overgeneralization and eventual recovery from such errors; and attainment of established knowledge that allows for productive generalization. Here we discuss the reported behaviour and underlying mechanisms of our model in the context of alternative accounts of argument structure acquisition and early use of such knowledge.

3.7.1 The Emergence of Constructions

There are a number of differing views on how the knowledge of constructions is acquired. One perspective is that argument structure patterns are initially learned on a verb-by-verb basis; then, through a process of categorization and generalization over the input, constructions are formed by associating certain meaning components with common syntactic patterns (Goldberg, 1999; Tomasello, 2000a). In Goldberg's (1999) account, the generalization of constructional meaning is based largely on the meanings

of highly frequent and semantically general "light" verbs, such as *go*, *do*, *make*, *give*, and *put*. Goldberg (1995) shows that the usages of such light verbs correspond closely to the meanings associated with argument structure constructions. For example, the intransitive construction, in which the pattern "Subject$_X$ Verb Oblique$_Y$" is paired with the meaning'X moves Y', corresponds to a core usage of the light verb *go*.[3] Goldberg suggests that children record a correlation between a certain formal pattern and the meaning of a specific verb that is used earliest and most frequently in that pattern. Because light verbs are more frequent than other verbs and are learned early, they tend to be the ones around which constructional meaning centers. Thus, constructions associate the semantically general meanings of light verbs with basic syntactic forms.

However, Tomasello (2003) notes that the evidence from actual child language data is not so clearcut, and that a child may use a number of different verbs in a construction very early on in her learning, depending on the verbs she has been exposed to most frequently. This accords with our model, in which constructions arise regardless of the generality of any particular verb's semantic properties. Because a construction in our model is composed of a usage-based probabilistic association between syntactic and semantic properties, all verb usages with similar properties contribute to the establishment of the construction. Our experiments show that, while an increasing number of semantic primitives may, over time, become weakly associated with a construction given exposure to a greater variety of verbs, the most general semantic primitives will become strongly associated with the construction because they will be observed across many more frames. For example, we showed that the general primitive *act* has a very high probability given a transitive form, compared to more specific primitives like *cause* and *become*, and even more so compared to very fine-grained primitives like *consume*. Though constructions may be likely to reflect light verb usages due to the frequency of such verbs, the emergence

[3]Here, Oblique refers to a prepositional or adverbial argument indicating a direction or goal, as in *Hans goes to school* or *Ingrid went home*.

of a construction need not be dependent on the presence of a light verb as its core. Thus this same probabilistic mechanism can explain both the emergence of semantically general constructions, such as the transitive or intransitive, as well as specific constructions (as in *He poisoned/wangled/hit/pitched/battled his way to a promotion*).[4]

3.7.2 Generalization in Comprehension and Production

The development of abstract argument structure knowledge in children appears to be reflected differentially in comprehension and production processes. Recent psycholinguistic evidence shows that very young children use verbs conservatively and only in those constructions in which they have heard them used before (Tomasello, 2000a). It seems that children do not produce utterances in accord with general (verb-independent) regularities in the mapping of form to meaning. In contrast to such production evidence, however, experiments on language comprehension using preferential-looking techniques indicate that children show signs of being aware of such argument structure regularities much earlier (Naigles, 1990; Fisher, 2002). To explain this, Tomasello and Abbot-Smith (2002) suggest that children begin by learning *weak* constructions which enable only certain linguistic operations, such as comprehension. Over time, children develop a more robust construction representation that supports use in production.

This timing of construction use in different tasks can be observed in the behaviour of our model without recourse to a notion of qualitatively different construction representations. Because our model seamlessly integrates specific verb-based and more general construction-based knowledge, it shows an interplay of imitation and generalization over the timecourse of learning that can account for the observed effects. Common constructions emerge very early on (usually after 10 training input pairs), enabling the model to successfully recognize (comprehend) established form-meaning mappings, as shown in our simulation of the preferential-looking (comprehension) experiments of Naigles (1990).

[4]Examples adapted from Google searches on "his way to a promotion", 19 June 2006.

Appropriate use of such general knowledge in production is not established in the model until at least after 100 input pairs, as shown in our experiment that mimics that of Tomasello et al. (1997). Moreover, such general knowledge cannot consistently override contrasting verb-specific information until later (after 150 input pairs), as shown in our simulation of Akhtar's (1999) elicitation experiments. In this way, our model naturally accounts for the child acquisition data, without having to posit different mechanisms of construction representation and use over time.

3.7.3 Mechanisms for Recovery from Overgeneralization Errors

While generalization over the observed input is an important means for a language learner to attain productive linguistic competence, it can also lead to mistaken application of observed patterns. How children recover from such overgeneralization errors has been the subject of much debate. There are a number of reasonable explanations of how children learn the productive patterns and constructions of their language, and why they overgeneralize them at times. However, accounting for how children recover from making overgeneralization errors is a challenge, especially given the general agreement that children are not being guided in the recovery process by negative evidence, in the form of explicit corrective feedback (Marcus, 1993). The complexity of this problem has made it a focal point in theories of argument structure acquisition.

Researchers who support an item-based approach to language acquisition have proposed a number of learning mechanisms as factors in recovery from overgeneralization, including entrenchment (Braine and Brooks, 1995; Goldberg, 1995), competition (MacWhinney, 1987), and cue construction (MacWhinney, 2004, drawing on Bowerman, 1982 and Pinker, 1984). However, none of these mechanisms alone can explain all aspects of the timecourse of development of the productive use of language. We briefly present each in turn.

Degree of entrenchment (Braine and Brooks, 1995) simply refers to the token fre-

quency of an item. Once an argument structure pattern has been learned for a particular verb, that pattern tends to block the creative use of the verb in any other argument structure pattern, unless the competing pattern is also witnessed in the input. The role of entrenchment is more articulated in the Competition Model (MacWhinney, 1987). There the productive use of general linguistic patterns is viewed as a process governed by two types of pressures: the underlying analogic force that produces the generalization, and the growth in the rote episodic auditory representation of an observed form. The entrenchment of the rote representation over time ensures the recovery from errors when the analogic process yields overgeneralizations. Both of these proposals, however, focus on morphological and syntactic generalizations, dealing with such errors as *He falled the ball.* However, these accounts fail to explain errors that stem from semantic overgeneralization of a construction. This occurs in situations such as the application of a resultative construction in **The child played the game finished*, on analogy to *The child hammered the putty flat*. Here, the overgeneralized verb *played* is not semantically appropriate for this situation due to a lack of causation of a change of state (in contrast to *hammer*). To address this type of error, cue construction suggests that the process of recovery from overgeneralization leads the child to construct new semantic features to block the application of the analogic pattern. Given the complementarity of these various mechanisms, MacWhinney (2004) has suggested that all of them must be used by the learner, in order to account for different types of overgeneralization errors and recovery from them. He also proposes that the learner estimates the "typical" rate of generalization for each form, whose distribution serves as "indirect negative evidence" in recovery from overgeneralization. Goldberg (2006) similarly suggests combining multiple cues in a formal way to model productivity: each cue is represented as a "weak hypothesis"; a boosting algorithm evaluates the prediction of each hypothesis against training items and changes the weight of that hypothesis accordingly. The output of the algorithm is a weighted sum of the predictions made by all the hypotheses. However, evaluating the

hypotheses depends on the availability of some sort of feedback or negative evidence. Goldberg suggests that statistical preemption be used to estimate such feedback. (The process of statistical preemption is discussed further in the next section.)

The Bayesian approach we propose here uses a single core mechanism, the prediction model, to govern all aspects of the productivity of acquired constructions. This mechanism displays all the effects that the above diverse mechanisms collectively describe. For example, when faced with a language use task such as sentence production, the prediction model chooses the best syntactic form based on the probability of each of the possible forms. In calculating these probabilities, both of the "pressures" suggested by MacWhinney (1987) are involved: the underlying analogic force is represented in the high value for the posterior probabilities of the constructions that match the situation in hand, and the rote episodic memory of a correct form is reflected in the verb-specific knowledge learned by the model. While in MacWhinney's (2004) view, this competition process handles syntactic generalization and recovery, he posits that cue construction is required for recovery from semantic overgeneralization. In our model, however, we learn and store syntactic and semantic knowledge as a single unit—the argument structure frame—whose probabilistic associations are used to make both syntactic and semantic predictions. Therefore, what is called "cue construction" happens in our model analogously to the competition between syntactic forms: more fine-grained semantic features of each construction are learned and highlighted over time by processing more input data.

Both verb-specific and construction-based sources of predictions in our model are weighted by their corresponding frequencies in the input data—i.e., their degree of entrenchment. Higher frequency of a construction results in a higher overgeneralization rate, whereas receiving and processing more instances of a correct specific form leads to recovery from overgeneralization. This intuition is supported by Theakston (2004), who shows that human subjects rate sentences with argument structure errors containing low frequency verbs more grammatical than those containing high frequency verbs.

Because not seeing a particular usage in the input means both lower probability for that usage as well as higher probabilities for competing ones, indirect negative evidence in our model simply results from the differential frequencies of different usages in the input data, without any explicit calculation of generalization rates.

3.7.4 Comparing Recovery and Productive Generalization

The process of entrenchment, or hearing a pattern with sufficient frequency, plays a key role in constraining overgeneralization. However, despite entrenchment of observed forms, it is still the case that verbs can sometimes be used creatively in new argument structure patterns, without any trace of ill-formedness. Upon closer inspection, effects that have been ascribed solely to entrenchment may be better described as the outcome of statistical preemption, a more encompassing process that critically involves the role of semantic or pragmatic contrast (Goldberg, 2006). In this view, overgeneralizations are constrained by the fact that more specific knowledge preempts general knowledge in determining a usage in production, given sufficient statistical support. However, this preemption can only occur if the more specific usage satisfies the functional demands of the context as well as the usage supported by the general knowledge. On the other hand, if a more general form better fits the semantic and pragmatic constraints of the situation, it may be felicitously used even in the presence of differing entrenched knowledge of the specific verb.

Entrenchment and statistical preemption are integrated into a single mechanism in our prediction model. The probability of each usage, as mentioned before, is a function of its frequency. Therefore, the more entrenched (i.e., the more frequent) a usage is, the more weight it carries in the probabilistic comparison. Importantly, we compare *frames*, not just syntactic forms, so that the probabilities take into account the satisfaction of both syntactic and semantic properties of the situation. As a result, when a verb-specific frame becomes entrenched enough, it generally wins the competition with construction-

based predictions, resulting in a decline over time in overgeneralization errors. However, if the model encounters a situation for which none of the well-entrenched frames for a verb are suitable enough, the model uses a matching general construction which is compatible with the scene.

Unlike overgeneralization errors, the ability of the model to generate novel utterances for unusual situations (or comprehend the meaning of the unusual utterances) does not fade over time by processing more input. Crucially, in these "unusual" cases, the model has not learned a verb-specific frame that sufficiently matches the partial frame to be processed. Therefore, the only reliable knowledge source is a matching construction (if such a construction exists). This property of the model embodies the interaction of entrenchment and statistical preemption in construction use suggested by Goldberg (2006).

A question that arises here is, what are the limitations of using a verb in a novel construction? Although many of the innovative uses of verbs in unusual constructions are acceptable, many others are not. The distinction seems to come from the fundamental semantic properties of the verb, and their compatibility with the scene to be described. For example, one can say *The child hammered the putty flat* but not **The child played the game finished*, because *hammer* itself encodes a potential change of state to the object, whereas *play* does not. This is a distinction that our model cannot make in its current form. Addressing this problem in our model will require a more sophisticated, fine-grained semantic representation for verbs and scenes, augmented with appropriate semantic constraints.

3.8 Summary of Contributions

We have used the Bayesian model proposed in the previous chapter for simulating the early stages of verb argument structure acquisition in children. Our experimental results

demonstrate the emergence of the abstract argument structure constructions from verb usages, and the transformation of the properties of constructions from verb-specific to more general. In using the acquired knowledge of verbs and constructions in a variety of language tasks, we have shown that the general patterns observed in experimental child data can be detected in the behaviour of the model.

While the model reflects the verb-based nature of argument structure acquisition, a main advantage of our model is its explicit formulation of the interaction between verb-based and construction-based knowledge in comprehension and production, which contributes to the stages of learning exhibited by children. The suggested Bayesian mechanism not only accounts for the correct use of verbs in their appropriate argument structures and the gradual cessation of overgeneralization errors, it also shows the possibility of productive generalization of constructions in the adult language facility. Many factors have been suggested in the linguistic and psycholinguistic literature for explaining these effects, such as entrenchment, competition, cue construction, and statistical preemption. Our proposed prediction model of language use unifies these factors, demonstrating their smooth integration in the processes of generalization and of recovery from overgeneralization.

Despite the success of the model in accounting for observed behaviours over the time-course of child argument structure acquisition, a number of limitations must be addressed in future work. For example, while we have not built in the knowledge of possible constructions, or innate principles of syntax-semantics mappings, we do assume that the child can figure out the appropriate participant roles for observed entities, drawing on a predefined set of possible roles. In the next chapter, we directly address the learning and use of semantic roles.

As noted earlier, we also assume that the model has learned the mapping between each word and its meaning (i.e., semantic symbol) before argument structure acquisition begins. This is an unrealistic assumption since the two processes are more likely to

be interleaved in learning. Extensions to the model will require integrating a word-to-meaning mapping algorithm, similar to that of Siskind (1996) or Fleischman and Roy (2005). This approach may allow the model to use its acquired argument structure knowledge to facilitate learning the meaning of new words, which is a natural extension of the syntactic bootstrapping behaviour demonstrated here.

Currently, our model does not record the actual words that participate in the acquired argument structure frames, instead replacing words by their semantic categories. These categories may later be generalized to their superordinates in the semantic hierarchy, thus enabling the acquisition of general patterns of argument types from instances of verb usages. While this approach supports learning of general constructions, it entails that constructions that depend on specific word forms cannot be acquired. This includes constructions in the adult repertoire such as "to not $verb_1$ let alone $verb_2$", which depends on the words "let alone." Moreover, some researchers have hypothesized that children's early constructions may depend on specific words or morphemes, as in *I'm Ving it* or *It Ved* (Lieven et al., 1998; Childers and Tomasello, 2001).

Chapter 4

Learning and Use of Verb Semantic Roles

4.1 Overview

Semantic roles such as Agent, Theme, and Recipient in (1) and (2) below, are a critical aspect of linguistic knowledge because they indicate the relations of the participants in an event to the main predicate.

(1) Mom_{Agent} gave this_{Theme} to $\text{her}_{Recipient}$.

(2) Mom_{Agent} gave $\text{her}_{Recipient}$ this_{Theme}.

Moreover, it is known that people use the associations between roles and their syntactic positions to help guide on-line interpretation (e.g., Trueswell et al., 1994). How children acquire this kind of complex relational knowledge, which links predicate-argument structure to syntactic expression, is still not well understood. Fundamental questions remain concerning how semantic roles are learned, and how associations are established between roles and the grammatical positions the role-bearing arguments appear in.

We use our Bayesian model to investigate the problem of semantic role learnability. A shortcoming of the implemented version in Chapter 3 is that roles are explicit labels, such

as Agent, which were assumed to be "perceptible" to the child from the scene. In this chapter, we have extended our model to directly address the learning and use of semantic roles. Here, we draw on the semantic properties of the arguments that appear in different verb usages to gradually learn the general semantic roles. Our Bayesian model associates each argument of a predicate with a *semantic profile*, which is a probability distribution over a set of semantic properties that each argument can take. We show that, initially, the semantic profiles of an argument position yield verb-specific conceptualizations of the role associated with that position. As the model is exposed to more input, these verb-based roles gradually transform into more abstract representations that reflect the general properties of arguments across the observed verbs. We show that intuitive profiles can be learned for each role from examples of its use. We further establish that such representations can be useful in guiding the argument interpretation of ambiguous input, an ability experimentally demonstrated in children in recent work (Nation et al., 2003).

Section 4.2 discusses the linguistic theories of semantic role learning, and lays out the major questions concerning the nature of semantic roles and their learnability. The existing computational models of semantic roles are reviewed in Section 4.3. In Section 4.4 we put forward an adaptation of our Bayesian model for the representation and learning of semantic roles. We describe our experimental set-up in Section 4.5, and present the experimental results in Section 4.6. Section 4.7 presents a general discussion about the experimental results, and relates them to the important questions on the learnability of semantic roles. Finally, we summarize our contributions in Section 4.8.

4.2 Theories of Semantic Role Learning

The notion of thematic roles was first introduced by semanticists as the relationship between a predicate and its arguments (Fillmore, 1968; Jackendoff, 1972). However, this notion was extensively used by syntacticians as a theoretical device to explain argu-

ment indexing (i.e., linking grammatical relations to semantic structure) and grammatical generalization (Chomsky, 1981; Pinker, 1984). In many theories of syntax such as Government and Binding Theory (Chomsky, 1981) and Lexical-Functional Grammar (Falk, 2001; Bresnan, 2001), thematic roles are believed to be discrete, limited in number, and universal. The mapping between roles and the sentence structure are defined through a set of universal "linking rules". These rules are argued to be innate, and to help children in learning the syntax of their language. A strong version of these rules suggests that the mapping of a thematic role to a particular grammatical function is rigid (Pinker, 1984; Baker, 1988). A weaker position proposes that thematic roles and syntactic positions are matched by means of a hierarchy, such that the highest-ranked thematic role occupies the highest-ranked syntactic position (i.e., the subject), and one works one's way down the two hierarchies in parallel until one runs out of arguments (Grimshaw, 1990; VanValin and LaPolla, 1997; Jackendoff, 1990).

