

# A Computational Model for Early Argument Structure Acquisition

Afra Alishahi and Suzanne Stevenson  
Department of Computer Science  
University of Toronto

## Abstract

How children go about learning the general regularities that govern language, as well as keeping track of the exceptions to them, remains one of the challenging open questions in the cognitive science of language. Computational modeling is an important methodology in research aimed at addressing this issue. We must determine appropriate learning mechanisms that can grasp generalizations from examples of specific usages, and that exhibit patterns of behaviour over the course of learning similar to those in children. Early learning of verb argument structure is an area of language acquisition that provides an interesting testbed for such approaches due to the complexity of verb usages. A range of linguistic factors interact in determining the felicitous use of a verb in various *constructions*—associations between syntactic forms and properties of meaning, that form the basis for a number of linguistic and psycholinguistic theories of language. We present a computational model for the representation, acquisition, and use of verbs and constructions. Our Bayesian framework is founded on a novel view of constructions as a probabilistic association between syntactic and semantic features. The computational experiments reported here demonstrate the feasibility of learning general constructions, and their exceptions, from individual usages of verbs. The behaviour of the model over the timecourse of acquisition mimics in relevant aspects the stages of learning exhibited by children. Our proposal thus sheds light on the possible mechanisms at work in forming linguistic generalizations and maintaining knowledge of exceptions.

## 1 Introduction

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.

One aspect of language that provides a rich testbed for exploring such issues is the usage of verbs—specifically, what syntactic constructions a verb can occur in. 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. In particular, verb argument structure is a complex aspect of language for a child to master, as it requires learning the relations of arguments to a verb, and how those arguments are mapped into

valid expressions of the language. In this paper, we describe a computational model that mimics early acquisition of verb argument structure in a child, and sheds light on the possible mechanisms at work in forming generalizations and maintaining knowledge of exceptions.

## 1.1 Child Learning of Verb Argument Structure

Appropriate usage of verbs is guided by the knowledge of verb argument structure. 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).

- |  |  |
|--|--|
| <b>1a)</b> The chain dropped into the box. | <b>1b)</b> The girl dropped the chain. |
| <b>2a)</b> The ball rolled into the room.  | <b>2b)</b> The child rolled the ball.  |
| <b>3a)</b> The book fell into the fire.    | <b>3b)</b> *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, 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 does not play a role in children’s cessation of incorrect forms.

## 1.2 Theories of Argument Structure Acquisition

The observed pattern of argument structure acquisition leads to two questions that must be answered in a theory of early verb learning:

- 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 (i.e., the U-shaped learning curve)?

In addition to theoretical proposals, explicit computational models that instantiate a coherent set of answers to these questions must be explored, in order to demonstrate the cognitive feasibility of an account of the data.

With regard to the first question above, learning in this domain has been suggested to rely on rich innate knowledge of argument structure regularities—in particular, mappings of semantics arguments to syntactic positions (e.g., Pinker, 1989; Grimshaw, 1990). However, a number of usage-based proposals have argued that children learn such regularities from the input alone, without guidance in the form of innate principles (e.g., Bowerman, 1982; Akhtar, 1999; Tomasello, 2000). Moreover, recent psycholinguistic evidence supports the view that children initially learn verb-argument patterns on an item-by-item (verb-by-verb) basis, before forming a conceptualization of more general syntactic structures (e.g., MacWhinney, 1982, 1987; Tomasello, 2000, 2003). Given this evidence, the proposed innate principles that draw on general structural knowledge are less plausible.

With regard to the second question above, of particular interest has been how general and specific knowledge 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 negative evidence (corrective feedback) plays little to no role in the process of recovery; only additional positive evidence of correct usages is necessary for “unlearning” of overgeneralized rules (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.

A family of usage-based theories that has grown in prominence in work on the acquisition of verb argument structure 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 (4) and (5) from Goldberg (1995).

(4)	Subj V Obj Obj <sub>2</sub> ⇔ X CAUSE Y RECEIVE Z Ex: <i>Pat faxed Bill the letter.</i>
(5)	Subj V Oblique ⇔ X MOVE Y Ex: <i>The fly buzzed into the room.</i>

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. In particular, explicit models must be explored, both of the underlying learning mechanisms and of the use of the acquired knowledge. 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 the use of negative evidence.

### 1.3 A Bayesian Model for Early Argument Structure Acquisition

We present an implemented computational model of the early stages of argument structure acquisition in children. We take a statistical usage-based approach, in which 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. 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 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; 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 U-shaped learning curve observed in children.

A preliminary version of this model was originally presented in Alishahi and Stevenson (2005a,b). In this paper, we describe an improved version of the model, present additional experiments, and discuss the results and their import for modeling of child language acquisition. We explain the model and its modules in more detail in the immediately following sections. Successive sections present and discuss computational experiments that demonstrate the match between the behaviour of the model and that of children in learning argument structure regularities and in using that knowledge in language comprehension and production.

## 2 The Representation of Probabilistic Constructions

The input to our model consists of natural language utterances, each paired with a logical representation of the meaning that the utterance conveys. In Section 2.1, we describe our assumptions regarding the input. The details of the input pairs are explained in Section 2.2. The model extracts an argument structure frame from each input pair, as described in Section 2.3. Whereas the input pairs are a simulation of the data that children use in language learning, the argument structure frames are the internal representation of the model for that data, which might differ for different tasks. Therefore, all through the experiments in this paper, we use the scene-utterance pairs as the original form of the input to our model. In Section 2.4, we describe the probabilistic representation of a construction in our model, which is a group of such argument structure frames which share similar features.

### 2.1 Basic Assumptions

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, they can establish a link between the linguistic description of the event and its conceptual representation. We use such pairings of utterance and meaning representation as the basic input to our model.