Many researchers have proposed lists of thematic roles, mostly different in size and granularity. But there is little consensus about the "correct" set of thematic roles. Even the familiar roles are subject to disagreement (for example, whether Theme is the same role as Patient or not). That is mainly due to the fact that, in order for the universal linking rules to be useful, it should be possible to assign each argument of every verb in the language to one and only one thematic role. That is, what a verb semantically entails about each of its arguments must permit us to assign the argument, clearly and definitely, to some role or other, and what the meaning entails about every argument must always be distinct enough that two arguments clearly do not fall under the same role definition. However, it seems that there is no cross-linguistically consistent definition of thematic roles that satisfies this criteria.

Dowty (1991) proposes a different theory of thematic roles: the Proto-Role Hypothesis. According to his theory, thematic roles draw from a pool of more basic semantic properties such as sentience, volition, and movement. No one thematic role necessarily

has all of these properties, and some have more than others. The proposed proto-roles are based on analysis of linguistic data. However, Dowty (1991) does not give any explicit account on whether these proto-roles are innate or children actually learn them. Kako (2006) provides an experimental test of the Proto-Roles Hypothesis, and shows that human subjects assign the suggested proto-role properties to grammatical roles such as subject and object, even for a novel verb (or a familiar verb in an unusual construction).

Although several experimental studies have been performed to explore the role of thematic roles in language processing (see, for example, Trueswell et al., 1994), there is little agreement on what the nature of thematic roles is. However, McRae et al.'s (1997) experiments on human subjects' ranking of role/filler featural similarity for Agent and Patient roles, as well as their ambiguity resolution in the presence of featural bias, suggests that thematic roles might best be viewed as verb-specific concepts. Based on such results, pioneers of the usage-based framework for language acquisition and use have suggested that children do not have access to a pre-defined set of thematic roles or proto-role properties. Instead, children learn thematic roles gradually and from the input they receive, through a process of categorization and generalization (e.g., Lieven et al., 1997; Tomasello, 2000a). For instance, Tomasello (2000a) claims that, initially, there are no general labels such as Agent and Theme, but rather verb-specific concepts such as 'hitter' and 'hittee,' or 'sitter' and 'thing sat upon.' It remains unexplained, though, precisely how verb-specific roles metamorphose to general semantic roles.

There are few experiments on how children learn general semantic roles. However, Shayan and Gershkoff-Stowe (2007) show that children indeed demonstrate a pattern of gradually learning thematic roles, and that both age and having access to linguistic cues affect the learning process. In their experiments, 3- and 5- year old children's knowledge of Agent and Patient roles is tested in simple transitive events with two participants. It is shown that older children have a better understanding of the relational similarity across Agents and Patients in animated scenes presented side-by-side on a computer screen. It

is also shown that providing linguistic cues (i.e., using a novel verb in a transitive frame to describe the scenes) makes it easier for both age groups to grasp relational similarity across semantic roles. Moreover, experiments with children have revealed the use of verb-specific biases in argument interpretation (Nation et al., 2003), as well as of strong associations between general roles and syntactic positions (e.g., Fisher, 1996, and related work). However, the how and when of the emergence and learning of thematic roles is yet to be explored through further psycholinguistic experiments and computational modeling.

4.3 Related Computational Models

A number of computational models learn verb-specific roles that are not generalized. Chang (2004) presents a model that learns associations between form relations (typically word order) and meaning relations (typically role-filler bindings) from input data, and uses them in language comprehension. However, the model does not make any generalizations beyond the scope of one lexical entry, and therefore does not learn a conception for a general semantic role. A number of computational approaches for learning the selectional preferences of a verb first initialize WordNet concepts with their frequency of use as the particular argument of a verb, and then find the appropriate level in the WordNet hierarchy for capturing the verb's restrictions on that argument (e.g., Resnik, 1996; Clark and Weir, 2002). However, none of these models generalize their acquired verb-based knowledge to a higher level, yielding constraints on the arguments of general constructions such as the transitive or intransitive. We will examine these models in more detail in the next chapter.

Many computational systems model human learning of the assignment of general predefined roles to sentence constituents, using a multi-feature representation of the semantic properties of arguments. McClelland and Kawamoto (1986) present a connectionist model

for assigning roles to constituents of sentences, simply by using the surface structure of the sentence as the input. Words are represented as vectors of semantic features (*human, softness, gender,* etc., for nouns, and *cause, touch,* etc., for verbs). The output of the model is the assignment of a limited number of fixed thematic roles such as Agent and Instrument to the arguments of a verb. The model can also guess certain properties (semantic features) of a missing argument in a sentence, but the roles themselves are not learned. Morris et al. (2000) also present a connectionist model that assigns nouns (mostly proper names) to the appropriate roles for a number of sentence structures. Again, the model is trained on sentences annotated with the correct role assignment for their arguments, where a limited number of pre-defined roles (Agent, Indirect Patient, etc.) are used. Allen (1997) presents yet another connectionist model for thematic role assignment which simulates the integration of syntactic, semantic and lexical information. A distributed representation of nouns is used to allow for generalization over semantic features, and arguments are labeled with the appropriate role from a set of traditional thematic roles. Allen's model treats the representation of thematic roles differently in that each role is further elaborated by additional proto-role units. The network can distinguish the grammatical and ungrammatical utterances, and predict the semantic features of novel verbs and the semantic roles (as well as the proto-role properties) for each argument in the input. However, the explicit labeling of the arguments is critical to the model, and it is not clear whether the model can learn these roles only based on the semantic properties of the arguments and the set of proto-role properties specified in the training data.

Others have looked at learning the appropriate cues of the language for role assignment, such as word order, noun animacy, and case inflection. It is an interesting problem because such cues might or might not be present in every sentence, and one cue may conflict with another cue as to the correct role assignment. Models proposed by Bates and MacWhinney (1989) and Matessa and Anderson (2000) can both learn the cue dom-

inance hierarchy of the language. The Competition Model of Bates and MacWhinney (1989) uses a learning-on-error mechanism, where a strength counter is maintained for each cue, and in deciding a role, the noun with the largest total cue strength is assigned to that role. The ACT-R Model of Matessa and Anderson (2000) uses both declarative and procedural rules for role assignment. The ACT-R model better predicts the roles assigned by human subjects, but neither of the models learn an explicit profile for each role.

In the computational linguistics literature, semantic role labeling of text corpora has been well studied. Many supervised models (e.g., Gildea and Jurafsky, 2002; Hacioglu et al., 2004) have been proposed which use manually annotated corpora of PropBank (Palmer et al., 2005) or FrameNet (Baker et al., 1998) as training data. The unsupervised methods either use bootstrapping to learn frames for verbs (e.g., Riloff and Schmelzenbach, 1998), or a verb lexicon such as VerbNet (Kipper et al., 2000) and a probability model to estimate the best role tag for each sentence constituent (Swier and Stevenson, 2004).

As in our previous work on argument structure acquisition, all of these computational models require explicit labelling of the arguments that receive the same role in order to learn the association of the roles to semantic properties and/or syntactic positions. In this chapter, we show that our extended model can learn general semantic profiles of arguments, as well as their association with grammatical positions, without the need for role-annotated training data, or even a list of known roles to draw from.

4.4 A Computational Model for Learning Verb Semantic Roles

We use our Bayesian framework, presented in Chapter 2, for modeling learning and use of the semantic roles in children. As in the implementation presented in Chapter

3, our model learns the argument structure frames for each verb, and their grouping across verbs into constructions. However, in contrast with our model of verb argument structure acquisition, we no longer assume that the child can perceive the "correct" semantic role for each argument upon watching an event. Instead, we assume that the child can infer certain semantic properties for the arguments. As the model processes the input, it associates each argument of a predicate with a semantic profile, which is a probability distribution over a set of such semantic properties. Moreover, the model forms probabilistic associations between the semantic properties of the arguments, their syntactic positions, and the semantic primitives of the predicate. These associations are generalized (through the constructions) to form more abstract notions of role semantics, dependent on argument position and verb primitives.

In Section 4.4.1, we describe the basic properties regarding the input to the model, and the frame structure and features used in this variation of the model. In Section 4.4.2, the measure of similarity between two arguments based on their semantic properties is explained. Finally, in Section 4.4.3, a formal definition of a semantic profile is given.

4.4.1 The Input and the Argument Structure Frames

The basic properties of the input are very similar to those described in Section 3.4.1: we use a set of scene-utterance pairs as input, where the meaning of the scene is represented by a simple logical structure. The only difference is that, instead of representing each word as a semantic category in a hierarchical ontology, we keep track of the actual words that appear in an utterance. The lexical entry for each word contains its semantic properties, which are extracted from WordNet (version 2.0) as follows. We hand-pick the intended sense of the word, extract all the hypernyms (ancestors) for that sense, and add all the words in the hypernym synsets to the list of the semantic properties. Fig. 4.1 shows an example of the hypernyms for *cake*, and its resulting set of semantic

cake →baked goods →food →solid →substance, matter →entity
cake: {baked goods, food, solid, substance, matter, entity}

Figure 4.1: Semantic properties for *cake* from WordNet

properties.[1] These properties are later used to induce a semantic profile for each argument position in a construction. Ideally, such profile should contain lexical properties such as *animate* and *solid* as well as event-specific properties such as *cause to move* or *affected*. However, automatic acquiring of the event-based properties for each scene is not feasible. Therefore, such properties must be added manually, which is labour-intensive. For now, we only use the WordNet-based lexical properties for each argument.

As before, the model extracts an argument structure frame from each input pair. Each frame records the head word (verb or preposition), the number of arguments, the semantic primitives of the predicate, the syntactic pattern of an utterance, and the semantic properties of each of the arguments. The extracted frame is stored in the lexical entry of the head word, and its frequency count is set to one. In contrast to the previous implemented version (see §3.4.1), similar frames are not merged together. Instead, each frame is kept separately in the lexical entry for the head predicate of that frame. An extracted frame is input to our Bayesian learner which forms groupings of frames into constructions. Fig. 4.2 shows a sample input pair, and the corresponding extracted frame.

[1]We do not remove alternate spellings of a term in WordNet; this will be seen in the profiles in the results section.

Scene-Utterance Input Pair	
Scene: $\mathbf{Make}_{[cause,become]}(\mathsf{TIM}, \mathsf{CAKE})$	
Utterance: *Tim made cake*	
Extracted Frame	
head verb	*make*
semantic primitives of verb	[*cause, become*]
number of arguments	2
semantic properties of argument 1	{ *man, male, person, individual, ...* }
semantic properties of argument 2	{ *baked goods, food, solid, substance, ...* }
syntactic pattern	arg1 *verb* arg2

Figure 4.2: An input pair and its corresponding frame.

4.4.2 Probabilities of Semantic Properties

As explained in Section 2.3.2 of Chapter 2, in both learning and prediction, the probability of value j for feature i in construction k is estimated using a (smoothed) maximum likelihood formula (this is equation (2.6) from page 24):

$$P_i(j|k) = \frac{\mathbf{count}_i^k(j) + \lambda}{n_k + \lambda\alpha_i} \tag{4.1}$$

where n_k is the number of frames participating in construction k, $\mathbf{count}_i^k(j)$ is the number of those with value j for feature i, and λ and α_i are smoothing factors.

For most features, $\mathbf{count}_i^k(j)$ is calculated by simply counting those members of construction k whose value for feature i exactly matches j. However, for the semantic properties of words, counting only the number of exact matches between the sets is too strict, since even highly similar words very rarely have the exact same set of properties. We instead compare the set of semantic properties of a particular argument in the observed frame, S_1, and the set of semantic properties of the same argument in a member

frame of a construction, S_2, using the Jaccard similarity score:[2]

$$\textbf{sem_score}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \tag{4.2}$$

For example, assume that the new frame represents the verb usage *John ate cake*, and one of the members of the construction that we are considering represents *Mom got water*. We must compare the semantic properties of the corresponding arguments *cake* and *water*:

cake: {baked goods, food, solid, substance, matter, entity}

water: {liquid, fluid, food, nutrient, substance, matter, entity}

The intersection of the two sets is {food, substance, matter, entity}, and the union is {baked goods, liquid, fluid, food, solid, nutrient, substance, matter, entity}, therefore the **sem_score** for these sets is $\frac{4}{9}$.

In general, to calculate the conditional probability for the set of semantic properties, we set $\textbf{count}_i^k(j)$ in equation (4.1) to the sum of the **sem_score**'s for the new frame and every member of construction k, and normalize the resulting probability over all possible sets of semantic properties in our lexicon:

$$P_{properties}(S_j|k) = \frac{\sum_{f \in k} \textbf{sem_score}(S_j, S_f) + \lambda|S_j|}{\sum_{l \in \text{Lexicon}} \sum_{f \in k} \textbf{sem_score}(S_l, S_f) + \lambda\alpha_{prop}} \tag{4.3}$$

where S_j is a set of semantic properties, S_f represents the semantic properties of an argument in a frame f that belongs to construction k, S_l represents the semantic properties of a lexeme l, $|S_j|$ is the size of the set S_j, and λ and α_{prop} are the smoothing factors (as in equation (2.6) on page 24). Intuitively, the sum factor in the numerator represents how much the meaning of the new set S_j resembles that of frames grouped in construction k.

[2]The selected semantic properties and the corresponding similarity score are not fundamental to the model, and could in the future be replaced with an approach that is deemed more appropriate to child language acquisition.

The double sum factor in the denominator, on the other hand, represents how distinct the semantic properties of the frames in construction k are from all the existing words in the language (since each word in the lexicon is compared against all the words in the construction). The latter factor is computationally expensive, and we use a pre-calculated similarity matrix to estimate it. However, we can assume that humans have an idea of how general or distinct the meanings associated with a lexeme or a construction are, and do not have to perform such extensive calculations every time they process a new input.

4.4.3 Representation of Semantic Profiles

The above probability, $P_{properties}(S_j|k)$, is the probability of a particular set of semantic properties as a particular argument of a construction. But in generally describing the properties of an argument, we want to use a semantic profile, as noted above, that instead tells us the probability of every semantic property (individually) occurring in that argument position. This requires looking at the probability of all the individual properties, j_p, rather than the probability of the full set, j. We use a modified version of equation (4.1) in which $\mathbf{count}_i^k(j_p)$ is the number of frames in construction k that include property j_p for the argument whose set of semantic properties is the i^{th} feature of the frame:

$$P_i(j_p|k) = \frac{\mathbf{count}_i^k(j_p) + \lambda}{n_k + \lambda\alpha_i} \tag{4.4}$$

A vector of the resulting probabilities over all j_p form the semantic profile of that argument.

A semantic profile contains all the semantic properties ever observed in an argument position. As learning proceeds, a profile may include a large number of properties with very low probability. In order to display the profiles we obtain in our results section below, we create truncated profiles which list the properties with the highest probabilities, in

decreasing order of probability value. To avoid an arbitrary threshold, we cut the ordered list of properties at the widest gap between two consecutive probabilities across the entire list.

4.5 Experimental Set-Up

Through a number of computational experiments, we demonstrate that our model can learn intuitive profiles for general semantic roles. We also show that the model can use the learned profiles in language processing tasks. We describe the creation of the input corpora for our simulations in Section 4.5.1, and the setting of the parameters of the model in Section 4.5.2.

4.5.1 The Input Corpora

The corpora for our experimental simulations are generated from an extended version of the input-generation lexicon used in our earlier work. As described in Section 3.5.2, we recorded the frequencies of a small set of verbs and their possible argument structure frames from a portion of the CHILDES database (MacWhinney, 1995) in the input-generation lexicon. However, in the input data used in the previous chapter, the nouns and prepositions in the lexicon were chosen arbitrarily. Here, we randomly selected 100 sentences containing each verb; from these, we extracted a list of head nouns and prepositions that appear in each argument position of each frame, and added these to the lexicon.

For each simulation in our set of experiments, an input corpus of scene-utterance pairs is automatically randomly generated using the frequencies in the input-generation lexicon to determine the probability of selecting a particular verb and argument structure. Arguments of verbs are also probabilistically generated based on the word usage information for the selected frame of the verb. The resulting corpora are further processed to

simulate noise and incomplete data, as described in Section 3.5.2.

4.5.2 Setting the Parameters of the Model

For the purpose of implementing the model and running simulations, we need to assign values to parameters α_i for each feature i in equation (2.6) (page 24 in Chapter 2). In the model presented in this chapter, there are five α_i parameters that need to be set, each of which correspond to a different feature in a frame: the head word, the semantic primitives of the head word, number of arguments, the semantic properties of the arguments, and syntactic pattern. We assign these parameters in the same way described in Section 3.5.3. The only new feature here is the semantic properties of the argument, or α_{prop}, introduced in equation (4.3). We set this value to the total number of (distinct) semantic properties extracted from WordNet for all the words in our initial lexicon, which is 493. We also set the parameter λ the same way as described in Section 3.5.3.

4.6 Experimental Results

In this section, we report experimental results from a number of simulations of our model. In Section 4.6.1, we look at some of the semantic profiles that our model acquires for argument positions in various constructions, which are traditionally associated with general thematic roles such as Agent, Theme, and Destination. We show that the generated profiles match the intuitive semantic properties of these roles. In Section 4.6.2 we show that our model can learn multiple semantic roles for a single argument position such as subject or object, depending on the semantic primitives of the verb that participates in the syntactic pattern. In Section 4.6.3, we demonstrate the process of role generalization, where the model starts from learning verb-specific profiles for a particular argument position in a construction, and later moves to a more general profile for the same position as a result of exposure to more input. Finally, in Section 4.6.4, we examine how the

acquired semantic profiles can be used in language tasks such as sentence comprehension and ambiguity resolution. Appendix B shows the constructions learned by the model in one of simulations, after processing 200 input pairs.

4.6.1 Formation of Semantic Profiles for Roles

Psycholinguistic experiments have shown that children are sensitive at an early age to the association between grammatical positions, such as subject and object, and the properties of the roles that are typically associated with those positions (e.g., Fisher, 1996). Here we show that for each argument position in a construction, our model learns a general semantic profile from instances of verb usage. For some common constructions, we study the semantic profile of the arguments by predicting the most probable semantic properties for that position, as detailed above. Although these semantic profiles do not include any event-specific knowledge, they can be considered as a weak form of the semantic roles that the corresponding arguments receive in that construction.

We train our model on 200 randomly generated input pairs,[3] and then present it with a test input pair containing a novel verb *gorp* appearing in a familiar construction, with unknown nouns appearing as its arguments. As an example, a test pair for a novel verb appearing in a transitive construction looks as follows:

$\mathbf{Gorp}_{[cause,become]}(\mathsf{X},\mathsf{Y})$

x gorp y

We then have the prediction model produce a semantic profile for each of the unknown arguments, to reveal what the model has learned about the likely semantic properties for that position in the corresponding construction. We average the obtained probabilities over 5 simulations on different random input corpora.

Our model learns semantic profiles for argument positions in a range of constructions.

[3] In most experiments, receiving additional input after 200 pairs ceases to make any substantial difference in the output.

Transitive Subject	*Transitive Object*
entity	entity
object	object
physical object	physical object
being	artifact
organism	artefact
living thing	whole
animate thing	whole thing
causal agent	unit
cause	
causal agency	
person	
individual	
someone	
somebody	
mortal	
human	
soul	

Figure 4.3: Semantic profiles of argument positions *Agent* and *Theme*.

Destination
location
destination
goal
entity
point
residence
abode
address
geographic point
geographical point
top

State
quality
abstraction
attribute
bad
badness
cleanness

Figure 4.4: Semantic profiles for the semantic roles *Destination* and *State*.

Fig. 4.3 shows the top portion of the predicted semantic profiles for the arguments in the subject and object positions of a transitive construction (corresponding to X and Y in the *gorp* test input above).[4] Even though we use semantic properties from WordNet, which lack any event-specific features, the emerging semantic profile for each argument position demonstrates the intuitive properties that the argument receiving that role should possess. For example, the semantic profile for an argument that appears in the subject position in a transitive construction (the left box of Fig. 4.3) demonstrates the properties of an animate entity, most likely a human. In contrast, the semantic profile for an argument in the object position (the right box of Fig. 4.3) most likely corresponds to a physical entity. These results are consistent with Kako (2006), who finds that, given unknown verbs and nouns in the transitive sentence, adults attribute more Agent-like

[4]We remove the probability values associated with each semantic property, since the values are very small and close to each other.

Intransitive Subject
entity
object
physical object
being
organism
living thing
animate thing
causal agent
cause
causal agency
person
individual
someone
somebody
mortal
human
soul

Figure 4.5: Semantic profile of the subject position in an intransitive construction.

semantic properties to subjects, and more Theme-like properties to objects.

Similarly, Fig. 4.4 demonstrates the predicted semantic profiles for the traditional roles *Destination* and *State*, as in *Joe went home*$_{\text{Destination}}$ and *She makes me angry*$_{\text{State}}$. Again, the profiles show intuitively plausible properties for each role.

Fig. 4.5 shows the semantic profile for the subject of an intransitive. This argument can receive an Agent role (*John went*) or a Theme role (*the ball fell*). We think that the semantic profile represents more Agent-like characteristics because, in our input data, the probability of an Agent in that position is much higher than a Theme. We explore this issue next.

4.6.2 Multiple Possible Roles for a Position

Our model's failure to distinguish different roles assigned to the intransitive subject position might simply be a consequence of the bias of the input corpora. Alternatively, it might be due to an inherent deficiency of the model when faced with input that lacks explicit role labels—i.e., an inability of the model to distinguish the arguments of different types of verbs when those arguments occur in the same syntactic position.

To test this, we created an input corpus with an artificially increased frequency of *fall*, the only intransitive verb in our lexicon that can have a Theme (rather than an Agent) in the subject position, so that the model would be given sufficient examples of such a usage. We then tested the model with two kinds of novel verbs: one with semantic primitives [*act,move*] (associated with agentive intransitives like *come* and *go*), and one with semantic primitives [*move,direction,vertical*] (like the Theme-assigning verb *fall*). The predicted profiles are shown in Fig. 4.6. In response to the former (*go*-type) input, the model still predicts a semantic profile very similar to the one shown in the left box of Fig. 4.3. In contrast, for the latter (*fall*-type) input, the predicted semantic profile shows a combination of that of the transitive subject, shown in the left box of Fig. 4.3, and the transitive object, shown in the right box of Fig. 4.3.

Go-type input
entity
object
physical object
being
organism
living thing
animate thing
person
individual
someone
somebody
mortal
human
soul
causal agent
cause
causal agency

Fall-type input
entity
object
physical object
being
organism
living thing
animate thing
person
individual
someone
somebody
mortal
human
soul
causal agent
cause
causal agency
artifact
artefact
whole
whole thing
unit

Figure 4.6: The predicted semantic profile for the subject position of an intransitive usage. The left profile is predicted for a usage with semantic primitives similar to verb *go*, whereas the right profile is for a usage with semantic primitives similar to verb *fall*.

This experiment is crucial in showing that the model does not simply associate a single semantic profile with a particular argument position. If this were the case, the model would never be able to distinguish, e.g., transitive verbs that assign Agent and Theme, from those that assign Experiencer and Stimulus. This experiment demonstrates that the model forms a complex association among a syntactic pattern, an argument position, and the semantic primitives of the verb, allowing it to make a distinction between different roles assigned to the same position in the same syntactic pattern.

4.6.3 Verb-Based vs. General Semantic Profiles

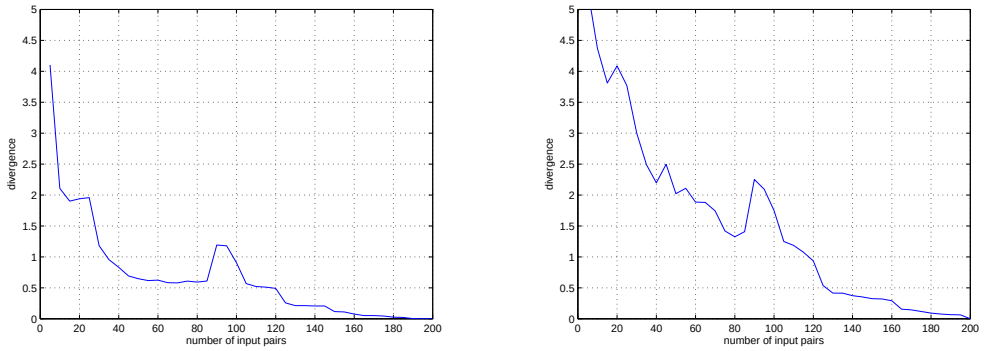
In the previous section, we showed that our model can learn general conceptions of semantic roles as a result of processing enough input. However, because the model learns these general profiles from instances of verb usage, we expect these profiles to go through a gradual generalization process, where they initially reflect the properties of specific verb arguments, and become more general over time. In this section, we investigate this hypothesis.

We tracked the generalization process for each semantic profile, to see how it moves from a verb-based profile to a more general one. Fig. 4.7 (left box) shows the semantic profile for the argument in the object position right after the first transitive usage. In this particular simulation, the first transitive verb in the corpus is *eat*, and its second argument in that usage is *pie*. The semantic profile thus reflects the properties of a pie, and not the general properties of that argument position. The profile becomes more general after processing 50 and 100 input pairs, shown in the middle and right boxes of Fig. 4.7, respectively. (Recall that Fig. 4.3 shows a profile for transitive object after 200 inputs, and averaged over 5 simulations.)

To observe the trend of moving from a more specific to a more general semantic profile for each argument position, we used relative entropy to measure the divergence between the semantic profile for an argument position at a given point in learning, and

After 5 pairs	*After 50 pairs*	*After 100 pairs*
entity	entity	entity
substance	object	object
matter	physical object	physical object
food	artifact	being
solid	artefact	organism
baked goods	whole	living thing
pastry	whole thing	animate thing
	unit	quality
	instrumentality	artifact
	instrumentation	artefact
	quality	whole
	equipment	whole thing
	electronic equipment	unit
	receiver	animal
	receiving system	animate being
	.	.
	. (additional	. (additional
	. properties)	. properties)

Figure 4.7: The evolution of the transitive object role.



Transitive Subject Transitive Object

Figure 4.8: Learning curves for semantic profiles. The x-axis is time (#inputs), and the y-axis is divergence from the profile that the model eventually converges to.

the profile for that position that the model eventually converges to at the end of each simulation.[5] We measured the profile divergence for subject and object positions of a transitive construction after every 5 input pairs over a total of 200 pairs, averaged over 5 simulations. Fig. 4.8 shows that the profile for the subject position (i.e., the Agent) is learned faster than the profile for the object position (i.e., the Theme), which is a much less constrained role. The curves show that the model stabilizes on the final profiles at around 150 input pairs, when receiving more inputs ceases to make any substantial difference in the profiles.

4.6.4 Using Semantic Profiles in Comprehension

Semantic roles are helpful in on-line ambiguity resolution, by guiding adults to the interpretation that best matches the role expectations of a verb for a given position (e.g., Trueswell et al., 1994). Nation et al. (2003) have shown that young children also draw on verb-specific biases in on-line argument interpretation.[6] Here we demonstrate the ability

[5] $RelativeEntropy(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$, where P and Q are probability distributions.

[6] This work complements that of Fisher (1996) and others, who demonstrate that children use associations between argument properties and syntactic positions to choose the interpretation of an unknown verb in a full sentence.

of our model to use its acquired semantic profiles to predict the best interpretation of an ambiguous partial input.

We consider cases using the verb *give*, in which an utterance beginning *you give* ⟨*noun*⟩ can continue with either a second object (*you give* ⟨*noun*⟩ *something*) or a prepositional phrase (*you give* ⟨*noun*⟩ *to someone*). In the first case, ⟨*noun*⟩ is the Recipient of the action, and in the second case, it is the Theme (the thing given). We vary ⟨*noun*⟩ to be either *her*, which is a likely Recipient, or *this*, which is a likely Theme. We then observe in each case which interpretation our model prefers.

We set up the experiment as one where we compare the probability of the following two input pairs:

her as Recipient	$\mathbf{Give}_{[cause,possess]}(\mathsf{YOU}, \mathsf{X}, \mathsf{HER})$ *you give her x*
her as Theme	$\mathbf{Give}_{[cause,possess]}(\mathsf{YOU}, \mathsf{HER}, \mathsf{X})$ *you give her x*

We give each of these inputs to the model, have it extract the corresponding frame F, and then have it calculate

$$\mathbf{match_score}(F) = \max_k P(k)P(F|k) \tag{4.5}$$

where $P(k)$ and $P(F|k)$ are from the learning model (equation (2.3) on page 23 and equation (2.5) on page 24, respectively). The measure **match_score** corresponds to how well the model thinks the given input pair matches an existing construction. Having a higher **match_score** for the first pair means that the model has recognized *her* as the Recipient, while a higher score for the second pair means that the model has recognized *her* as the Theme.

We also compared analogous inputs pairs, using *this* instead of *her*. Since the only difference in these two sets of input is the use of *her* vs. *this*, differences in interpretation

Table 4.1: log(**match_score**) for Recipient-first and Theme-first frames after processing 20 and 200 input pairs. The score of the best match (highest score) is shown in bold.

	20 Pairs		200 Pairs	
Utterance	Recipient	Theme	Recipient	Theme
You gave her x	−23.08	**−20.47**	**−15.06**	−19.89
You gave this x	−24.23	**−20.78**	−16.68	**−16.55**

of the first object noun (as Recipient or Theme) depend only on the match of the noun's semantic properties to the semantic profiles of that argument position.

Table 4.1 shows the results after processing 20 and 200 input pairs, averaged over 5 simulations; a higher number (lower absolute value) indicates the preferred interpretation. After processing 20 pairs, the model displays a strong preference towards treating the first object as the Theme, for both *her* and *this*. This occurs because *give* is the only verb in our lexicon that appears in a ditransitive (Recipient-first) construction, whereas many high frequency verbs (e.g., *put* and *get*) appear in the competing (Theme-first) prepositional construction. Thus, at the early stages of learning, the ditransitive construction is relatively weak. However, after processing 200 input pairs, the model shows a preference for the "correct" interpretation in both cases. (The difference between the two frames for *you gave this x* is small, but consistently indicates a Theme preference across all simulations.)

The model is thus able to use its learned associations between semantic properties and argument positions to appropriately guide interpretation of an ambiguity. These results predict that very early on, children (like our model) would experience some difficulty in this type of task, when drawing on the knowledge of a less commonly observed construction.

4.7 Discussion of the Results

4.7.1 Where Do Abstract Roles Come From?

As reported in Section 4.6.1, the emerging semantic profiles predicted by our model for the argument positions in different constructions demonstrate the intuitive properties that the argument receiving that role should possess. For example, the semantic profile for an argument that appears in the subject position in a transitive construction demonstrates the properties of an animate entity. In contrast, the semantic profile for an argument in the object position most likely corresponds to a physical entity. As mentioned before, these results are consistent with Kako (2006), who finds that, given unknown verbs and nouns in a transitive pattern, adults attribute more Agent-like semantic properties to subjects, and more Patient-like properties to objects (specifically, a variation on Dowty's (1991) proto-properties).

Kako (2006) also reports similar results using a *known* verb in an incompatible construction. He used a set of verbs that can only be used in an intransitive construction (such as *appear, arrive,* etc.) and placed them in a transitive frame. Human subjects were asked whether the subject and object have certain proto-role properties. The results show that, even for such unusual usages, subjects are rated as more Agent-like than objects by participants, while objects are rated more Patient-like than subjects. Kako argues that, because (unlike his previous experiments) the verbs used in this experiments are known to participants, one cannot assume that they assign proto-role properties for each argument position by "borrowing" another familiar transitive verb, and therefore the thematic properties must be inferred from grammatical roles alone. However, no clear claim is being made on how these abstract properties are learned for each role.

To simulate this, we gave our model a transitive input like the *gorp* pair above, but with an intransitive-only verb *dance*:

$\textbf{Dance}_{[act,manner,cause]}(\mathsf{X},\mathsf{Y})$

x dance y

We then had the prediction model produce a semantic profile for each of the unknown arguments. The predicted profiles for both subject and object positions are shown in Figures 4.9 and 4.10 (the predicted probabilities for each property are shown in parentheses). We found a similar profile for the nouns in both grammatical positions as in a transitive usage of a novel verb, as displayed in Fig. 4.3. Since *dance* has never been seen with an object, this profile must come from the model's learned associations over existing verbs in the transitive construction.

4.7.2 Age-Dependence of Semantic Roles

Tomasello (2000a) (among others) has proposed that children initially learn verb-specific roles such as 'hitter' and 'hittee,' and only later move to more general roles. Moreover, Shayan and Gershkoff-Stowe (2007) show that general notions like Agent and Patient develop over time: children indeed demonstrate a pattern of gradually learning thematic roles, and both age and having access to linguistic cues affect the learning process. It is shown that older children have a better understanding of the relational similarity across Agents and Patients in animated scenes presented side-by-side on a computer screen.

In spite of these suggestions on the learnability of thematic roles, there have been few attempts to give a detailed account of how general roles can be learned from individual verb usages. Our model illustrates how such role generalization might come about. Although the semantic profiles of our model reflect the general properties that a particular role-bearing argument must have, they are formed from input that contains only argument properties for specific verb usages. The generalization occurs as more and more semantic properties are associated with an argument position, and only the most general properties are seen sufficiently frequently to have high probability. The effect of age (or exposure to more input) in forming stable profiles for semantic roles can be seen

Transitive Subject			
gorp		*dance*	
entity	(0.056)	entity	(0.057)
object	(0.055)	object	(0.056)
physical object	(0.055)	physical object	(0.056)
being	(0.053)	being	(0.053)
organism	(0.053)	organism	(0.053)
living thing	(0.053)	living thing	(0.053)
animate thing	(0.053)	animate thing	(0.053)
causal agent	(0.053)	causal agent	(0.053)
cause	(0.053)	cause	(0.053)
causal agency	(0.053)	causal agency	(0.053)
person	(0.053)	person	(0.053)
individual	(0.053)	individual	(0.053)
someone	(0.053)	someone	(0.053)
somebody	(0.053)	somebody	(0.053)
mortal	(0.053)	mortal	(0.053)
human	(0.053)	human	(0.053)
soul	(0.053)	soul	(0.053)

Figure 4.9: Semantic profiles for Subject position of a transitive usage for a novel verb *gorp*, and an intransitive verb *dance*

.

Transitive Object			
gorp		*dance*	
entity	(0.028)	entity	(0.024)
object	(0.022)	object	(0.017)
physical object	(0.022)	physical object	(0.017)
artifact	(0.017)	location	(0.013)
artefact	(0.017)	artifact	(0.011)
whole	(0.017)	artefact	(0.011)
whole thing	(0.017)	whole	(0.011)
unit	(0.017)	whole thing	(0.011)
		unit	(0.011)

Figure 4.10: Semantic profiles for Object position of a transitive usage for a novel verb *gorp*, and an intransitive verb *dance*

.

in the learning curves of Agent and Theme roles in Fig. 4.8, where the semantic profile predicted by the model for each role shows fluctuations at the beginning, but eventually stabilizes after receiving enough input.

4.8 Summary of Contributions

We have shown that our Bayesian model of early verb learning, extended to include sets of semantic properties for arguments, can acquire associations between those properties, the syntactic positions of the arguments, and the semantic primitives of verbs. These probabilistic associations enable the model to learn general conceptions of roles, based only on exposure to individual verb usages, and without requiring explicit labelling of the roles in the input. The acquired role properties are a good intuitive match to the expected properties of various roles, and are useful in guiding comprehension in the model to the most likely interpretation in the face of ambiguity.