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, and that for each word, the child knows the corresponding meaning (i.e., 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). It has been argued that, for certain verbs or situations, using cross-situational learning is not enough.<sup>1</sup> However, for the purpose of learning the simple verbs discussed in this paper (see Section 5.2, we make the simplifying assumption that the child can learn the core (idiosyncratic) meaning of simple verbs before learning their full argument structure, although the detailed meaning (which we call the *semantic primitives* of the verb) are learned later.

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.) However, we recognize that resolving referential indeterminacy, learning word-to-meaning

---

<sup>1</sup>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”.

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

<sup>a</sup>Extracted from a representation of the child’s ontology.

Figure 1: An input pair and extracted frame for a verb argument structure.

mappings, and determining potential argument roles 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. In fact, we address the problem of learning semantic roles in Alishahi and Stevenson (2007), where we present a modified version of the model that acquires a general conception of the semantic roles of predicates based only on exposure to individual verb usage.

Another 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 certain features from each input pair in the form of an argument structure frame (see Section 2.3 for details), 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).

## 2.2 Scene-Utterance Input Pairs

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. 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.)

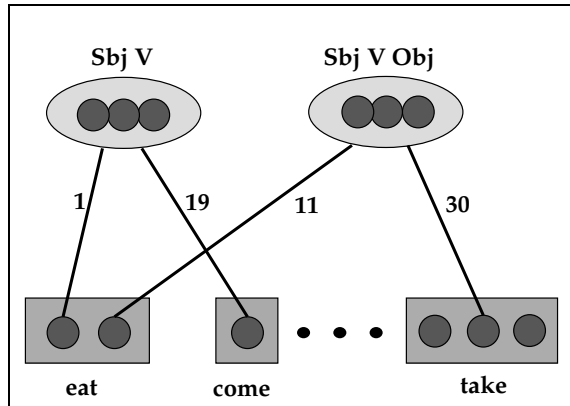


Figure 2: A portion of the lexicon showing an intransitive construction, with syntactic pattern Subj(ect) V(erb), and a transitive construction, with syntactic pattern Subj(ect) V(erb) Obj(ect). Numbers on links represent the observed frequency of the verb in a frame that participates in that construction. For example, *eat* has been seen once in an intransitive construction and 11 times in a transitive construction.

### 2.3 Argument Structure Frames

From each scene-utterance pair, our system extracts an argument structure frame. The bottom panel of Fig. 1 shows the frame corresponding to the main predicate, **Put**, of the input pair in the top panel of the figure. 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.

### 2.4 Constructions as Groups of Frames

A construction in our model is a group of argument structure frames that are “similar enough,” according to the probabilities over their features, to be grouped together. The notion of “similar enough” is described in detail in the next section of the paper; 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

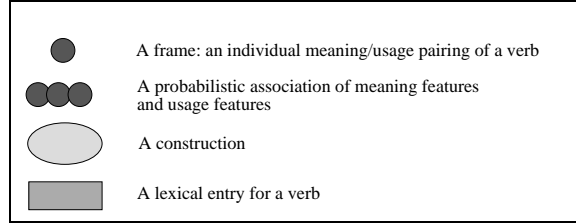


Figure 3: The legend for symbols used in all the diagrams representing verbs and constructions throughout the paper.

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 represent 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 construction 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 illustrates a portion of the acquired lexicon, showing the lexical entries of verbs containing frames, and their links to constructions. (See Fig. 3 for the legend of symbols used in all depictions of the lexicon throughout the paper.) In the next section, we describe how the link between a frame and a construction is established by our Bayesian learning algorithm.