The semantic profiles that we use are drawn from a standard lexical resource (WordNet; Miller, 1990), so that our results are not biased toward any theory of semantic roles. One limitation of this approach is that the profiles fail to reflect any event-specific properties that an argument might have. Such properties (like "causally affected") are almost certainly required in an accurate representation of roles, as in Dowty (1991). Despite their absence, we are able to show that intuitive profiles can be learned for each role from examples of its use. We further establish that such representations can be useful in guiding the argument interpretation of ambiguous input, an ability experimentally demonstrated in children in recent work (Nation et al., 2003).

Because of the model's Bayesian formulation, the roles naturally metamorphose from verb-specific to highly general properties. However, because the prediction of the semantic roles is solely based on the groupings of verbs, it cannot make use of verb-specific knowledge in generating expectations about a particular verb's arguments. That is, once

it is exposed to a range of verbs, it no longer has access to the verb-specific information, only to generalizations over clusters of verbs. The semantic properties that a particular verb imposes on its arguments, or its *selectional preferences*, are an important characteristics of a verb, and a comprehensive model of verb learning should acquire such information. We study this aspect in the next chapter.

Chapter 5

Learning Verb Selectional Preferences

5.1 Overview

Many verbs show strong preferences concerning the semantic properties of their arguments. For example, *eating food* and *drinking water* are acceptable, whereas **eating water* and **drinking food* are normally not. The computational study of learning verb selectional preferences is heavily influenced by Resnik (1996), who introduced a statistical approach to learning and use of verb selectional preferences. In this framework, a semantic class hierarchy for words is used, together with statistical tools, to induce a verb's selectional preferences for a particular argument position in the form of a probability distribution over all the classes that can occur in that position. Resnik's model was proposed as a model of human learning of selectional preferences that made minimal representational assumptions; it showed how such preferences could be acquired from usage data and an existing conceptual hierarchy. However, his and later computational models (see Section 5.3) have properties that do not match with certain cognitive plausibility criteria for a child language acquisition model. All of these models use the training

data in "batch mode", and most of them use information-theoretic measures that rely on total counts from a corpus. Therefore, it is not clear how the representation of selectional preferences could be updated incrementally in these models as the person receives more data. Moreover, the assumption that children have access to a full hierarchical representation of semantic classes may be too strict. We propose an alternative view in this paper which is more plausible in the context of child language acquisition.

In Chapter 4, we showed that our model can incrementally learn a semantic profile for the semantic roles of the predicates based on exposure to individual verb usages. We also showed that these profiles initially yield verb-specific properties, but eventually transform into more abstract representations that reflect the general properties of arguments across the observed verbs. A shortcoming of the model was that, because the prediction of the semantic roles was based only on the groupings of verbs, it could not make use of verb-specific knowledge in generating expectations about a particular verb's arguments. That is, once it was exposed to a range of verbs, it no longer had access to the verb-specific information, only to generalizations over clusters of verbs.

In this chapter, we propose a new version of our model that, in addition to learning general semantic roles for constructions, can use its verb-specific knowledge to predict intuitive selectional preferences for each verb argument position. We introduce a new notion, a *verb semantic profile*, as a probability distribution over the semantic properties of an argument for each verb. Our proposed prediction model makes appropriate generalizations over the observed properties, and captures expectations about previously unseen arguments.

As in other work on selectional preferences, as well as our own semantic role learning model, the semantic properties that we use in our representation of arguments are drawn from a standard lexical ontology (WordNet; Miller, 1990). It must be noted that our model does not require knowledge of the hierarchical structure of the WordNet concepts. From the computational point of view, this makes use of an available resource, while

from the cognitive view, this avoids ad hoc assumptions about the representation of a conceptual hierarchy. However, we do require some properties to be more general (i.e., shared by more words) than others, which eventually enables the model to make appropriate generalizations. Otherwise, the selected semantic properties are not fundamental to the model, and could in the future be replaced with an approach that is deemed more appropriate to child language acquisition. Each argument contributes to the semantic profile of the verb through its (potentially large) set of semantic properties instead of its membership in a single class. As input to our model, we use an automatically parsed corpus, which is very noisy. However, as a result of our novel representation, the model can induce and use selectional preferences using a relatively small set of noisy training data.

Section 5.2 discusses the treatment of verb selectional constraints or preferences in the linguistic and computational linguistic literature. The existing computational models of selectional preferences are reviewed in Section 5.3. In Section 5.4 we present an adaptation of our Bayesian model for representation and learning of verb selectional preferences. We describe our experimental set-up in Section 5.5, and present the experimental results in Section 5.6. Section 5.7 discusses these results, and the suitability of the model for this task. Finally, we summarize our contributions in Section 5.8.

5.2 Verb Selectional Preferences

Selectional preferences, or constraints, are viewed as limitations on the applicability of natural language predicates to arguments. In their semantic theory, Katz and Fodor (1964) characterized selectional constraints as restrictions in terms of the defining features of the arguments: they outlined a decompositional theory of word meaning in which lexical entries specified the features applicable to a particular lexical item. For words that denote predicates, Katz and Fodor proposed that the arguments in their lexical entries

be annotated with restrictions identifying the necessary and sufficient conditions that a semantically acceptable argument must meet. Such conditions were represented as Boolean functions of semantic features, such as HUMAN or HIGHER ANIMAL for the subject of the verb *hit*, and PHYSICAL OBJECT for its object. Jackendoff's (1983) lexical theory, on the other hand, situates selectional constraints as information appearing in the context of a rich representation of the predicate's meaning, such as the annotation LIQUID appearing as a constraint on one argument of the verb *drink*. Selectional constraints were also explicitly integrated into grammar, as in Generative Lexicon Theory (Pustejovsky, 1995; Copestake and Briscoe, 1991). However, identifying restrictions that are both necessary and sufficient, and choosing the primitives themselves, is viewed by many to be an insurmountable problem.

Resnik (1993) instead emphasized the view of the restrictions a verb places on its arguments as selectional preferences, and proposed a new approach to their representation and learning, which was followed by many in the computational linguistics community. In this approach, the knowledge of words, or concepts, is represented as a pre-defined semantic class hierarchy, and statistical tools are used to learn selectional preferences from examples in a corpus. As opposed to a Boolean interpretation of selectional constraints, here the selectional preferences are viewed as probability distributions over various semantic classes. For example, the preferred objects of *eat* are represented not as the black-and-white class FOOD but rather as a gray probability distribution over all nouns or various classes thereof.

Many theories of lexical acquisition make use of selectional constraints (Gleitman and Gillette, 1995; Pinker, 1994). Gleitman and Gillette (1995) show that selectional constraints provide adult subjects with significant constraints on the possible meanings of unknown verbs: the subjects identified a verb 80% of the time if they were given the syntactic frame of the verb together with the nouns that appear as the verb arguments; however, the syntactic frame alone or the noun arguments alone (without specifying

their position) did not help subjects to identify the verb half the time. This shows that the semantic properties of the verb arguments (or verb selectional preferences) are more informative than simply the semantic associations between a verb and a group of nouns, or the syntactic properties of the verb. Moreover, selectional constraints play an important role in many aspects of language processing: they influence the syntactic structure of a sentence, especially in the face of ambiguity; they affect selecting the likely word in a sequence of speech signals; and they can be drawn on for the task of word sense disambiguation. Therefore, the process by which the acquisition of selectional preferences takes place has to be made explicit.

5.3 Related Computational Models

Two central questions for the automated treatment of selectional preferences are: what *representation* to use, and how to *induce* preferences from available data. A variety of computational models of verb selectional preferences have been proposed, which use different statistical models to induce the preferences of each verb from corpus data. Most of these models, however, use the same representation for verb selectional preferences: the preference can be thought of as a mapping, with respect to an argument position for a verb, of each class to a real number (Light and Greiff, 2002). The induction of a verb's preferences is, therefore, modeled as using a set of training data to estimate that number.

Resnik (1993) is the first to model the problem of induction of selectional preferences using a a pre-existing semantic class hierarchy, WordNet (Miller, 1990). He defines the selectional preference strength of a verb as the divergence between two probability distributions: the prior probabilities of the classes, and the posterior probabilities of the classes given that verb. The selectional association of a verb with a class is also defined as the contribution of that class to the total selectional preference strength. Resnik

estimates the prior and posterior probabilities based on the frequencies of each verb and its relevant argument in a corpus.

Following Resnik (1993, 1996), a number of methods were presented that make use of WordNet and a text corpus, together with a variety of statistical models, to induce selectional preferences. Li and Abe (1998) model selectional preferences of a verb (for an argument position) as a set of nodes in WordNet with a probability distribution over them. They use the Minimum Description Length (MDL) principle to find the best set for each verb and argument based on the usages of that verb in the training data. Clark and Weir (2002) also find an appropriate set of concept nodes to represent the selectional preferences for a verb, but do so using a χ^2 test over corpus frequencies mapped to concepts to determine when to generalize from a node to its parent. Ciaramita and Johnson (2000) use a Bayesian network with the same topology as WordNet to estimate the probability distribution of the relevant set of nodes in the hierarchy. Abney and Light (1999) use a different representational approach: they train a separate hidden Markov model for each verb, and the selectional preference is represented as a probability distribution over words instead of semantic classes.

In contrast to the WordNet-based methods above, Erk (2007) proposes a similarity-based model that does not rely on a hierarchical representation of semantic classes. Instead, her model estimates the selectional preference of a predicate argument position for a possible head word as a weighted sum of the similarities between that word and the seen head words for that argument position in a corpus. The similarity between the seen and the potential head word is computed based on a corpus-based semantic similarity metric.

All these models use the training data in "batch mode", and most of them use statistical measures that rely on total counts from a corpus. Therefore, it is not clear how the representation of selectional preferences could be updated incrementally in these models as the person receives more data. Moreover, the assumption that children have access

to a full hierarchical representation of semantic classes, as in most of these models, may be too strict. In the context of child language acquisition, an alternative, cognitively plausible view of learning selectional preferences is lacking.

It is not easy to evaluate the acquired selectional preferences on their own, since there is no "gold standard" set of examples against which to compare the outcome of a method (Light and Greiff, 2002). The existing models of verb selectional preference have been evaluated through a wide range of computational linguistic tasks, including word sense disambiguation (Resnik, 1997; Abney and Light, 1999; Ciaramita and Johnson, 2000; McCarthy and Carroll, 2003), PP-attachment disambiguation (Li and Abe, 1998); a pseudo-disambiguation task of choosing the best verb-argument pair (Clark and Weir, 2002), and semantic role labelling (Erk, 2007). Resnik (1996) also evaluated his method through two other tasks that are more interesting from a human language acquisition point of view: the simulation of plausibility judgements elicited from adult subjects, and the analysis of the way that arguments are syntactically realized for a class of English verbs. We refer to these two tasks in our experimental results.

5.4 A Computational Model of Learning Verb Selectional Preferences

We use our Bayesian framework, presented in Chapter 2, for modeling learning and use of the verb selectional preferences. As before, our model learns the argument structure frames for each verb, and their grouping across verbs into constructions. For each verb, the model associates an argument position with a probability distribution over a set of semantic properties—a verb semantic profile. In doing so, the model uses the knowledge that it has learned for that verb, as well as the grouping of frames for that verb into constructions.

The basic properties regarding the input to the model, as well as the frame structure

and features used in this variation of the model, are detailed in Section 5.4.1. In Section 5.4.2, a formal definition of a verb semantic profile is given. Finally, in Section 5.4.3, a measure of compatibility between a verb and a noun for a particular argument position is presented.

5.4.1 The Input and the Argument Structure Frames

For the previous versions of the model, we used a set of scene-utterance pairs that link a relevant aspect of an observed scene to the utterance that describes it. However, that relied on manually-compiled data about verbs and their argument structure frames from the CHILDES database (MacWhinney, 1995). To evaluate the new version of our model for the task of learning selectional preferences, we need a wide selection of verbs and their arguments that is impractical to compile by hand.

Instead, for the current version of the model, we generate the input for our experiments from corpus data as follows. For each verb usage in a sentence, we construct a frame by recording the verb in root form, the number of the arguments for that verb, and the syntactic pattern of the verb usage (i.e., the word order of the verb and the arguments). We also record in the frame the semantic properties of the verb and each of the argument heads (each noun is also converted to root form). The extracted frame is stored in the lexical entry of the head word, and is also input to our Bayesian learner to be classified. Fig. 5.1 shows a sample usage and the corresponding extracted frame.

As in the previous chapter, the semantic properties of words are taken from WordNet (version 2.0). However, because the size of the lexicon in the current version is considerably increased, it is not feasible to choose the right sense for each word in a sentence. Therefore, we extract all the hypernyms (ancestors) for all the senses of a word, and add all the words in the hypernym synsets to the list of the semantic properties. Fig. 5.2 shows an example of the hypernyms for *dinner*, and its resulting set of semantic properties. Note that what we called the semantic primitives of the predicate in the previous

Verb Usage	
Rose made dinner.	
Extracted Frame	
head verb	*make*
semantic properties of verb	[*change, alter, modify, make, create, ...*]
number of arguments	2
semantic properties of argument 1	{ *woman, female, person, individual, ...* }
semantic properties of argument 2	{ *meal, repast, nutriment, ...* }
syntactic pattern	arg1 *verb* arg2

Figure 5.1: A verb usage and its corresponding frame.

chapters, which were manually devised to reflect the properties of the event, are now represented by the automatically extracted semantic properties of verbs from WordNet. For prepositions, however, we still use a manually selected set of semantic primitives.

5.4.2 Predicting Semantic Profiles for Verbs

We represent the selectional preferences of a verb for an argument position as a verb semantic profile, which is a probability distribution over all the semantic properties. Unlike semantic roles, we do not want a verb semantic profile to be a generalization over the argument properties across all the usages in a relevant construction. Instead, we want the verb profile to reflect these semantic properties of the arguments that appear in the usages of that verb. However, we still need to take into consideration the groupings of the verb usages into constructions, since the more entrenched constructions are expected to have a higher influence on the profile.

```
Sense 1
dinner
       => meal, repast
           => nutriment, nourishment, nutrition, sustenance,
              aliment, alimentation, victuals
               => food, nutrient
                   => substance, matter
                       => entity
Sense 2
dinner, dinner party
       => party
           => social gathering, social affair
               => gathering, assemblage
                   => social group
                       => group, grouping
```

dinner: {meal, repast, nutriment, nourishment, nutrition, substance, aliment, alimentation, victuals, food, nutrient, substance, matter, entity, party, social gathering, social affair, gathering, assemblage, social group, group, grouping }

Figure 5.2: Semantic properties for *dinner* from WordNet

To predict the profile of a verb v for an argument position *arg*, we need to estimate the probability of each semantic property j separately.

$$\begin{aligned} P_{arg}(j|v) &= \sum_k P_{arg}(j,k|v) \\ &\propto \sum_k P(k,v) P_{arg}(j|k,v) \end{aligned} \tag{5.1}$$

Here, j ranges over all the possible semantic properties that an argument can have, and k ranges over all constructions. The prior probability of having verb v in construction k, or $P(k,v)$, is mainly determined by the frequency with which v participates in k, or f_k^v. However, we want to emphasize the influence of the well-entrenched constructions in the prior probability. To determine the relative entrenchment of a construction k among all the constructions, we calculate its weight w_k by taking its number of frames n_k over the total number of frames:

$$w_k = \frac{n_k}{\sum_{k'} n_{k'}} \tag{5.2}$$

The prior probability $P(k,v)$ is then calculated by:

$$P(k) = \frac{w_k \times f_k^v}{\sum_{k'} w_{k'} \times f_{k'}^v} \tag{5.3}$$

The posterior probability $P_{arg}(j|k,v)$ is calculated analogously to $P_i(j|k)$ in equation (4.1) (see Chapter 2, page 93), but limiting the count of matching features to those frames in k that contain v:

$$P_{arg}(j|k,v) = \frac{\textbf{verb_count}_{arg}^k(j,v) + \lambda \alpha_{prop_j}}{n_{kv} + \lambda \alpha_{prop}} \tag{5.4}$$

where n_{kv} is the number of frames for v participating in construction k, $\textbf{verb_count}_{arg}^k(j,v)$ is the number of those with semantic property j for argument *arg*. As before, λ and α_{prop}

are the smoothing factors. However, because the semantic property j can be more general (i.e., shared by more words) or more specific, we no longer assume that the default probability for different values of j is uniform, since the more general properties are applicable to a wider range of words, and therefore should have a higher default probability. Instead, we introduce a new smoothing factor α_{prop_j}, where the default probability of j is $\alpha_{prop_j}/\alpha_{prop}$.

5.4.3 Verb-Argument Compatibility

To estimate verb argument plausibility, we need a measure of compatibility of a particular noun n for an argument position *arg* of some verb v. That is, we need to estimate how much the semantic properties of n conform to the acquired semantic profile of v for *arg*. We formulate the compatibility as the conditional probability of observing n as an argument *arg* of v:

$$\textbf{compatibility}(v, n) = \log(P_{arg}(S_n|v)) \tag{5.5}$$

where S_n is the set of the semantic properties for word n, and $P_{arg}(S_n|v)$ is estimated as in equation (5.1). However, since S_n here is a set of properties (as opposed to j in equation (5.1) being a single property), $\textbf{verb_count}^k_{arg}$ in equation (5.4) should be modified as described in Section 4.4.2:

$$\textbf{verb_count}^k_{arg}(S_n, v) = \sum_{f \in k} \textbf{sem_score}(S_n, S_{arg(f)}) \tag{5.6}$$

where f is a frame that belongs to construction k, and $S_{arg(f)}$ is the set of the semantic properties for argument *arg* of frame f. The measure **sem_score** uses Jaccard similarity score to compare two sets.

5.5 Experimental Set-up

Through a number of computational experiments, we demonstrate that our model can learn intuitive semantic profiles for different verbs. We also show that the model can use the learned profiles in different language tasks. In Section 5.5.1, we describe how the input data for each simulation in our set of experiments is extracted from text corpora. In Section 5.5.2, we explain how the parameters of the model are set.

5.5.1 The Input Corpora

The training data for our experiments here are generated as follows. We use 20,000 sentences randomly selected from the British National Corpus (BNC, 2000), to extract verb–noun pairs, along with information on the syntactic patterns they appear in. We use Fazly's (2007) subcategorization frames, which are extracted as follows. The BNC is automatically parsed using the Collins parser (Collins, 1999), and further processed using TGrep2 (Rohde, 2004), in order to extract syntactic dependencies. For each instance of a transitive verb, heuristics are used to extract the noun phrase (NP) in each argument position. A head extraction software[1] is then used to get the head noun of the extracted NP.

For each verb usage in a sentence, a frame is extracted as described in Section 5.4.1. If the head verb or any of its arguments cannot be found in WordNet, that frame is discarded. This process results in 16,300 frames which serve as input data to our model.

5.5.2 Setting the Parameters of the Model

For the purpose of implementing the model and running simulations, we need to assign values to parameters α_i for each feature i, as well as λ, in equation (2.6) (page 24 in

[1]We use a modified version of the software provided by Eric Joanis based on heuristics from Collins (1999).

Chapter 2). We set the parameter $\lambda = 10^{-4}$ as described in Section 3.5.3. In the model presented in this chapter, there are four α_i parameters that need to be set, each of which corresponds to a different feature in a frame: the semantic primitives of the head word, the number of arguments, the semantic properties of the arguments, and the syntactic pattern. Since each α_i represents the maximum number of distinct values that feature i can take, the appropriate values of these parameters are determined based on the input data: $\alpha_{prim} = 3390$, $\alpha_{arg_num} = 6$, and $\alpha_{patt} = 15$. We set α_{prop_j} in equation (5.4) to the relative frequency of each property j among all nouns in the input, and $\alpha_{prop} = \sum_j \alpha_{prop_j}$. There are a total of 8890 distinct semantic properties in our training data.

5.6 Experimental Results

In the following sections, we report computational experiments that demonstrate the ability of our model to learn the selectional preferences of a wide range of verbs. In accordance with other computational models, we focus here on the verb preferences for the direct object position.[2] We first provide a qualitative analysis of our model through examination of the semantic profiles for a number of verbs. As noted earlier, we then evaluate our model through two tasks of simulating verb-argument plausibility judgment, and analyzing the implicit object alternation, following Resnik (1996).

5.6.1 Formation of Semantic Profiles for Verbs

After training our model on the training data (see Section 5.5.1), we use equation (5.1) to predict the semantic profile of the direct object position for a range of verbs. Some of these verbs, such as *write* and *sing*, have strong selectional preferences, whereas others, such as *want* and *put*, can take a wide range of nouns as direct object (as confirmed

[2]To our knowledge, the only work that considers selectional preferences of subjects and prepositional phrases as well as direct objects is that of Brockmann and Lapata (2003).

by Resnik's (1996) estimated strength of selectional preference for these verbs). The semantic profiles for *write* and *sing* are displayed in Fig. 5.3, and the profiles for *want* and *put* are displayed in Fig. 5.4. Due to the large number of semantic properties in each profile (more than 8800), we only include the 25 properties that have the highest probability in each profile.

Because we extract the semantic properties of words from WordNet, which has a hierarchical structure, the properties that come from nodes in the higher levels of the hierarchy (such as *entity* and *abstraction*) appear as the semantic property for a very large set of words, whereas the properties that come from the leaves in the hierarchy are specific to a small set of words. Therefore, the general properties are more likely to be associated with a higher probability in the semantic profiles for most verbs. In fact, a closer look at the semantic profiles for *want* and *put* reveals that the top portion of the semantic profile for these verbs consists solely of such general properties that are shared among a large group of words. However, this is not the case for the more restrictive verbs. The semantic profiles for *write* and *sing* show that the specific properties that these verbs demand from their direct object appear amongst the highest-ranked properties, even though only a small set of words share these properties (e.g., *content, message, written communication, written language, ...* for *write*, and *auditory communication, music, opus, ...* for *sing*).

write		*sing*	
(0.024)	abstraction	(0.020)	abstraction
(0.022)	entity	(0.015)	relation
(0.021)	location	(0.015)	communication
(0.020)	substance	(0.015)	social relation
(0.019)	destination	(0.013)	act
(0.018)	relation	(0.013)	human action
(0.015)	communication	(0.013)	human activity
(0.015)	social relation	(0.013)	auditory communication
(0.013)	content	(0.012)	music
(0.011)	message	(0.010)	entity
(0.011)	subject matter	(0.010)	piece
(0.011)	written communication	(0.009)	composition
(0.011)	written language	(0.009)	musical composition
(0.010)	object	(0.009)	opus
(0.010)	physical object	(0.009)	piece of music
(0.010)	writing	(0.009)	psychological feature
(0.010)	goal	(0.008)	cognition
(0.010)	unit	(0.008)	knowledge
(0.009)	whole	(0.008)	noesis
(0.009)	whole thing	(0.008)	activity
(0.009)	artifact	(0.008)	content
(0.009)	artefact	(0.008)	grouping
(0.009)	state	(0.008)	group
(0.009)	amount	(0.008)	amount
(0.009)	measure	(0.008)	measure

Figure 5.3: Semantic profiles of *write* and *sing* for the direct object position.

want		*put*	
(0.016)	entity	(0.015)	entity
(0.015)	object	(0.015)	object
(0.015)	physical object	(0.013)	physical object
(0.014)	abstraction	(0.013)	abstraction
(0.013)	act	(0.011)	unit
(0.012)	human action	(0.011)	whole
(0.012)	human activity	(0.011)	whole thing
(0.012)	relation	(0.011)	artifact
(0.011)	unit	(0.011)	artefact
(0.011)	whole	(0.010)	act
(0.011)	whole thing	(0.009)	relation
(0.011)	artifact	(0.008)	human action
(0.011)	artefact	(0.008)	human activity
(0.008)	communication	(0.008)	communication
(0.008)	social relation	(0.008)	social relation
(0.008)	activity	(0.007)	substance
(0.007)	cause	(0.007)	content
(0.007)	state	(0.007)	instrumentality
(0.007)	instrumentality	(0.007)	instrumentation
(0.007)	instrumentation	(0.007)	measure
(0.007)	event	(0.006)	amount
(0.006)	being	(0.006)	quantity
(0.006)	living thing	(0.006)	cause
(0.006)	animate thing	(0.006)	causal agent
(0.006)	organism	(0.006)	causal agency

Figure 5.4: Semantic profiles of *want* and *put* for the direct object position.

A novel verb	
(0.021)	entity
(0.017)	object
(0.017)	physical object
(0.015)	abstraction
(0.010)	act
(0.010)	human action
(0.010)	human activity
(0.010)	unit
(0.009)	whole
(0.009)	whole thing
(0.009)	artifact
(0.009)	artefact
(0.009)	being
(0.009)	living thing
(0.009)	animate thing
(0.009)	organism
(0.008)	cause
(0.008)	causal agent
(0.008)	causal agency
(0.008)	relation
(0.008)	person
(0.008)	individual
(0.008)	someone
(0.008)	somebody
(0.008)	mortal

Figure 5.5: Semantic profile of a novel verb for the direct object position.

The examination of the semantic profiles for fairly frequent verbs in the training data shows that our model can use the verb usages to predict an appropriate semantic profile for each verb. When presented with a novel verb (for which no verb-based information is available), equation (5.1) predicts a semantic profile which reflects the relative frequencies of the semantic properties among all words (due to the smoothing factor added to equation (5.4)), modulated by the prior probability of each construction. The predicted profile is displayed in Fig. 5.5. It shows similarities with the profiles for *want* and *put* in Fig. 5.4, but the general properties in this profile have an even higher probability. Since the profile for the novel verb is predicted in the absence of any evidence (i.e., verb usage) in the training data, we later use it as the base for estimating other verbs' strength of selectional preference.

5.6.2 Verb-Argument Plausibility Judgments

Holmes et al. (1989) evaluate verb argument plausibility by asking human subjects to rate sentences like *The mechanic warned the driver* and *The mechanic warned the engine*. Resnik (1996) used this data to assess the performance of his model by comparing its judgments of selectional fit against the plausibility ratings elicited from human subjects. He showed that his selectional association measure for a verb and its direct object can be used to select the more plausible verb-noun pair among the two (e.g., $\langle$*warn, driver*$\rangle$ vs. $\langle$*warn, engine*$\rangle$ in the previous example). That is, a higher selectional association between the verb and one of the nouns compared to the other noun indicates that the former is the more plausible pair. Resnik (1996) used the Brown corpus as training data, and showed that his model arrives at the correct ordering of more and less plausible arguments in 11 of the 16 cases.

We repeated this experiment, using the same 16 pairs of verb-noun combinations. For each pair of $\langle v, n_1 \rangle$ and $\langle v, n_2 \rangle$, we calculate the compatibility measure using equation (5.5) on page 126; these values are shown in Fig. 5.6. (Note that because these are

log-probabilities and therefore negative numbers, a lower absolute value of **compatibility**(v, n) shows a better compatibility between the verb v and the argument n.) For example, $\langle$*see, friend*$\rangle$ has a higher compatibility score (-30.50) than $\langle$*see, method*$\rangle$ (-32.14). Like Resnik's model, our model detects 11 plausible pairs out of 16. However, these results are reached with a much smaller training corpus (around 500,000 words), compared to the Brown corpus used by Resnik (1996) which contains one million words. Moreover, whereas the Brown corpus is tagged and parsed manually, the portion of the BNC that we use is parsed automatically, and as a result our training data is very noisy. Nonetheless, the model achieves the same level of accuracy in distinguishing plausible verb-argument pairs from implausible ones.

5.6.3 Implicit Object Alternations

In English, some inherently transitive verbs can appear with or without their direct objects (e.g., *John ate his dinner* as well as *John ate*), but others cannot (e.g., *Mary made a cake* but not **Mary made*). It is argued that implicit object alternations involve a particular relationship between the verb and its argument. In particular, for verbs that participate in the implicit object alternation, the omitted object must be in some sense inferable or *typical* for that verb (Levin, 1993, among others).

Resnik (1996) used his model of selectional preferences to analyze implicit object alternations, and showed a relationship between his measure of selectional preference strength and the notion of typicality of an object. He calculated this measure for two groups of Alternating and Non-alternating verbs, and showed that, on average, the Alternating verbs have a higher strength of selectional preference for the direct object than the Non-alternating verbs. However, there was no threshold separating the two groups of verbs.

To repeat Resnik's experiment, we need a measure of how "strongly constraining" a semantic profile is. We can do this by measuring the similarity between the semantic

Verb	Plausible		Implausible	
see	**friend**	**−30.50**	**method**	**−32.14**
read	**article**	**−32.76**	**fashion**	**−33.33**
find	**label**	**−32.05**	**fever**	**−33.30**
hear	**story**	**−32.11**	**issue**	**−32.40**
write	**letter**	**−31.37**	**market**	**−32.46**
urge	daughter	−36.73	contrast	−35.64
warn	**driver**	**−33.68**	**engine**	**−34.42**
judge	contest	−39.05	climate	−38.23
teach	language	−45.64	distance	−45.11
show	sample	−31.75	travel	−31.42
expect	visit	−33.88	mouth	−32.87
answer	**request**	**−31.89**	**tragedy**	**−33.95**
recognize	**author**	**−32.53**	**pocket**	**−32.62**
repeat	**comment**	**−33.80**	**journal**	**−33.97**
understand	**concept**	**−32.25**	**session**	**−32.93**
remember	**reply**	**−33.79**	**smoke**	**−34.29**

Figure 5.6: Compatibility scores for plausible vs. implausible verb-noun pairs. The correctly detected plausible pairs are shown in boldface.

profile we generate for the object of a particular verb and some "default" notion of the argument for that position across all verbs. We use the semantic profile predicted for the object position of a novel verb, shown earlier in Fig. 5.5, as the default profile for that argument position. Because this profile is predicted in the absence of any evidence in the training data, it makes the minimum assumptions about the properties of the argument and thus serves as a suitable default. We then assume that verbs with weaker preferences have semantic profiles more similar to the default profile than verbs with stronger preferences. We use the cosine measure to estimate the similarity between two profiles p and q:

$$\text{cosine}(p, q) = \frac{p \times q}{||p|| \times ||q||} \tag{5.7}$$

The similarity values for the Alternating and Non-alternating verbs are shown in Fig. 5.7. The larger values represent more similarity with the base profile, which means a weaker selectional preference. The means for the Alternating and Non-alternating verbs were respectively 0.76 and 0.81, which confirms the hypothesis that verbs participating in implicit object alternations select more strongly for the direct objects than verbs that do not. However, like Resnik (1996), we find that it is not possible to set a threshold that will distinguish the two sets of verbs.

5.7 General Discussion

The model presented in this chapter for learning and use of verb selectional preferences is a natural continuation of the semantic role learning model that was presented in Chapter 4. Together, these two models show that a semantic profile, or a probability distribution over semantic properties, can be learned both for argument positions in a general construction, and for an individual verb's arguments. Unlike other existing models of selectional preferences, our model is incremental, and does not rely on having access to a full hierarchical representation of semantic classes prior to the acquisition

Alternating verbs		**Non-alternating verbs**	
write	0.61	*hang*	0.56
sing	0.67	*wear*	0.71
drink	0.67	*say*	0.75
eat	0.74	*catch*	0.76
play	0.74	*show*	0.77
pour	0.76	*make*	0.78
watch	0.77	*hit*	0.78
pack	0.78	*open*	0.81
steal	0.80	*take*	0.83
push	0.80	*see*	0.87
call	0.80	*like*	0.87
pull	0.80	*get*	0.87
explain	0.81	*find*	0.87
read	0.82	*give*	0.88
hear	0.87	*bring*	0.89
		want	0.89
		put	0.90
Mean:	0.76	Mean:	0.81

Figure 5.7: Similarity with the base profile for Alternating and Non-alternating verbs.

of verb selectional preferences. These two properties make our model more cognitively plausible than other existing models

The training data set that we use in this chapter is considerably larger in size than those used in the experiments reported in the previous chapters. We investigate the effect of a larger data set on the performance of the model in Section 5.7.1. In Section 5.7.2, we look at the possibility of using the current model for acquiring the selectional preferences for other parts of speech, especially nouns.

5.7.1 The Effect of Processing More Input Data

Unlike the previous versions of the model, the current version uses corpus data for training. In Chapters 3 and 4, we used at most 1000 input pairs in a simulation, whereas here we used more than 16,000 frames as input. This gives us an opportunity to investigate the scalability of the model. As expected, as a result of the Bayesian formulation that we use, both learning and prediction (especially learning) are computationally expensive: to find the best construction for a new frame, we have to compare the frame with every frame in every construction. This limitation prevents us from studying the effect of more input on the performance of the model by easily increasing the size of the corpus. We will return to this issue in the next chapter.

A central problem for induction of the selectional preferences is noise in the training data. Noise can be due to errors in part of speech tagging or syntactic analysis, or due to metaphorical usage. Another problem is word sense ambiguity in the training data, which can lead to incorrect generalization, especially because we merge the hypernyms of multiple senses of a word to construct its set of semantic properties. (However, McCarthy and Carroll (2003) show that integrating a word sense disambiguation module into a selectional preference induction model does not significantly improve the performance.) Both these problems are expected to be improved by increasing the size of training data, something we could not explore due to the computational expensiveness of our model.

However, our model shows promise in learning intuitive semantic profiles for verbs and in simulating human plausibility judgements using small and highly noisy input data.

In examining the emergent constructions in the simulations of this chapter, we found that the learned constructions are more coarse-grained than the ones learned in the previous chapters. That is, the number of the acquired constructions is relatively small, where each construction contains a very large group of frames. Such constructions are less effective in controlling the outcome of the prediction model. In future research, we need to further explore this issue through different parameter settings that better suit the characteristics of the current input data.

5.7.2 Noun Selectional Preferences

It has been argued that selectional preferences can be applied for nouns too, since they "prefer" to select certain predicates over others (Pustejovsky, 1995). For example, *cake* prefers to be baked, but not written. We can model this reverse preference as predicting the semantic primitives of the head word (verb) based on the properties of the arguments and the syntactic pattern of the usage. Moreover, the compatibility measure introduced in equation (5.5), Section 5.4.3, can also be used to compare two pairings of a noun with two different verbs, and to measure the preference of that noun for each of the verbs (e.g., $\langle$*bake, cake*$\rangle$ vs. $\langle$*write, cake*$\rangle$). This possibility of extending to other parts of speech is lacking in the existing models of selectional preferences.

5.8 Summary of Contributions

We have proposed a cognitively plausible model for learning selectional preferences from instances of verb usage. The model represents verb selectional preferences as a semantic profile, which is a probability distribution over the semantic properties that an argument can take. One of the strengths of our model is the incremental nature of its learning

mechanism, in contrast to other approaches which learn selectional preferences in batch mode. Here we have only reported the results for the final stage of learning, but the model allows us to monitor the semantic profiles during the course of learning, and compare it with child data for different age groups, as we do with semantic roles in Chapter 4 (however, to our knowledge, there is no such experimental data reported in the literature). We have shown that the model can predict appropriate semantic profiles for a variety of verbs, and use these profiles to simulate human judgments of verb-argument plausibility, using a small and highly noisy set of training data. The model can also use the profiles to measure verb-argument compatibility, which was used in analyzing the implicit object alternation.