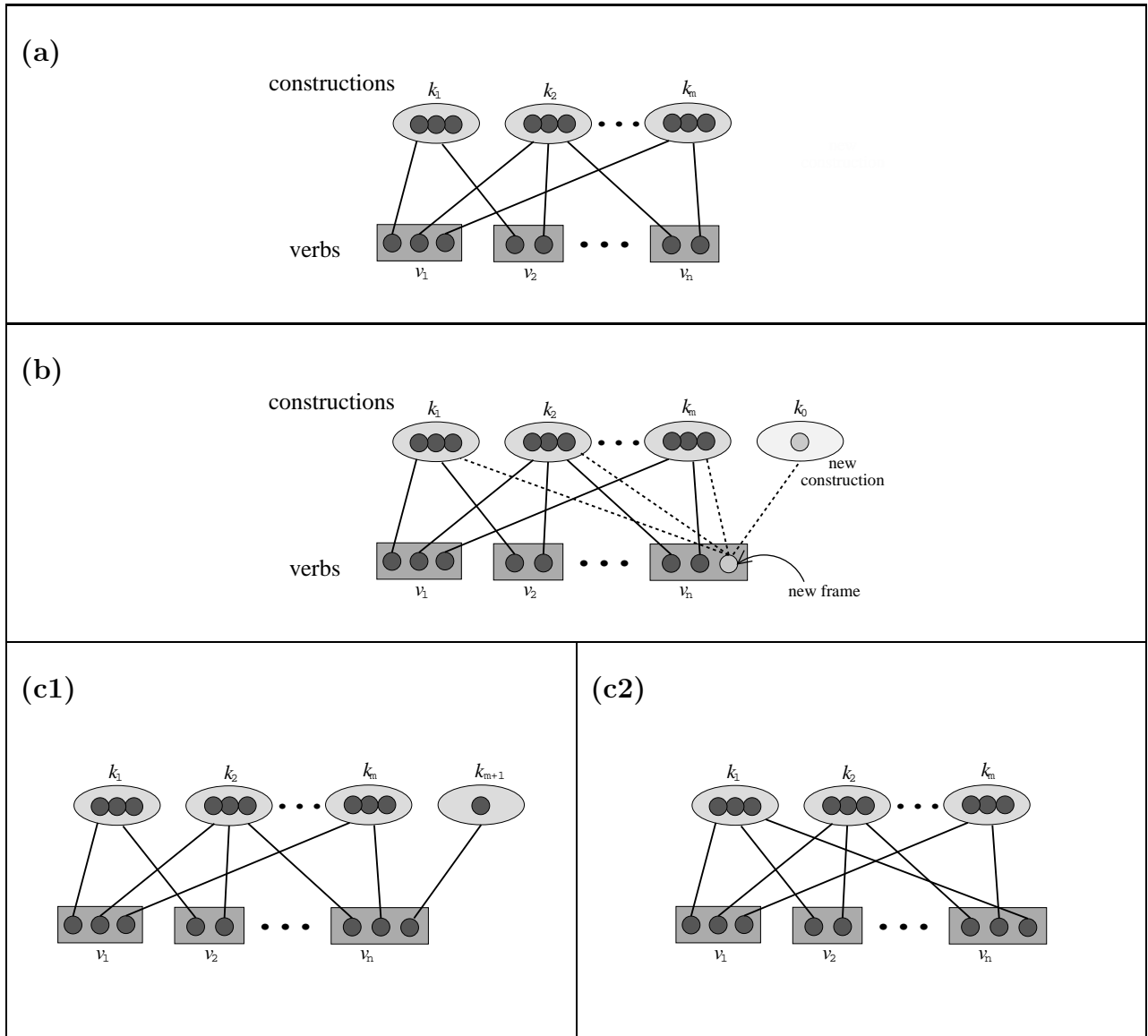


Figure 4: Processing a new frame: (a) A snapshot of the lexicon before the new frame is processed. (b) Addition of the new frame to the lexical entry for its main verb, in this case  $v_n$ . All existing constructions and a potential new construction are considered as a match for the new frame, as indicated by the dashed lines. Two alternatives are possible, as shown in panels c1 and c2. (c1) The new construction  $k_0$  is the best match, hence it is added to the lexicon, 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.

### 3 Acquisition of Constructions

#### 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).<sup>2</sup> 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. 4.

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

$$\mathbf{BestConstruction}(F) = \underset{k}{\operatorname{argmax}} P(k|F) \tag{1}$$

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  $k$ :

$$P(k|F) = \frac{P(k)P(F|k)}{P(F)} \propto P(k)P(F|k) \tag{2}$$

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).

#### 3.2 Details of the Probability Model for Learning

The following sections describe in detail the calculation in equation (2) of the prior and conditional probabilities, respectively, used in learning.

##### 3.2.1 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

---

<sup>2</sup>In Alishahi and Stevenson (2005a), we referred to these categories as ‘classes’; here we use the term ‘construction’ to refer to a group of similar frames.

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:

$$P(k) = \frac{n_k}{n + 1} \tag{3}$$

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:<sup>3</sup>

$$P(0) = \frac{1}{n + 1} \tag{4}$$

Thus, the prior probability for an existing construction is proportional to the frequency of its frames (equation (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 (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.

### 3.2.2 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:

$$P(F|k) = \prod_{i \in \text{FrameFeatures}} P_i(j|k) \tag{5}$$

where  $j$  is the value of the  $i^{\text{th}}$  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):<sup>4</sup>

$$P_i(j|k) = \frac{\text{count}_i^k(j) + \lambda}{n_k + \lambda\alpha_i} \tag{6}$$

where  $n_k$  is the number of frames participating in construction  $k$ , for  $k \geq 0$  (note that  $n_0 = 0$ ), and  $\text{count}_i^k(j)$  is the number of those with value  $j$  for feature  $i$ . The smoothing parameter  $\alpha_i$  is

---

<sup>3</sup>The formulas used in equation (3) and equation (4) are similar 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.

<sup>4</sup>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.

set to (an estimate of) the number of possible values that feature  $i$  can 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  $\lambda$  in equation (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.

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  $\lambda$  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 (3)), if the feature value of the 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.

## 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:

**Kick**<sub>[cause,contact]</sub>(**TIM**<sub>[agent]</sub>, **BALL**<sub>[theme]</sub>)

we ask the model to predict the “missing” syntactic pattern 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:

$$\begin{aligned} \mathbf{BestValue}_i(F) &= \operatorname{argmax}_j P_i(j|F) \\ &= \operatorname{argmax}_j \sum_k P_i(j|k)P(k|F) \end{aligned} \tag{7}$$

Here again,  $F$  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 (6). To calculate the factor  $P(k|F)$ , we use Bayes rule and drop the constant  $P(F)$  term (cf. equation (2) on page 14), to yield:

$$P(k|F) \propto P(k)P(F|k) \tag{8}$$

Here the prior probability,  $P(k)$ , and the conditional probability,  $P(F|k)$ , are simply determined as in our learning module (using equation (3) on page 15, and equation (5) on page 16, respectively).

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 other constructions. 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.

## 5 Experimental Set-Up

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 paper. Moreover, using automatically generated input allows us to test the model through a number of simulations, and evaluate it across minor variations in corpora.

We explain the basic properties of the input required by our model in Section 5.1. 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 5.2. In Section 5.3, we describe how the input-generation lexicon is used to create the input corpora for each simulation

U(tterance):	<i>Mary put ball on table.</i>
S(cene):	<b>Put</b> <sub>[cause,move]</sub> (MARY <sub>&lt;agent&gt;</sub> , BALL <sub>&lt;theme&gt;</sub> , <b>On</b> <sub>∅</sub> (TABLE <sub>&lt;destination&gt;</sub> ) <sub>&lt;destination&gt;</sub> )

Figure 5: A sample input pair

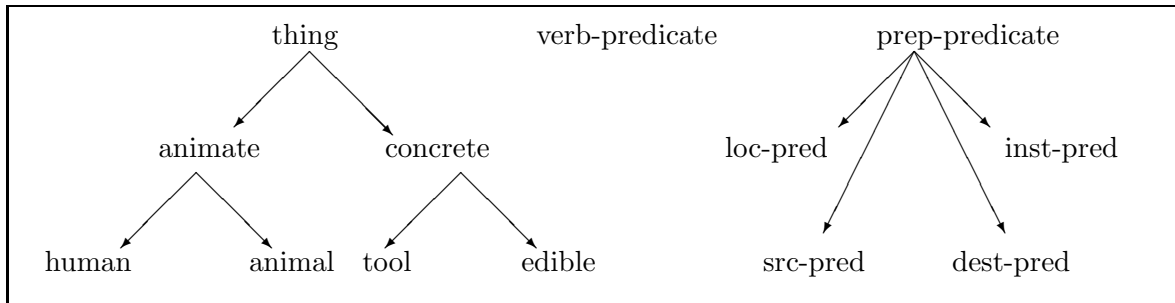


Figure 6: The semantic hierarchy

in our set of experiments. Finally, in Section 5.4, we explain how the parameters of the model are set.

### 5.1 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. 5, 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)). Each argument is assigned a participant role in the event (e.g., *agent*, *theme*, *destination*) and 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 are 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: *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;

---

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 7: A sample lexeme

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. 6. (As noted earlier in Section 2.3, these categories may be used in merging highly similar argument structure frames.) 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.

## 5.2 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 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

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

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

Table 2: Predicate terms in the lexicon.

of the same three children. Prepositions used in these conversations were also added to the lexicon. The selected words are shown in Table 2, and a sample lexical entry is shown in Fig. 7.

For each semantic category from Fig. 6 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  $\langle edible \rangle$  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 and a semantic symbol.

### 5.3 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 2.1).

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 (not so perfect) 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 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.<sup>5</sup>

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.

## 5.4 Setting the Parameters of the Model

For the purpose of implementing the model and running simulations, we need to assign values to parameter  $\alpha_i$  for each feature  $i$  in equation (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, and syntactic pattern.

The parameter  $\alpha_i$  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  $\alpha_i$  value for other features is determined by the number of semantic primitives, thematic roles, and semantic categories, respectively, as defined in Section 5.1:  $\alpha_{prim} = 12$ ,  $\alpha_{role} = 10$ , and  $\alpha_{arg\_cat} = 13$ .

We also need to set the parameter  $\lambda$  in equation (6). For all features  $i$  in frame  $F$ , we need  $\lambda$  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 (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

---

<sup>5</sup>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.

$k$  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 an upper bound of  $\lambda < \prod_l \frac{1}{\alpha_l}$ . This gives us an upper-bound of  $10^{-7}$  for the value of  $\lambda$ . We set  $\lambda$  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  $\alpha_i$  is determined by the number of the possible values that feature  $i$  can take, setting  $\lambda$  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  $\lambda$  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.

## 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.

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. However, 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.

### 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

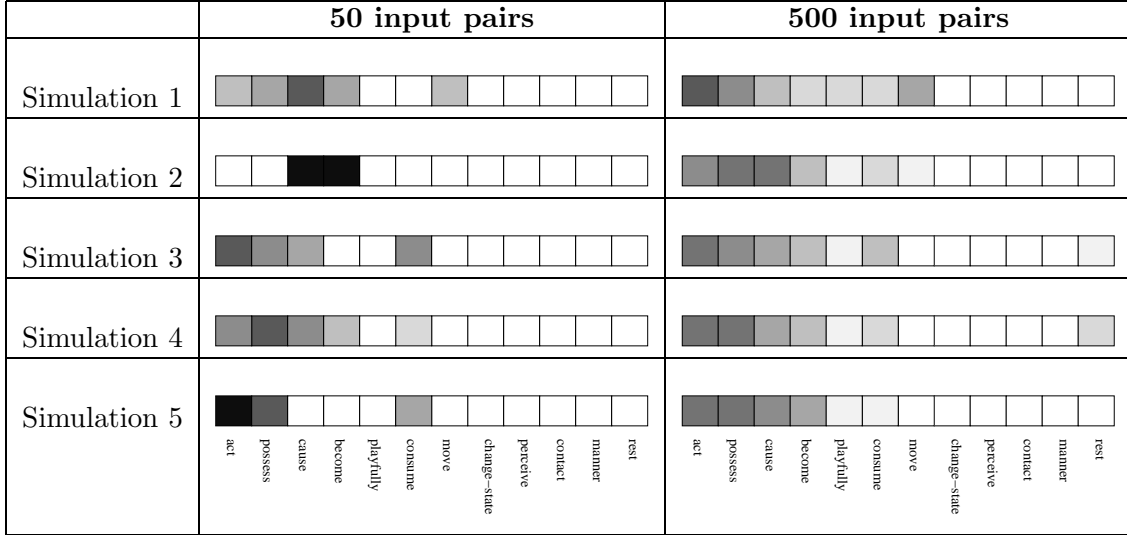


Figure 8: The probability distribution of semantic primitives in the transitive construction at two different points in learning.

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).

### 6.1.1 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. 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. 8, Fig. 9 and Fig. 10 show the probabilities of primitives associated with the transitive, intransitive, and ditransitive constructions, respectively, across the five simulations. 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. 8 and Fig. 9. 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. 9 shows that, after processing 50 input pairs, the probabilistic distribution of semantic primitives differs across the five simulations.