Chapter 6

Summary and Future Directions

6.1 Summary of Contributions

Developing computational algorithms that capture the complex structure of natural languages is an open problem. In particular, learning the abstract properties of language only from usage data remains a challenge. In this dissertation, we have presented a computational, usage-based model of verb argument structure acquisition. We have shown that our model can successfully learn abstract knowledge of language from instances of verb usage, and use this knowledge in various language tasks. We have proposed a novel representation for argument structure constructions, as well as a new, unified formulation for both language learning and use.

The main focus of our work is on learning and use of verb argument structure. We have shown through computational simulations that the model can successfully learn the semantic relationship between a verb and its arguments, and their syntactic realization in a natural language sentence. We have also investigated the learnability of verb semantic roles, a critical aspect of linking the syntax and semantics of verbs. In contrast to many existing linguistic theories and computational models, which assume that semantic roles are innate and fixed, we have shown that general conceptions of semantic roles can be

learned from the semantic properties of the verb arguments in the input usages. We have extended this view to model the learning and use of verb selectional preferences, a phenomenon usually viewed as separate from verb semantic roles. We review our contributions with regard to each aspect of our model in the following sections.

6.1.1 Representation

The main goal of our work is to show that it is possible to learn general regularities of language from verb usage only, and use these regularities in language use the way children do. We draw on Construction Grammar in representing the regularities in verb behaviour in the form of argument structure constructions, where the surface form and the elements of meaning are learned and used as a single unit. However, we propose a new view of argument structure construction as a probabilistic association between syntactic and semantic features of verb usages. The probabilistic nature of argument structure constructions in the model enables it to capture both statistical effects in language learning, and adaptability in language use.

As for verb semantic roles and selectional preferences, we use the same representation to model both phenomena: a probability distribution over the semantic properties that an argument can take. Whereas general roles are represented as the semantic properties that are preferred by a general construction for a particular argument position, verb selectional preferences are viewed as the preferred semantic properties for an argument position by a particular verb.

6.1.2 Underlying mechanisms

For modelling linguistic abstraction, we have extended a general model of categorization (Anderson, 1991) which employs some of the most basic cognitive skills: analogy and categorization. We model the learning of the abstract knowledge of language as detecting similar usages, and grouping them together. As mentioned before, each group of usages

represents a general construction in language. We use a probabilistic measure of similarity between verb usages, and a Bayesian framework for clustering them. Learning in our model is incremental: upon receiving each new verb usage as input, the model finds the most suitable construction to link to the new item. The Bayesian learning model finds the best construction based on two factors: the degree of entrenchment of each construction, and the similarity between the new input usage and the usages that have already been grouped into that construction.

We propose a creative way of modelling language use as a prediction problem: each language task is viewed as finding the best value for a missing feature in a usage, based on the available features in that same usage and the acquired knowledge of language so far. For example, in sentence production, the most suitable syntactic pattern must be found for expressing the semantics of a verb and its relation with its arguments. Similarly, language comprehension can be seen as finding the most appropriate values for certain semantic features given the available syntactic and semantic properties of a usage. In formulating prediction, we use the same Bayesian framework that we used for learning, a formulation that takes into account both the general knowledge of language (i.e., constructions), and the specific behaviour of each individual verb.

6.1.3 Properties of the model

The novel representations used in our model, together with the unified formulation for language learning and use, lead to some desirable properties in the behaviour of the model. First of all, the model satisfies the major cognitive plausibility criteria that are expected from any model of child language acquisition. Learning in the model takes place in an incremental fashion: the input items are processed one by one, and the representation of argument structure constructions, as well as the profiles that represent verb semantic roles and selectional preferences, are updated accordingly. This fundamental property is lacking from many existing computational models of language learning, which use data in

batch mode to induce abstract patterns of language. Moreover, the model incorporates basic cognitive skills that human beings use in a range of tasks: analogy, categorization, and sensitivity to statistical properties of input.

In learning general regularities of language and their applicability on novel or low-frequency situations, the model relies solely on the input it receives from the environment. No innate knowledge of language is used for guiding the learning process: the general conceptions of verb argument structure constructions, semantic roles and selectional preferences are learned from verb usages only. In that regard, the proposed model is truly usage-based, and demonstrates the plausibility of a usage-based account in the domain of argument structure acquisition. Moreover, in learning the applicability of the acquired abstract knowledge, the model does not use any form of corrective feedback or negative evidence.

It has been shown through computational simulation that the behaviour of the model mirrors that of young children in some relevant aspects. The model goes through the same learning stages as children do: the conservative use of the more frequent usages for each individual verb at the beginning, followed by a phase when general patterns are grasped and applied overtly, which leads to occasional overgeneralization errors. Such errors cease to be made over time as the model processes more input. In learning the semantic roles, it is shown that, at first, the profile for each role reflects the properties of the particular arguments that participate in the early usages. This is in line with the early item-specific nature of the semantic roles proposed by the usage-based theories. However, the profiles transform to a more general probability distribution over the semantic properties of the arguments, matching the intuitive properties expected from each role. The same profile generalization process takes place for verb selectional preferences, only this time the final profiles represent the general preferences for a particular verb. Profiles learned for both semantic roles and selectional preferences are shown to be useful in various language tasks as observed in reported experimental data on human subjects, such as resolving

ambiguity in language comprehension and simulating human plausibility judgements.

In summary, we have presented a unified, probabilistic model of language learning and use, which demonstrates the feasibility of a usage-based account of language acquisition, and provides concrete explanation for the observed patterns in child language acquisition. Despite having many desirable properties, the model has certain limitations that should be addressed in future extensions. We will discuss some of these extensions in Section 6.2. However, the robust and flexible framework used in this model makes it an appropriate testbed for investigating a range of phenomena in the domain of child language acquisition, and suggests many new research directions. We will mention a number of these directions in Section 6.3.

6.2 Short-term Extensions

6.2.1 Using Child Data as Input

As input, our model requires a corpus of child-directed natural language sentences, annotated with the meaning that each sentence conveys. This annotation must include the predicate structure of the event described by the scene, as well as the semantic primitives of each predicate and the semantic properties of each argument in the described scene. Large scale corpora of utterances paired with such meaning representations do not currently exist. Therefore, in Chapters 3 and 4, we artificially generate the needed input based on the properties of existing child data. In Chapter 5, however, we automatically reconstruct the logical representation of the meaning of each sentence based on its parse tree from the British National Corpus (BNC, 2000). Recently, attempts have been made to develop appropriate tools for syntactic annotation of child-directed speech. For example, Sagae (2006) and Sagae et al. (2007) have developed an annotation scheme designed to represent syntactic information in CHILDES corpora, and a syntactic analysis system for annotating utterances from CHILDES transcripts. In the future, we can use such an-

notated data for reconstructing the predicate structure of the meaning of child-directed utterances, and evaluate the behaviour of our model in a more realistic setting.

6.2.2 Compressed Representation of Constructions

We represent each argument structure construction as a group of similar verb usages. From each such grouping of usages, the Bayesian formulation that we use induces a probabilistic association between syntactic and semantic features. Such associations are necessary both in learning, when the best cluster of usages has to be found for a new input item, and in use, when the best value for a missing feature in a usage must be found based on the groupings of the previously seen usages into different constructions. Therefore, every time the model receives a new verb usage to process, or every time it needs to perform some task of language use, it has to re-examine every single usage (or frame) that it has processed in the past. This property seriously affects the efficiency of the model, and makes it difficult to test the model with large amount of input, as discussed in Section 5.7. Moreover, it might be argued that this property makes the model less cognitively plausible, since the model must remember every usage it has seen in its lifespan.

A possible solution to this problem is to represent each construction not as a cluster of frames but instead as a profile in the form of a probability distribution over all the possible values for each feature in a frame. The distribution for each feature should reflect the observed values for that feature among all the usages that were at some point linked to this construction, and the construction profile can be updated every time a new usage is linked to it. As a result, the conditional probability $P_i(j|k)$ of a value j for a feature i in a construction k will be 'looked up' and not calculated every time, a factor that will improve the performance of the model significantly (especially for large constructions, and for expensive features such as semantic properties of an argument).

6.2.3 Sensitivity to the Order of Input

In the context of modelling child language acquisition, one of the strengths of the proposed model is its incremental nature. The model processes one input item at a time, and the constructions emerge and grow as a response to the input. However, this strategy results in a formation of constructions that is very sensitive to the order of the input items: the similarity between the early verb usages determines the number of the new constructions that are created. It often happens that one (intuitive) argument structure construction is divided between two clusters, or one single cluster contains usages from what we would recognize as two different argument structure constructions. The Bayesian formulation of the prediction model in language use allows for contribution from all of the constructions to each task, and therefore reduces the unwanted effect of an imperfect construction structure on the behaviour of the model. Nevertheless, the structure of the constructions does affect the output, and sometimes it is not possible for the model to overcome the early mistakes it has made even when provided with more input, a fact that was briefly discussed in Section 3.6.2.

Although it is not reasonable to assume that the structure of constructions should be revised after processing *each* new usage, it can be assumed that some constructions are revised and reorganized at certain intervals (a variation of this proposal is implemented in Parisien (2008)). In fact, it has been suggested that children consolidate and reorganize what they learn during their sleep (Dumay et al., 2004; Walker, 2005; Orban et al., 2006). Revision might consist of examining the coherence of a cluster, and considering the possibility of breaking it into two different clusters. It can also consist of measuring the similarity between two clusters, and considering the possibility of merging them together. The resulting model may be more robust against the order of input and noise.

Another way to reduce sensitivity to the order of input is to consider more than one partitioning of the observed frames into clusters at each point of time. That is, instead of permanently assigning each new frame to a cluster, we might use various ways of

clustering the input frames in our prediction model. A complete Bayesian formulation for prediction is not possible in this case for large data sets. However, the result of prediction can be approximated using repeated random sampling, or Monte Carlo method. Although most of the existing sampling methods process the data set in batch, incremental sampling methods have been proposed that can be used in cognitive models such as ours (Sanborn et al., 2006). Using a sampling method instead of the complete Bayesian formulation can also improve the efficiency of the model (see the previous section).

6.2.4 The Generalization Factor

In our model, generalization of any sort takes place when a construction that is not linked to the head verb of a frame is entrenched enough and contains frames that are similar enough to the current frame that it dominates the output of prediction. Therefore, generalization happens naturally, and as a result of the properties of the new item and of the input processed so far. However, it has been argued that the rate of generalization might be different among children. For example, it has been suggested that gender affects the rate of generalization of a familiar construction to a novel situation, as girls show more tendency towards using the general patterns than boys do (Hartshorne and Ulman, 2006). Therefore, a model such as ours must have an explicit parameter whose value influences the generalization rate in language use. In our model, the only parameter that can play such a role is λ in equation (2.6) on page 24. This parameter was originally introduced as a smoothing factor, but it also implicitly determines the importance of a distinct feature value in creating a new construction. Thus a smaller value for λ results in a higher probability for creating a new construction when processing a new frame that contains a previously unseen value for one of its features. This property influences generalization in our model. In a more coarse-grained construction structure, the possibility of generalization is higher: more well-entrenched constructions are formed, and thus there is a higher chance that a construction that is not linked to the head verb of

the current frame but contains similar enough frames to the current frame dominates the outcome of the prediction model. However, careful investigation of such a hypothesis is needed through monitoring various experiments with different values for λ, and analyzing the consequent behaviour of the model.

6.2.5 Argument Indexing in Syntactic Patterns and Semantic Representations

In Chapter 4, we attempted to show that it is possible to learn intuitive profiles for semantic roles without explicitly labeling each argument with the correct role in the semantic representation of the input scene. For this purpose, we removed the role labels from the input of the model. However, in the semantic representation that we use, there is a consistency in the ordering of the arguments receiving certain roles that can help the model to detect those roles in the input. For example, whenever there is an Agent and a Theme in a scene, the argument that plays the Agent role always appears first in the predicate structure representing the meaning, and the Theme argument appears second, as in the following examples:

John made cake.	$\mathbf{Make}_{[cause,become]}(\mathsf{JOHN}, \mathsf{CAKE})$
Mary ate cake.	$\mathbf{Eat}_{[act,consume]}(\mathsf{MARY}, \mathsf{CAKE})$

In all such cases, the syntactic pattern for the utterance is represented as 'arg1 *verb* arg2.' It is also assumed that 'arg1' in the syntactic pattern refers to the first argument that appears in the semantic representation (i.e., Agent), and 'arg2' refers to the second semantic argument (i.e., Theme). However, there is no evidence for such consistency of mapping between predicate structure and syntax in the representation available to children. In fact, in cases such as *Jim sold the house* and *Kim bought the house*, it is more reasonable to assume that the same underlying meaning representation is described in two different ways, and therefore 'arg1' in the following syntactic patterns should not

point to the same referent in the semantic representation:

TransferProperty$_{[act,change_possession]}$(JIM, HOUSE, KIM)

Jim sold the house. Syntactic Pattern: 'arg1 *verb* arg2'

Kim bought the house. Syntactic Pattern: 'arg1 *verb* arg2'

But in our model, these two sentences should either be linked to two different semantic representations so that the indexing of the arguments is consistent, or two different syntactic patterns should be used ('arg1 *sell* arg2' vs. 'arg3 *buy* arg2'). To make our claim on the learnability of semantic roles stronger, we need to employ a new representation for a syntactic pattern that is independent of the order of the arguments in the meaning representation.

6.3 Future Directions

6.3.1 Learning Word-Meaning Mapping

As one of our basic assumptions, we rely on the fact that the model knows the correct mapping between a word and its idiosyncratic meaning before embarking on the process of argument structure acquisition. For example, the model knows that the word *put* in an utterance represents the semantic symbol **Put** in the corresponding predicate structure, which itself represents the core meaning of the action 'putting'. In reality, this assumption is far from the truth. Children continue learning new words long after they have achieved the mastery of the structure of language. In fact, it has been suggested that knowing about the basic constructions of language can lead the child to pick the right meaning for a novel word in an ambiguous situation, a hypothesis known as Syntactic Bootstrapping (Landau and Gleitman, 1985). Some go as far as to propose that it is not possible for some words (especially verbs) to be learned without using any syntactic knowledge (e.g., Fisher, 1996). For example, more than one verb can be used to describe the same event from different perspectives (as in *the fox chased the rabbit* vs. *the rabbit fled the fox*).

In such situations, the only way a child can decide which perspective the unknown verb takes is to use the syntactic cues in the utterance, i.e., the word order or the preposition used (in addition to the semantic cues from the scene).

Our model provides a suitable testbed for demonstrating this phenomenon. The experimental setup should be modified to let the model process input usages that contain novel words as well as the familiar ones. The prediction model can be used for finding the more probable meanings for the novel word based on the available evidence. For example, consider the *chase/flee* case. Many verbs describe an event where the focus is on an Agent/Actor who moves away from a place/object/person. By receiving enough unambiguous usages of such verbs (i.e., *He left the room* or *She escaped the house*), the model learns that 'the participant who moves away' is indicated as the first argument preceding the verb. Similarly, by processing enough unambiguous usages of the verbs that focus on running after something, the model learns that for such verbs, 'the participant that runs after' is the argument preceding the verb. Therefore, upon hearing an utterance such as *The rabbit blicked the fox*, the model can use the fact that *rabbit* precedes *blicked* to deduce that the unknown verb must focus on the 'moving away from something' aspect of the scene, i.e., the fleeing interpretation. Therefore, it can pick the correct meaning for the novel verb. The model can be even more powerful in guessing (parts of) the meaning of a noun, using the knowledge it has acquired about the selectional preferences of the verb that takes the unknown noun as its argument. For example, if the model receives *She ate the blick*, it can use the semantic profile for the direct object position for *eat* to deduce highly probable properties of *blick* such as being a physical object and being edible.

Some have argued that the Syntactic Bootstrapping Hypothesis is circular: learning the basic constructions of a language depends on knowing the meaning of certain verbs and their arguments. But learning the meaning of verbs is impossible without knowing anything about syntax, since in each situation there are many different perspectives that

a verb can describe, and there is no easy way for the child to guess the intended meaning of the verb (e.g., Gleitman and Gillette, 1995; Fisher et al., 1994). However, it has been shown that single-word utterances form a significant portion of infant-directed speech (Brent and Siskind, 2001), which can help children acquire an initial vocabulary of both nouns and verbs without using any syntactic cues. Such a vocabulary can help them to start learning the argument structure constructions. This situation is similar to what happens in our model: the model has access to an initial vocabulary of basic words, each mapped with a meaning symbol, before it starts learning the argument structure constructions. Also, cross-situational learning, or detecting the meaning elements that are common across all the situational contexts of a verb usage, has proved to be helpful in disambiguating the intended meaning of a verb (Siskind, 1996; Yu et al., 2005). However, to our knowledge, no usage-based computational model has been proposed that shows the bi-directional bootstrapping between word-meaning learning and construction learning.

6.3.2 Language Processing

In some of the experiments reported in the previous chapters, we simulated limited language comprehension in the form of predicting the missing semantic primitives for a verb or semantic properties of an argument, or choosing between two possible semantic representations for an utterance. However, in order to simulate real language comprehension, we need a sentence processing module that is able to reconstruct a conceptual structure for a given natural language sentence.