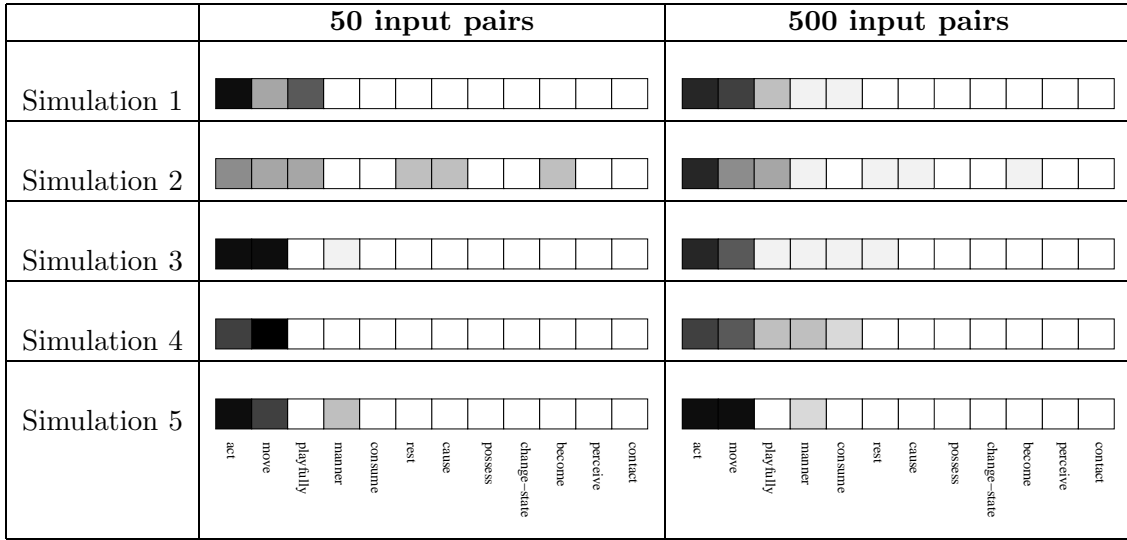


Figure 9: The probability distribution of semantic primitives in the intransitive construction at two different points in learning.

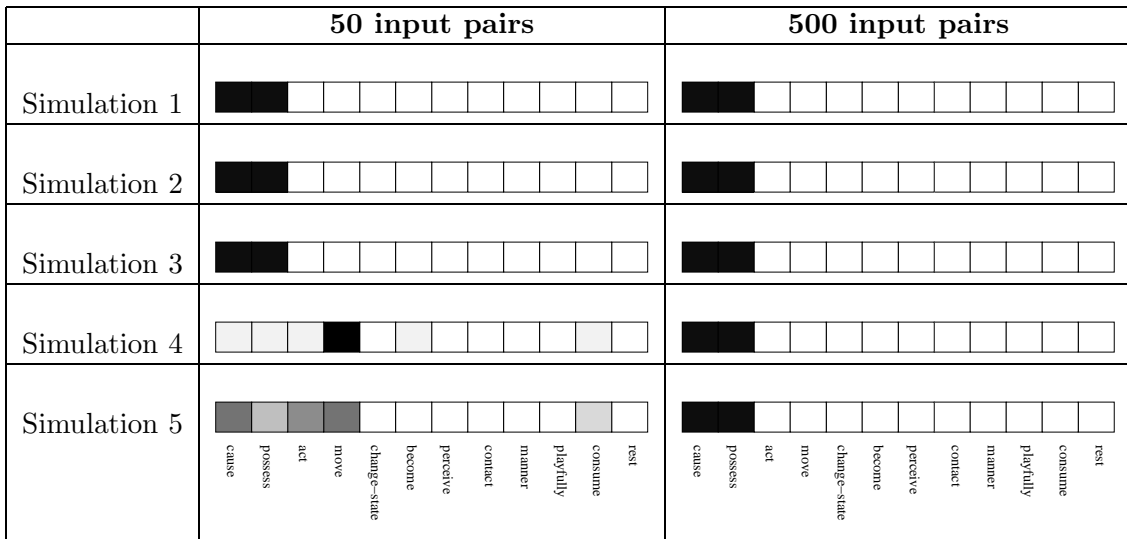


Figure 10: The probability distribution of semantic primitives in the ditransitive construction at two different points in learning.

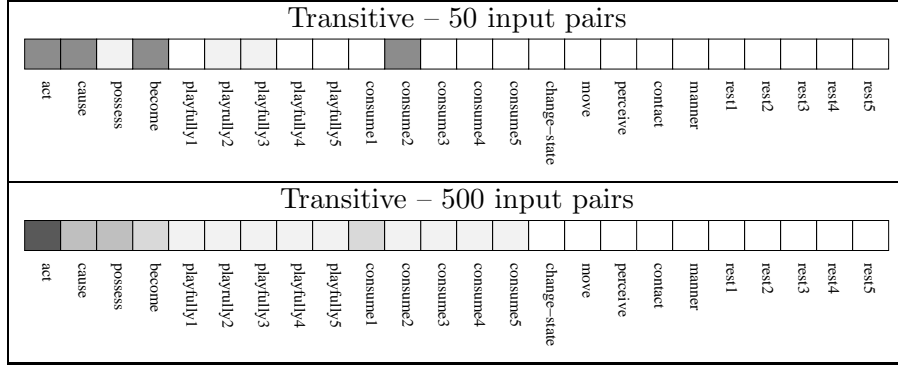


Figure 11: Transitive construction for the extended lexicon

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. 10 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.

### 6.1.2 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

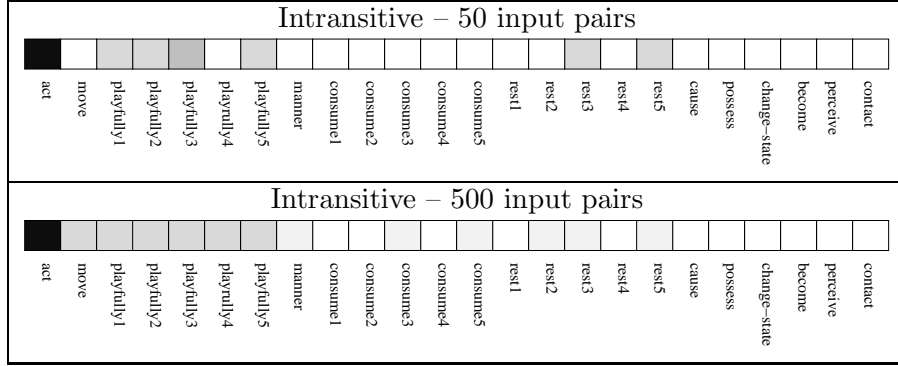


Figure 12: Intransitive construction for the extended lexicon

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. For example, we added to the lexicon five verbs,  $sit_1$  to  $sit_5$ , each of which had the semantic primitive *act* associated with it, but in place of the fine-grained primitive *rest* that occurs with *sit*, each verb  $sit_i$  is associated with the primitive  $rest_i$ . 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. 11 and Fig. 12 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 primitives and the form. In the long run, this is determined by the frequencies of all the verbs that appear in that construction.

## 6.2 U-Shaped 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. As a usage-based model that incorporates a generalization mechanism, we expect an overall U-shaped learning curve from our system, but also expect variation among individual verbs. 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.

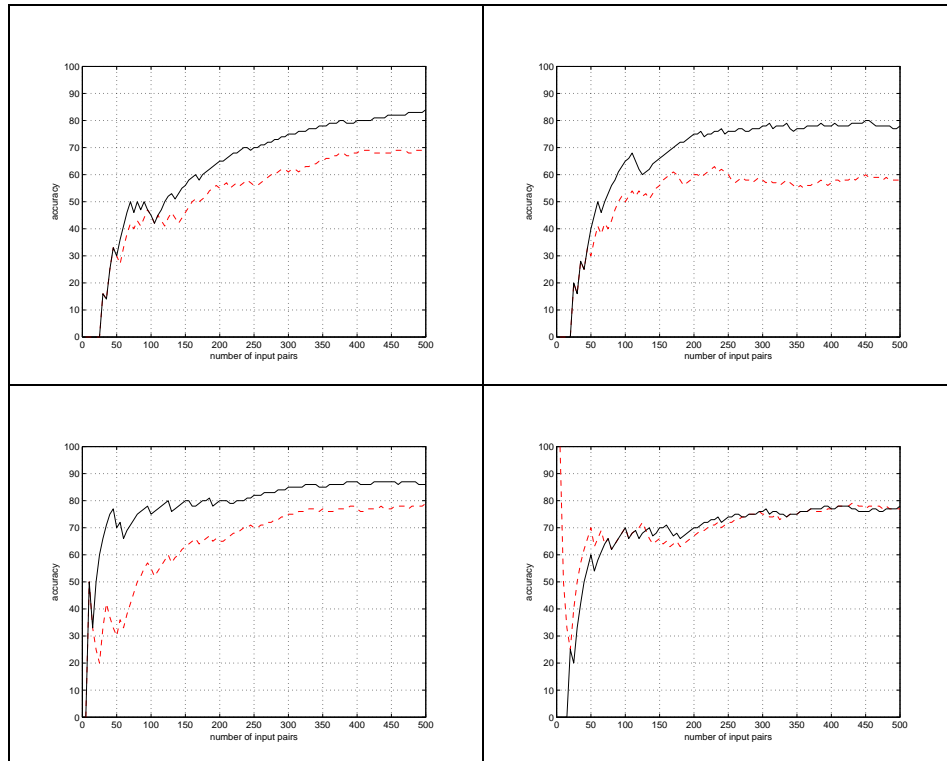


Figure 13: Cumulative accuracy of the model for predicting syntactic pattern. Solid lines show performance with noiseless input data, and dashed lines show performance with noisy input data.

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<sup>th</sup> test input in which the utterance was removed. Our prediction model was used to find the syntactic pattern with the highest probability for the 6<sup>th</sup> (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. 13 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 5.3); 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 more accurate performance followed by less accurate performance, before more consistent performance is eventually achieved, reflecting the general aspects of a U-shaped learning curve. 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

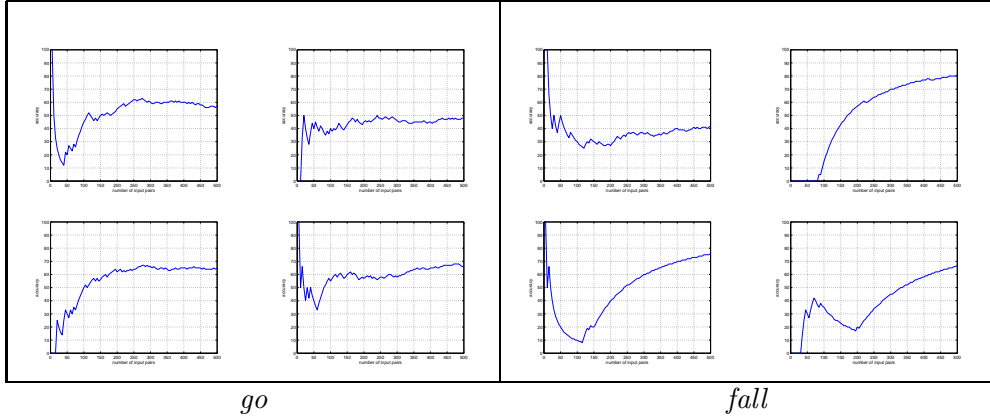


Figure 14: Sample learning curves for different verbs, showing cumulative accuracy over time.