Human language processing is a well-studied problem. Many consistent patterns have been observed in the experimental data gathered from both children and adults, which provide important hints to the underlying mechanisms. One of these phenomena is the garden-path effect: sentences such as *the horse raced pass the barn fell* are grammatical, but humans find them difficult to parse. On the other hand, certain clues such as the thematic roles or the selectional preferences of the verbs have been shown to affect sen-

tence processing (Trueswell et al., 1994; Nation et al., 2003). These findings can guide the development of a computational model of human language processing, especially in making critical decisions such as which parsing algorithms to use, how to resolve local and global ambiguities, and which information sources to integrate into the parsing process.

We believe that the probabilistic nature of both the verb-specific knowledge and the general constructions acquired by our model provides an interesting testbed for studying sentence processing in children. The entrenchment of the constructions, as well as their match with a partial sentence, can help the processing module to prune the partial parses, and find the most probable interpretations for a sentence at each point. Moreover, the acquired profiles for general semantic roles can be matched against the first words of a sentence and guide parsing even before the main verb of the sentence is heard.

6.3.3 Second Language Acquisition

A promising domain for applying the current model is second language acquisition. Many computational models have been proposed for studying the acquisition of bilingual lexicons and lexical categories (e.g., French, 1998; Li and Farkas, 2002). However, the computational modelling of the acquisition of the argument structure constructions for two different languages has received less attention (but see Rappoport and Sheinman, 2005).

Because the constructions acquired by our model are explicit and can be examined at different stages of learning, our model can be an ideal framework for studying many different factors in second language acquisition and their effect on the structure of the learned constructions. For example, it is interesting to see how the age of the acquisition of the second language is reflected in the acquired constructions: since usages of the different languages contain the same semantic features but different syntactic patterns, it is expected that at the very early stages of learning, distinct constructions are formed for each of the two languages (as a result of the languages having different values for

the syntactic pattern feature). However, if the second language usages are introduced to the model in a later stage, when familiar constructions of the first language are already established, it is more probable that the second language usages will be linked to the existing constructions. Other interesting phenomena that can be examined are gradual changes in language dominance, bilingual aphasia and recovery, and language decay, all of which can be traced in the output of the model.

6.3.4 Language Evolution

Natural languages undergo substantial change over generations. It has been suggested that language evolution occurs as a result of imperfect learning from the linguistic data received from the speech community: the dramatic changes in a language are due to the accumulation of small changes made by the individual speakers of that language (e.g., Mitchener and Nowak, 2004; Griffiths and Kalish, 2007). A usage-based, probabilistic model such as ours is an appropriate testbed for investigating this hypothesis. Generations of language learners can be simulated as sequenced instances of the model, each processing the production output of the previous instance as input data. The distributional properties of the 'evolved language', learned and produced by the last instance in the sequence, can be examined and compared against the original input to the first instance. Interesting patterns to look for would be whether some low-frequency words or constructions are dropped after a few generations, and whether new constructions, which at some point are introduced to one instance due to noise, are gained strength over time and passed through the generations.

Appendix A

Constructions in an Argument Structure Acquisition Simulation

A snapshot of the constructions learned by the model in one of the simulations in Chapter 3 is shown below. The constructions are learned after processing a total of 200 noisy input pairs. Each line below a construction header represents a frame that is grouped in that construction. The head word of the frame is shown in italic font. The following keywords are used to indicate different features in a frame:

- **freq**: the frequency of the frame
- **head-cat**: the semantic category of the head word
- **args**: the number of the arguments
- **roles**: the semantic roles of the arguments
- **arg-cats**: the semantic categories of the arguments
- **prims**: the semantic primitives of the head word
- **patt**: the syntactic pattern

Construction 1

get(**freq**=2;**head-cat**=predicate;**args**=2;**roles**={theme state};**arg-cats**={animal adjective};**prims**={changestate};**patt**='arg1 word arg2')
get(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={theme state};**arg-cats**={human adjective};**prims**={changestate};**patt**='arg1 word arg2')

Construction 2

see(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={human concrete};**prims**={perceive};**patt**='')
and(**freq**=1;**head-cat**=and-pred;**args**=2;**roles**={unknown unknown};**arg-cats**={edible edible};**prims**={};**patt**='arg1 word arg2')
and(**freq**=1;**head-cat**=and-pred;**args**=2;**roles**={unknown unknown};**arg-cats**={animal meal};**prims**={};**patt**='kitty word doggy and arg2')
look(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={animal unknown};**prims**={act perceive};**patt**='arg1 word arg2')
see(**freq**=3;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={animal unknown};**prims**={perceive};**patt**='arg1 word arg2')
see(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={animal game};**prims**={perceive act};**patt**='arg1 word arg2')
see(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={human unknown};**prims**={perceive};**patt**='arg1 word arg2')
and(**freq**=1;**head-cat**=and-pred;**args**=2;**roles**={unknown unknown};**arg-cats**={unknown meal};**prims**={};**patt**='arg1 word arg2')
look(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={human tool};**prims**={act perceive};**patt**='arg1 word arg2')
see(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={human human};**prims**={perceive act};**patt**='arg1 word arg2')
see(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={unknown stimulus};**arg-cats**={human tool};**prims**={perceive};**patt**='arg1 word arg2')
look(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={human at-pred};**prims**={};**patt**='arg1 word arg2')
see(**freq**=15;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={animate thing};**prims**={perceive};**patt**='arg1 word arg2')
and(**freq**=1;**head-cat**=and-pred;**args**=2;**roles**={unknown unknown};**arg-cats**={tool edible};**prims**={};**patt**='arg1 word arg2')
and(**freq**=1;**head-cat**=and-pred;**args**=2;**roles**={unknown unknown};**arg-cats**={tool animal};**prims**={};**patt**='arg1 word arg2')
see(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={unknown meal};**prims**={perceive};**patt**='arg1 word arg2')
see(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={animal edible};**prims**={};**patt**='arg1 word arg2')
look(**freq**=2;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={animal at-pred};**prims**={act perceive};**patt**='arg1 word arg2')
look(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={human at-pred};**prims**={act perceive};**patt**='arg1 word arg2')
look(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={experiencer stimulus};**arg-cats**={animal edible};**prims**={act perceive};**patt**='arg1 word arg2')

Construction 3

in(**freq**=1;**head-cat**=loc-pred;**args**=1;**roles**={destination};**arg-cats**={unknown};**prims**={};**patt**='word arg1')
on(**freq**=1;**head-cat**=loc-pred;**args**=1;**roles**={destination};**arg-cats**={animal};**prims**={};**patt**='word arg1')
on(**freq**=1;**head-cat**=loc-pred;**args**=1;**roles**={destination};**arg-cats**={tool};**prims**={};**patt**='word arg1')
at(**freq**=1;**head-cat**=at-pred;**args**=1;**roles**={destination};**arg-cats**={game};**prims**={};**patt**='word arg1')
to(**freq**=1;**head-cat**=dest-pred;**args**=1;**roles**={destination};**arg-cats**={human};**prims**={};**patt**='word arg1')
into(**freq**=1;**head-cat**=dest-pred;**args**=1;**roles**={destination};**arg-cats**={game};**prims**={};**patt**='word arg1')
to(**freq**=1;**head-cat**=dest-pred;**args**=1;**roles**={destination};**arg-cats**={game};**prims**={act};**patt**='word arg1')
to(**freq**=1;**head-cat**=dest-pred;**args**=1;**roles**={destination};**arg-cats**={animal};**prims**={};**patt**='word arg1')
on(**freq**=1;**head-cat**=loc-pred;**args**=1;**roles**={destination};**arg-cats**={unknown};**prims**={};**patt**='word arg1')
at(**freq**=1;**head-cat**=at-pred;**args**=1;**roles**={destination};**arg-cats**={tool};**prims**={};**patt**='word arg1')
to(**freq**=4;**head-cat**=dest-pred;**args**=1;**roles**={destination};**arg-cats**={concrete};**prims**={};**patt**='word arg1')
into(**freq**=1;**head-cat**=dest-pred;**args**=1;**roles**={destination};**arg-cats**={unknown};**prims**={};**patt**='word arg1')
in(**freq**=1;**head-cat**=loc-pred;**args**=1;**roles**={destination};**arg-cats**={animal};**prims**={};**patt**='word arg1')
in(**freq**=7;**head-cat**=loc-pred;**args**=1;**roles**={destination};**arg-cats**={concrete};**prims**={};**patt**='word arg1')

Construction 4

play(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={human};**prims**={act move playfully cause};**patt**='arg1 word')
come(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animal};**prims**={act move};**patt**='arg1 word')
play(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animal};**prims**={act playfully};**patt**='')
play(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={human};**prims**={act playfully};**patt**='')

come(**freq**=2;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={human};**prims**={act move};**patt**='arg1 word')
go(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animal};**prims**={act move};**patt**='arg1 word')
eat(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animal};**prims**={act playfully move cause};**patt**='arg1 word')
go(**freq**=2;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={human};**prims**={act move};**patt**='arg1 word')
go(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={unknown};**arg-cats**={human};**prims**={act move};**patt**='arg1 word')
come(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animal};**prims**={act move playfully cause};**patt**='arg1 word')
play(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animal};**prims**={};**patt**='arg1 word')
dance(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={unknown};**arg-cats**={animal};**prims**={act manner};**patt**='arg1 word')
play(**freq**=4;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animate};**prims**={act playfully};**patt**='arg1 word')
eat(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={animal};**prims**={act consume};**patt**='arg1 word')
eat(**freq**=1;**head-cat**=predicate;**args**=1;**roles**={agent};**arg-cats**={human};**prims**={act consume};**patt**='arg1 word')

Construction 5

fall(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animal unknown};**prims**={move direction vertical};**patt**='arg2 word')
fall(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={human tool};**prims**={move direction vertical};**patt**='arg2 word')

Construction 6

get(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animal tool};**prims**={};**patt**='arg1 word arg2')
get(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animal meal};**prims**={act possess};**patt**='arg1 word arg2')
get(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={human tool};**prims**={act possess};**patt**='arg1 word arg2')
eat(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={human meal};**prims**={act consume};**patt**='arg1 word arg2')
get(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animal tool};**prims**={act possess};**patt**='arg1 word arg2')
eat(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={human edible};**prims**={act consume};**patt**='arg1 word arg2')
eat(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={unknown theme};**arg-cats**={human meal};**prims**={act consume};**patt**='arg1 word arg2')
get(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animal edible};**prims**={act possess};**patt**='arg1 word arg2')
get(**freq**=4;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={human edible};**prims**={act possess};**patt**='arg1 word arg2')
make(**freq**=6;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animate concrete};**prims**={cause become};**patt**='arg1 word arg2')
get(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animal concrete};**prims**={act possess};**patt**='')
take(**freq**=6;**head-cat**=predicate;**args**=2;**roles**={agent theme};**arg-cats**={animate concrete};**prims**={cause possess};**patt**='arg1 word arg2')

Construction 7

give(**freq**=3;**head-cat**=predicate;**args**=3;**roles**={agent theme beneficiary};**arg-cats**={animal meal human};**prims**={cause possess};**patt**='arg1 word arg3 arg2')
give(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme beneficiary};**arg-cats**={human tool animal};**prims**={cause possess};**patt**='arg1 word arg3 arg2')
give(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme beneficiary};**arg-cats**={human meal animal};**prims**={cause possess};**patt**='')
give(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme beneficiary};**arg-cats**={animal tool human};**prims**={cause possess};**patt**='arg1 word arg3 arg2')
give(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme beneficiary};**arg-cats**={human meal animal};**prims**={cause possess};**patt**='arg1 word arg3 arg2')
give(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme beneficiary};**arg-cats**={human tool dest-pred};**prims**={};**patt**='arg1 word arg2 arg3')

Construction 8

put(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={human concrete game};**prims**={cause move};**patt**='arg1 word arg2 arg3')
put(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={human tool loc-pred};**prims**={cause move};**patt**='arg1 word arg2 arg3')
put(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={human unknown loc-pred};**prims**={cause move};**patt**='arg1 word arg2 arg3')
take(**freq**=2;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={animal meal dest-pred};**prims**={cause move};**patt**='arg1 word arg2 arg3')
take(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={human edible dest-pred};**prims**={cause move};**patt**='arg1 word arg2 arg3')

go(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={human concrete loc-pred};**prims**={};**patt**='arg2 word arg3')
take(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={animal concrete dest-pred};**prims**={cause move};**patt**='arg1 word arg2 arg3')
put(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={human meal loc-pred};**prims**={move cause};**patt**='arg1 word arg2 arg3')
put(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={animal edible unknown};**prims**={cause move};**patt**='arg1 word arg2 arg3')
put(**freq**=7;**head-cat**=predicate;**args**=3;**roles**={agent theme destination};**arg-cats**={animal concrete loc-pred};**prims**={cause move};**patt**='arg1 word arg2 arg3')

Construction 9
go(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent destination};**arg-cats**={animal unknown};**prims**={act move};**patt**='arg1 word arg2')
go(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent destination};**arg-cats**={unknown dest-pred};**prims**={act move};**patt**='arg1 word arg2')
go(**freq**=8;**head-cat**=predicate;**args**=2;**roles**={agent destination};**arg-cats**={animate dest-pred};**prims**={act move};**patt**='arg1 word arg2')
come(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent destination};**arg-cats**={human dest-pred};**prims**={act move};**patt**='')
come(**freq**=1;**head-cat**=predicate;**args**=2;**roles**={agent destination};**arg-cats**={human dest-pred};**prims**={act move};**patt**='arg1 word arg2')

Construction 10
for(**freq**=1;**head-cat**=for-pred;**args**=1;**roles**={theme};**arg-cats**={meal};**prims**={};**patt**='word arg1')
for(**freq**=1;**head-cat**=for-pred;**args**=1;**roles**={theme};**arg-cats**={human};**prims**={move act};**patt**='word arg1')

Construction 11
sit(**freq**=2;**head-cat**=predicate;**args**=2;**roles**={agent agent};**arg-cats**={animal loc-pred};**prims**={act rest};**patt**='arg1 word arg2')

Construction 12
make(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={stimulus theme state};**arg-cats**={tool concrete adjective};**prims**={cause changestate};**patt**='arg1 word arg2 arg3')
make(**freq**=1;**head-cat**=predicate;**args**=3;**roles**={stimulus experiencer state};**arg-cats**={meal human adjective};**prims**={cause changestate};**patt**='arg1 word arg2 arg3')

Construction 13
fall(**freq**=2;**head-cat**=predicate;**args**=1;**roles**={theme};**arg-cats**={meal};**prims**={move direction vertical};**patt**='arg1 word')

Appendix B

Constructions in a Semantic Role Learning Simulation

A snapshot of the constructions learned by the model in one of the simulations in Chapter 4 is shown below. The constructions are learned after processing a total of 200 noisy input pairs. Each line below a construction header represents a frame that is grouped in that construction. The head word of the frame is shown in italic font. The following keywords are used to indicate different features in a frame:

- **freq**: the frequency of the frame
- **args**: the number of the arguments
- **prims**: the semantic primitives of the head word
- **patt**: the syntactic pattern

Due to limited space, the semantic properties of the arguments are not shown.

Construction 1

to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
under(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
into(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
for(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
with(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
with(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
of(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
for(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
into(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
of(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
into(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
into(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
of(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
of(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
to(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
from(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
of(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
at(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
into(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
with(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
on(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')
in(**freq**=1; **args**=1; **prims**={}; **patt**='word arg1')

Construction 2

get(**freq**=1; **args**=2; **prims**={changestate}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={changestate}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={changestate}; **patt**='arg1 word arg2')

Construction 3

see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')

see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={perceive}; **patt**='arg1 word arg2')

Construction 4

make(**freq**=1; **args**=2; **prims**={cause become}; **patt**='arg1 word arg2')
make(**freq**=1; **args**=2; **prims**={cause become}; **patt**='arg1 word arg2')
take(**freq**=1; **args**=2; **prims**={cause possess}; **patt**='arg1 word arg2')
take(**freq**=1; **args**=2; **prims**={cause possess}; **patt**='')
make(**freq**=1; **args**=2; **prims**={cause become}; **patt**='arg1 word arg2')
make(**freq**=1; **args**=2; **prims**={cause become}; **patt**='arg1 word arg2')

Construction 5

come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
play(**freq**=1; **args**=2; **prims**={act playfully}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
play(**freq**=1; **args**=2; **prims**={act playfully}; **patt**='arg1 word arg2')
come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
play(**freq**=1; **args**=2; **prims**={act playfully}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='')
come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
look(**freq**=1; **args**=2; **prims**={act perceive}; **patt**='arg1 word arg2')
come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')

go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
play(**freq**=1; **args**=2; **prims**={act playfully}; **patt**='arg1 word arg2')
come(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')
go(**freq**=1; **args**=2; **prims**={act move}; **patt**='arg1 word arg2')

Construction 6

take(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
go(**freq**=1; **args**=3; **prims**={move}; **patt**='arg2 word to arg1')
take(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
go(**freq**=1; **args**=3; **prims**={move}; **patt**='arg2 word arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
take(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
give(**freq**=1; **args**=3; **prims**={}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='')
put(**freq**=1; **args**=3; **prims**={}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={}; **patt**='arg1 word arg2 arg3')
go(**freq**=1; **args**=3; **prims**={move}; **patt**='arg2 word arg3')
take(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
go(**freq**=1; **args**=3; **prims**={move}; **patt**='arg2 word arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
give(**freq**=1; **args**=3; **prims**={cause possess}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
take(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
give(**freq**=1; **args**=3; **prims**={cause possess}; **patt**='arg1 word arg2 arg3')
take(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')
put(**freq**=1; **args**=3; **prims**={cause move}; **patt**='arg1 word arg2 arg3')

Construction 7

and(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
and(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
take(**freq**=1; **args**=2; **prims**={cause possess}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
take(**freq**=1; **args**=2; **prims**={cause possess}; **patt**='')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='')
take(**freq**=1; **args**=2; **prims**={cause possess}; **patt**='arg1 word arg2')
eat(**freq**=1; **args**=2; **prims**={act consume}; **patt**='arg1 word arg2')

get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={act possess}; **patt**='arg1 word arg2')
take(**freq**=1; **args**=2; **prims**={cause possess}; **patt**='arg1 word arg2')

Construction 8

dance(**freq**=1; **args**=1; **prims**={act manner}; **patt**='arg1 word')
play(**freq**=1; **args**=1; **prims**={act playfully}; **patt**='arg1 word')
come(**freq**=1; **args**=1; **prims**={act move}; **patt**='arg1 word')
play(**freq**=1; **args**=1; **prims**={act playfully}; **patt**='arg1 word')
fall(**freq**=1; **args**=1; **prims**={move direction vertical}; **patt**='arg1 word')
fall(**freq**=1; **args**=1; **prims**={move direction vertical}; **patt**='arg1 word')
go(**freq**=1; **args**=1; **prims**={act move}; **patt**='arg1 word')
come(**freq**=1; **args**=1; **prims**={act move}; **patt**='arg1 word')
play(**freq**=1; **args**=1; **prims**={act playfully}; **patt**='arg1 word')
dance(**freq**=1; **args**=1; **prims**={}; **patt**='arg1 word')
come(**freq**=1; **args**=1; **prims**={act move}; **patt**='arg1 word')
play(**freq**=1; **args**=1; **prims**={act playfully}; **patt**='arg1 word')
play(**freq**=1; **args**=1; **prims**={act playfully}; **patt**='arg1 word')
go(**freq**=1; **args**=1; **prims**={act move}; **patt**='arg1 word')
come(**freq**=1; **args**=1; **prims**={act move}; **patt**='arg1 word')

Construction 9

give(**freq**=1; **args**=3; **prims**={cause possess}; **patt**='arg1 word arg3 arg2')
give(**freq**=1; **args**=3; **prims**={cause possess}; **patt**='arg1 word arg3 arg2')
give(**freq**=1; **args**=3; **prims**={cause possess}; **patt**='arg1 word arg3 arg2')

Construction 10

make(**freq**=1; **args**=3; **prims**={cause changestate}; **patt**='arg1 word arg2 arg3')
make(**freq**=1; **args**=3; **prims**={cause changestate}; **patt**='arg1 word arg2 arg3')

Construction 11

and(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')

Construction 12

see(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
get(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
make(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')
see(**freq**=1; **args**=2; **prims**={}; **patt**='arg1 word arg2')

Bibliography

Abney, S. and Light, M. (1999). Hiding a semantic hierarchy in a Markov model. In *Proceedings of the ACL Workshop on Unsupervised Learning in Natural Language Processing*.

Akhtar, N. (1999). Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26:339–356.

Allen, J. (1997). Probabilistic constraints in acquisition. In Sorace, A., Heycock, C., and Shillcock, R., editors, *Proceedings of the GALA97 Conference on Language Acquisition, Edinburth*, pages 300–305.

Allen, J. and Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In MacWhinney, B., editor, *Emergence of Language*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429.

Antoniak, C. (1974). Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics*, 2(6):1152–1174.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (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*.

Baker, M. C. (1988). *Incorporation: a theory of grammatical function changing*. Chicago: University of Chicago Press.

Bates, E. and MacWhinney, B. (1989). Functionalism and the competition model. In MacWhinney, B. and Bates, E., editors, *The crosslinguistic study of sentence processing*, pages 3–76. New York: Cambridge University Press.

Bencini, G. M. L. and Goldberg, A. E. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language*, 43:640–651.

BNC (2000). *Reference guide for the British National Corpus (world edition)*, second edition.

Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: implications of developmental errors with causative verbs. *Quaderni di semantica*, 3:5–66.

Bowerman, M. (1988). The 'no negative evidence' problem. In Hawkins, J., editor, *Explaining language universals*. Londen: Blackwell.

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: are children helped by innate linking rules? *Linguistics*, 28:1253–1289.

Bowerman, M. (1996). Argument structure and learnability: is a solution in sight? In *Proceedings of the Berkeley Linguistics Society*, volume 22, pages 454–468.

Braine, M. D. S. and Brooks, P. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In Tomasello, M. and Merriman, W. E., editors, *Beyond names for things: young children's acquisition of verbs*. Hillsdale, NJ: Lawrence Erlbaum.

Brent, M. R. and Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2):B33–B44.

Bresnan, J. (2001). *Lexical functional syntax*. Blackwell Textbooks in Linguistics. Blackwell Publishers.

Brockmann, C. and Lapata, M. (2003). Evaluating and combining approaches to selectional preference acquisition. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*.

Brown, R. and Hanlon, C. (1970). Derivational complexity and order of acquisition of syntax. *Cognition and the Development of Language, New York; Wiley*, pages 11–53.

Buttery, P. (2003). A computational model for first language acquisition. In *Proceedings of CLUK6, Edinburgh*, pages 1–8.

Buttery, P. (2004). A quantitative evaluation of naturalistic models of language acquisition: the efficiency of the Triggering Learning Algorithm compared to a Categorial Grammar Learner. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 1–6.

Cartwright, T. A. and Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition*, 63:121–170.

Chang, N. (2004). Putting meaning into grammar learning. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 17–24.

Childers, J. B. and Tomasello, M. (2001). The role of pronouns in young children's acquisition of the english transitive construction. *Developmental Psychology*, 37(6):739–748.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.

Chomsky, N. (1980). *Rules and representations*. Oxford: Basil Blackwell.

Chomsky, N. (1981). *Lectures on government and binding*. Mouton de Gruyter.

Chouinard, M. M. and Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30(3):637–669.

Ciaramita, M. and Johnson, M. (2000). Explaining away ambiguity: learning verb selectional preference with Bayesian networks. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*.

Clark, A. (2001a). Unsupervised induction of stochastic context free grammars with distributional clustering. In *Proceedings of CoNLL, Toulouse, France*, pages 105–112.

Clark, A. (2001b). *Unsupervised language acquisition: theory and practice*. PhD thesis, University of Sussex.

Clark, A. (2006). PAC-learning unambiguous NTS languages. In *Proceedings of the 8th international colloquium on grammatical inference*, pages 59–71. Springer, Berlin.

Clark, S. and Weir, D. (2002). Class-based probability estimation using a semantic hierarchy. *Computational Linguistics*, 28(2):187–206.

Collins, M. (1999). *Head-driven statistical models for natural language parsing*. PhD thesis, University of Pennsylvania.

Copestake, A. and Briscoe, T. (1991). Lexical operations in a unification-based framework. *Lecture Notes in Computer Science*, 627:101–119.

Demuth, K., Machobane, M., Moloi, F., and Odato, C. (2002). Rule learning and lexical frequency effects in learning verb-argument structure. In *Proceedings of the 26th Annual Boston University Conference on Language Development*, pages 142–153.

Desai, R. (2002). Bootstrapping in miniature language acquisition. In *Proceedings of the 4th International Conference on Cognitive Modelling*.

Dominey, P. F. (2003). Learning grammatical constructions in a miniature language from narrated video events. In *Proceedings of the 25nd Annual Conference of the Cognitive Science Society*.

Dominey, P. F. and Inui, T. (2004). A developmental model of syntax acquisition in the construction grammar framework with cross-linguistic validation in English and Japanese. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 33–40.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3):547–619.

Dumay, N., Gaskell, G., and Feng, X. (2004). A day in the life of a spoken word. In *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*, pages 339–344.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.

Elman, J. L. (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning*, 7:195–225.

Erk, K. (2007). A simple, similarity-based model for selectional preferences. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 216–223, Prague, Czech Republic.

Falk, Y. N. (2001). *Lexical-Functional Grammar: an introduction to parallel constraint-based syntax*. CSLI Publications.

Fazly, A. (2007). *Automatic acquisition of lexical knowledge about multiword predicates*. PhD thesis, Department of Computer Science, University of Toronto.

Fern, A., Givan, R., and Siskind, J. M. (2002). Specific-to-general learning for temporal

events with application to learning event definitions from video. *Journal of Artificial Intelligence Research*, 17:379–449.

Fillmore, C. (1968). The case for case. In Back, E. and Harms, R. J., editors, *Universals in linguistic theory*, pages 1–88. Holt, Rinehart and Winston.

Fillmore, C., Kay, P., and O'Connor, M. K. (1988). Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language*, 64:501–538.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, 31(1):41–81.

Fisher, C. (2002). Structural limits on verb mapping: the role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, 5(1):55–64.

Fisher, C., Hall, G., Rakowitz, S., and Gleitman, L. (1994). When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92(1):333–375.

Fleischman, M. and Roy, D. (2005). Intentional context in situated language learning. In *Ninth Conference on Computational Natural Language Learning*.

French, R. M. (1998). A simple recurrent network model of bilingual memory. In *Proceedings of the Twentieth Annual Cognitive Science Society Conference*, pages 368–373.

Gildea, D. and Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 23(3):245–288.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1:135–176.

Gleitman, L. and Gillette, J. (1995). The role of syntax in verb learning. In *Handbook of child language*. Oxford: Blackwell.

Gobet, F., Freudenthal, D., and Pine, J. M. (2004). Modelling syntactic development in a cross-linguistic context. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 53–60.

Goldberg, A. (2006). *Constructions at work: the nature of generalization in language*. Oxford University Press.

Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. The University of Chicago Press.

Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In MacWhinney, B., editor, *Emergence of language*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40:237–264.

Griffiths, T. and Kalish, M. (2007). Language Evolution by Iterated Learning With Bayesian Agents. *Cognitive Science*, 31(3):441–48040.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Hacioglu, K., Pradhan, S., Ward, W., Martin, J., and Jurafsky, D. (2004). Semantic role labeling by tagging syntactic chunks. In *Proceedings of CoNLL 2004 Shared Task*.

Hartshorne, J. K. and Ulman, M. T. (2006). Why girls say 'holded' more than boys. *Developmental Science*, 9(1):21–32.

Holmes, V. M., Stowe, L., and Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language*, 28:668–689.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. MIT Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

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

Jones, G., Gobet, F., and Pine, J. M. (2000). A process model of children's early verb use. In *Proceedings of the 22th Annual Conference of the Cognitive Science Society*.

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, 101:1–42.

Kipper, K., Dang, H. T., and Palmer, M. (2000). Class-based construction of a verb lexicon. In *Proceedings of the 17th AAAI Conference*, pages 691–696.

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

Landau, B. and Gleitman, L. R. (1985). *Language and experience: evidence from the blind child*. Cambridge, MA: Harvard UP.

Langacker, R. (1999). *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

Lappin, S. and Shieber, S. M. (2007). Machine learning theory and practice as a source of insight into universal grammar. *Journal of Linguistics*, 43:1–34.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. The University of Chicago Press.

Li, H. and Abe, N. (1998). Generalizing case frames using a thesaurus and the MDL principle. *Computational Linguistics*, 24(2):217–244.

Li, P. and Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In Heredia, R. and Altarriba, J., editors, *Bilingual sentence processing*. Elsevier.

Lieven, E., Pine, J., and Rowland, G. (1998). Comparing different models of the development of the English verb category. *Linguistics*, 36(4):807–830.

Lieven, E. V. M., Pine, J. M., and Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24:187–219.

Light, M. and Greiff, W. (2002). Statistical models for the induction and use of selectional preferences. *Cognitive Science*, 26(3):269–281.

MacWhinney, B. (1982). Basic syntactic processes. In Kuczaj, S., editor, *Language development: syntax and semantics*, volume 1, pages 73–136. Hillsdale, N.J., Lawrence Erlbaum.

MacWhinney, B. (1987). The competition model. In MacWhinney, B., editor, *Mechanisms of language acquisition*. Hillsdale, NJ: Erllbaum.

MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates, second edition.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31:883–914.

Manning, C. D. and Schütze, H. (2001). *Foundations of statistical natural language processing*. The MIT Press.

Maratsos, M., Gudeman, R., Gerard-Ngo, P., and DeHart, G. (1987). A study in novel word learning: the productivity of the causative. In MacWhinney, B., editor, *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, 46:53–85.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., and Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, 57(4, Serial No. 228).

Matessa, M. and Anderson, J. R. (2000). Modelling focused learning in role assignment. *Language and Cognitive Processes*, 15(3):263–292.

McCarthy, D. and Carroll, J. (2003). Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. *Computational Linguistics*, 29(4):639–654.

McClelland, J. L. and Kawamoto, A. H. (1986). *Mechanisms of sentence processing: assigning roles to constituents of sentences*, pages 272–325. MIT Press Cambridge, MA, USA.

McNeill, D. (1966). The creation of language by children. *Psycholinguistic papers*, pages 99–115.

McRae, K., Ferretti, T. R., and Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, 12(2/3):137–176.

Miller, G. (1990). WordNet: an on-line lexical database. *International Journal of Lexicography*, 17(3).

Mitchener, W. and Nowak, M. (2004). Chaos and language. *Proceedings: Biological Sciences*, 271(1540):701–704.

Morris, W. C., Cottrell, G. W., and Elman, J. L. (2000). A connectionist simulation of the empirical acquisition of grammatical relations. In Wermter, S. and Sun, R., editors, *Hybrid neural symbolic integration*. Springer Verlag.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17:357–374.

Nation, K., Marshall, C. M., and Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psych.*, 86:314–329.

Neal, R. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9(2):249–265.

Ninio, A. (2006). *Language and the learning curve: a new theory of syntactic development*. Oxford University Press.

Niyogi, S. (2002). Bayesian learning at the syntax-semantics interface. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*.

Onnis, L., Roberts, M., and Chater, N. (2002). Simplicity: a cure for overgeneralizations in language acquisition? In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 720–725.

Orban, P., Rauchs, G., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., and Peigneux, P. (2006). Sleep after spatial learning promotes covert reorganization of brain activity. *Proceedings of the National Academy of Sciences of the USA*, 103(18):7124–7129.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Parisien, C. (2008). An incremental bayesian model for learning syntactic categories. Master's thesis, Department of Computer Science, University of Toronto.

Pereira, F. (2000). Formal grammar and information theory: together again? In *Philosophical Transactions of the Royal Society*, pages 1239–1253. Royal Society, London.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics? *Lingua*, 92:377–410.

Pinker, S., Lebeaux, D. S., and Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition*, 26:195–267.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Quine, W. (1960). *Word and object*. Cambridge, MA: MIT Press.

Rappaport Hovav, M. and Levin, B. (1998). Building verb meanings. In *The projection of arguments: lexical and computational factors*, pages 97–134. CSLI Publications, Stanford, CA.

Rappoport, A. and Sheinman, V. (2005). A second language acquisition model using example generalization and concept categories. In *Proceedings of the Second Workshop on Psychocomputational Models of Human Language Acquisition*, pages 45–52.

Resnik, P. (1993). *Selection and information: A class-based approach to lexical relationships*. PhD thesis, University of Pennsylvania.

Resnik, P. (1996). Selectional constraints: an information-theoretic model and its computational realization. *Cognition*, 61:127–199.

Resnik, P. (1997). Selectional preference and sense disambiguation. In *Proceedings of ACL Siglex Workshop on Tagging Text with Lexical Semantics, Why, What and How?, Washington, April 4-5, 1997.*

Riloff, E. and Schmelzenbach, M. (1998). An empirical approach to conceptual case frame acquisition. In *Proceedings of the 6th Workshop on Very Large Corpora*, pages 49–56.

Rohde, D. L. T. (2004). TGrep2 user manual.

Sagae, K. (2006). *A multi-strategy approach to parsing of grammatical relations in transcripts of parent-child dialogs*. PhD thesis, Carnegie Mellon University.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., and Wintner, S. (2007). High-accuracy annotation and parsing of CHILDES transcripts. In *Proceedings of the ACL-2007 Workshop on Cognitive Aspects of Computational Language Acquisition*.

Sanborn, A., Griffiths, T., and Navarro, D. (2006). A more rational model of categorization. *Proceedings of the Twenty-eighth Annual Conference of the Cognitive Science Society*.

Saxton, M. (2000). Negative evidence and negative feedback: immediate effects on the grammaticality of child speech. *First Language*, 20(60):221.

Shayan, S. and Gershkoff-Stowe, L. (2007). How do children learn Agent and Patient roles? In *Poster presentation at SRCD Biennial Meeting Boston, Massachusetts, USA*.

Siskind, J. M. (1994). Lexical acquisition in the presence of noise and homonymy. In *Proceedings of the American Association of Artificial Intelligence AAAI-94*, pages 760–766.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91.

Solan, Z., Horn, D., Ruppin, F., and Edelman, S. (2004). Unsupervised context sensitive language acquisition from a large corpus. In Saul, L., editor, *Advances in Neural Information Processing*, volume 16. Cambridge, MA, MIT Press.

Swier, R. and Stevenson, S. (2004). Unsupervised semantic role labelling. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Barcelona, Spain*, pages 95–102.

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development*, 19:15–34.

Thompson, C. A. and Mooney, R. J. (1999). Automatic construction of semantic lexicons for learning natural language interfaces. In *Proceedings of the 17th National Conference on Artificial Intelligence*, Orlando, FL.

Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition*, 74:209–253.

Tomasello, M. (2000b). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, 4(4):156–163.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. and Abbot-Smith, K. (2002). A tale of two theories: response to Fisher. *Cognition*, 83:207–214.

Tomasello, M., Akhtar, N., Dodson, K., and Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language*, 24:373–387.

Trueswell, J. C., Tanenhaus, M. K., and Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33(3):285–318.

VanValin, R. D. and LaPolla, R. J. (1997). *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.

Walker, M. P. (2005). A refined model of sleep and the time course of memory formation. *Behavioural and Brain Sciences*, 28:51–104.

Yu, C., Ballard, D. H., and Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, 29:961–1005.