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. 13). In an actual child language situation with much more exposure to language, we expect that such scenarios would be avoided.

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. 14 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 even 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.

Fig. 13 and Fig. 14 show that 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.

### 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 7.2.

### 6.3.1 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 blinking*) 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 blinking 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 4, language use is modeled as choosing the best value  $j$  for a missing feature  $i$ , based on other observed features, using equation (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:

$$\begin{aligned} U_T &: \textit{Bunny blink duck} \\ U_I &: \textit{Bunny and duck blink} \end{aligned}$$

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,  $S_A$ :

$$\begin{aligned} S_{CA} &: \mathbf{Blick}_{[cause,act]}(\mathbf{BUNNY}_{(agent)}, \mathbf{DUCK}_{(theme)}) \\ S_A &: \mathbf{Blick}_{[act]}(\mathbf{And}_{\square}(\mathbf{BUNNY}, \mathbf{DUCK})_{(agent)}) \end{aligned}$$

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

$U_T-S_{CA}$	<i>Bunny blick duck.</i> <b>Blick</b> <sub>[cause,act]</sub> ( <b>BUNNY</b> <sub>(agent)</sub> , <b>DUCK</b> <sub>(theme)</sub> )	Matched
$U_T-S_A$	<i>Bunny blick duck.</i> <b>Blick</b> <sub>[act]</sub> ( <b>And</b> <sub>∅</sub> ( <b>BUNNY</b> , <b>DUCK</b> ) <sub>(agent)</sub> )	Unmatched
$U_I-S_{CA}$	<i>Bunny and duck blick.</i> <b>Blick</b> <sub>[cause,act]</sub> ( <b>BUNNY</b> <sub>(agent)</sub> , <b>DUCK</b> <sub>(theme)</sub> )	Unmatched
$U_I-S_A$	<i>Bunny and duck blick.</i> <b>Blick</b> <sub>[act]</sub> ( <b>And</b> <sub>∅</sub> ( <b>BUNNY</b> , <b>DUCK</b> ) <sub>(agent)</sub> )	Matched

Table 3: The four test input pairs used in the experiment of Section 6.3.1.

Test Input Pair	# Training Input Pairs				
	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 4:  $\log P(k|F)$  values for matched and unmatched scene-utterance pairs.

the intransitive form,  $U_I$ . The four resulting input pairs are shown in Table 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 (1) on page 14), 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 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). (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 ( $7.5 \times 10^{-7}$ ) is almost four times higher than the one for the unmatched pair ( $1.9 \times 10^{-7}$ ). 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 6.5).

		# 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 5: Average probability of the predicted patterns.

### 6.3.2 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! Gorp-ing!*, 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:

$S_{G1}$ : **Gorp**<sub>[cause,act]</sub>(KITTY<sub>(agent)</sub>, DOGGY<sub>(theme)</sub>)  
 $U_{G1}$ : ?

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 gorp-ing”, 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 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

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

Figure 15: Comparing the first eight uses of *fall* for Adam with the first eight sentences generated by our model.

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:

$S_{G2}$ : **Gorp**<sub>[cause,act]</sub>(KITTY<sub><agent></sub>, DOGGY<sub><theme></sub>)  
 $U_{G2}$ : *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 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.

## 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 the area of argument structure is when a non-causative (intransitive-only) verb is used causatively (in a transitive structure).

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

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. 15 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.

## 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

---

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

convey particular semantic properties (Goldberg, 1995). For instance, in *The fly buzzed into the room*, the sound emission verb *buzz* 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. 16. 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 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:

$S_{P1}$ : **Dance**<sub>[??]</sub>(KITTY<sub>[?]</sub>, DOGGY<sub>[?]</sub>, Under<sub>[]</sub>(TABLE<sub>[?]</sub>)<sub>[?]</sub>)  
 $U_{P1}$ : *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.

Verb Sem. Primitives	# Input Pairs		Verb Sem. Primitives	# Input Pairs	
	50	500		50	500
<i>act</i>	.44	.41	<i>cause</i>	.41	.46
<i>move</i>	.34	.31	<i>move</i>	.39	.33
<i>playfully</i>	.18	.14	<i>possess</i>	.04	.19
<i>manner</i>	.04	.09	<i>act</i>	.02	10 <sup>-2</sup>
<i>consume</i>	10 <sup>-4</sup>	.02	<i>become</i>	10 <sup>-3</sup>	10 <sup>-3</sup>
<i>rest</i>	10 <sup>-4</sup>	.01	<i>change-state</i>	.09	10 <sup>-4</sup>
<i>cause</i>	10 <sup>-4</sup>	10 <sup>-3</sup>	<i>playfully</i>	10 <sup>-3</sup>	10 <sup>-5</sup>
<i>change-state</i>	10 <sup>-4</sup>	10 <sup>-5</sup>	<i>consume</i>	10 <sup>-3</sup>	10 <sup>-5</sup>
<i>possess</i>	10 <sup>-4</sup>	10 <sup>-5</sup>	<i>rest</i>	10 <sup>-3</sup>	10 <sup>-6</sup>
<i>become</i>	10 <sup>-4</sup>	10 <sup>-5</sup>	<i>perceive</i>	10 <sup>-3</sup>	10 <sup>-6</sup>
<i>perceive</i>	10 <sup>-4</sup>	10 <sup>-5</sup>	<i>manner</i>	10 <sup>-3</sup>	10 <sup>-6</sup>
<i>contact</i>	10 <sup>-4</sup>	10 <sup>-5</sup>	<i>contact</i>	10 <sup>-3</sup>	10 <sup>-6</sup>

Typical usage

“Unusual” usage

Table 6: Probabilities of semantic primitives associated with a typical and “unusual” usage of the verb *dance*.

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. 17.

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:

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

$$U_{P2}: ?$$

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

## 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.

## 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, 2000). 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<sub>X</sub> Verb Oblique<sub>Y</sub>” is paired with the meaning “X moves Y”, corresponds to a core usage of the light verb *go*.<sup>6</sup> 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 comprised 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 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*).<sup>7</sup>

## 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, 2000). 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

---

<sup>6</sup>Here, Oblique refers to a prepositional or adverbial argument indicating a direction or goal, as in *Hans goes to school* or *Ingrid went home*.

<sup>7</sup>Examples adapted from Google searches on “his way to a promotion”, 19 June 2006.

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. Less-entrenched constructions emerge very early on (after 10 training input pairs, for a common construction), 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). 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.

### 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 (Bowerman, 1988). 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 frequency 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 (Bowerman, 1988). 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.

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.

#### 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, 2005). 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 (2005).

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.

## 8 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 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 (e.g., as in Siskind, 1996; Yu et al., 2005). However, 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.

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.” Li et al. (2004) use a self-organizing neural network to model early lexical development. Their model shows the emergence of linguistic categories, and lexical confusion and age-of-acquisition effects, but does not deal with any kind of ambiguity regarding the correct mapping between a form and a meaning representation. In our work, we simplify the assumptions along all of these lines, assuming a simple logical form representation for an action, and presupposing that the learner has determined the correct level of interpretation of the event. Future work in our framework will need to incorporate more extensive semantic representations, as well as the means for disambiguating the utterance with respect to the intended semantics.

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 to represent words by categorizing them as nouns or verbs, with further subcategorization of nouns as animate/inanimate, human/non-human, etc. 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, learning of general syntactic constructions is implicit, and the acquired knowledge cannot be used in any language task other than limited comprehension.

Other computational models are more similar to ours in honing in 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 of general grammatical knowledge (e.g., Jones et al., 2000; Clark, 2001a,b; Gobet et al., 2004; Solan et al., 2004). 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.

## 9 Conclusions

We have described a Bayesian model for the representation, acquisition and use of argument structure constructions in a usage-based framework. The proposed model suggests a novel description for the nature of constructions. Instead of formulating constructions as pairs of form and meaning, we view each construction as a probabilistic association between syntactic and semantic properties of verbs and their arguments. This probabilistic association emerges over time through a Bayesian acquisition process, in which similar verb usages are detected and grouped together to form general constructions, based on their syntactic and semantic properties. The model demonstrates the feasibility of learning a range of abstract constructions from examples of verb usage even in the presence of noisy or incomplete input data.

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 timecourse 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 an extended version of this work (Alishahi and Stevenson, 2007), we directly address the learning and use of semantic roles. The modified model associates each argument position of a predicate with a probability distribution over a set of semantic properties of the arguments that appear in that position. We call this distribution over semantic properties a *semantic profile*. The preliminary results 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, as has been suggested in the literature (Tomasello, 2000, e.g.).

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*<sub>1</sub> let alone *verb*<sub>2</sub>”, 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). We are currently incorporating into our model of early verb learning a mechanism for word-specific constructions in order to provide a fuller account of construction representation, acquisition and use.

## References

- Akhtar, N. (1999). Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26:339–356.
- Alishahi, A. and Stevenson, S. (2005a). The acquisition and use of argument structure constructions: a Bayesian model. In *Proceedings of the Second Workshop on Psycho-computational Models of Human Language Acquisition, ACL-2005, Ann Arbor*, pages 82–90.

- Alishahi, A. and Stevenson, S. (2005b). A probabilistic model of early argument structure acquisition. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Stresa, Italy.
- Alishahi, A. and Stevenson, S. (2007). A computational usage-based model for learning general properties of semantic roles. In *Proceedings of the 2nd European Cognitive Science Conference*.
- Allen, J. (1997). Probabilistic constraints in acquisition. In Sorace, A., Heycock, C., and Shillcock, R., editors, *Proceedings of the GALA97 Conference on Language Acquisition, Edinburgh*, pages 300–305.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429.
- 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*. London: Blackwell.
- 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.
- 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.
- Clark, A. (2001a). Unsupervised induction of stochastic context free grammars with distributional clustering. In *Proceedings of CoNLL, Toulouse, France*, pages 105–112.
- Clark, A. S. (2001b). *Unsupervised language acquisition: theory and practice*. PhD thesis, University of Sussex.
- 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 25th 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.
- 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.
- 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., 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. (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*.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1:135–176.
- 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. (2005). *Constructions in Context*. Oxford University Press. to appear.
- 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.
- Grimshaw, J. (1990). *Argument Structure*. 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*.
- Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.
- Langacker, R. (1999). *Grammar and conceptualization*. Berlin: Mouton de Gruyter.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. The University of Chicago Press.

- Li, P., Farkas, I., and MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17:1345–1362.
- Lieven, E., Pine, J., and Rowland, G. (1998). Comparing different models of the development of the english verb category. *Linguistics*, 36(4):807–830.
- 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: Erlbaum.
- MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates, 2 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).
- 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.
- Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17:357–374.
- 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.
- Pereira, F. (2000). *Formal grammar and information theory: Together again?*, 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., Lebeaux, D. S., and Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition*, 26:195–267.
- 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.
- 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.
- Theakston, A. L. (2004). The role of entrenchment in children’s and adults’ performance on grammaticality judgment tasks. *Cognitive Development*, 19:15–34.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74:209–253.
- 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.
- Yu, C., Ballard, D. H., and Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, 29:961–1005.