

# The Acquisition and Use of Argument Structure Constructions: A Bayesian Model

**Afra Alishahi**

Department of Computer Science  
University of Toronto  
afra@cs.toronto.edu

**Suzanne Stevenson**

Department of Computer Science  
University of Toronto  
suzanne@cs.toronto.edu

## Abstract

We present a Bayesian model for the representation, acquisition and use of argument structure constructions, which is founded on a novel view of constructions as a mapping of a syntactic form to a probability distribution over semantic features. Our computational experiments demonstrate the feasibility of learning general constructions from individual examples of verb usage, and show that the acquired knowledge generalizes to novel or low-frequency situations in language use.

## 1 Argument Structure Constructions

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., Fillmore et al., 1988; Lakoff, 1987). In particular, an argument structure construction is a mapping between fundamental verb-argument relations and the syntax used to express them, as in (1) and (2) from Goldberg (1995).<sup>1</sup>

<p>(1) Subj V Obj Obj<sub>2</sub> ⇔ X CAUSE Y RECEIVE Z Ex: <i>Pat faxed Bill the letter.</i></p>
<p>(2) Subj V Oblique ⇔ X MOVE Y Ex: <i>The fly buzzed into the room.</i></p>

Associations between argument configurations (such as the transitive) and semantic features

<sup>1</sup>Such constructions serve a similar purpose to linking rules (e.g., Pinker, 1989) or the event structure templates of Rappaport Hovav and Levin (1998).

(such as causation) capture important linguistic regularities, and appear to play a role in both child language acquisition (Gleitman, 1990; Naigles, 1990; Fisher, 2002) and adult sentence interpretation (Bencini and Goldberg, 2000). An established form-meaning mapping may even impose an “unusual” meaning when a verb is used in a manner that is not typical for it. For example, in (2) above, using the verb *buzzed* in a construction with a path argument induces a semantics of movement as well as the standard sound emission sense for the verb (Goldberg, 1995).

A theory of grammar that includes argument structure constructions (henceforth, simply “constructions”) as an organizing component of predicate knowledge must address a number of questions concerning the nature, acquisition and use of such constructions, namely:

- Precisely what constitutes the form-meaning mapping of a construction?
- How are general constructions learned from specific usages of verbs?
- What role do constructions play in language interpretation and production?

We have developed a Bayesian model of language acquisition and processing that has been shown to mimic children’s behaviour in forming word order generalizations, and in recovering from overgeneralizations without negative evidence (Alishahi and Stevenson, 2005). In this paper, we describe how the model enables a new view on the nature of constructions, thus providing an answer to the first question above which leads to interesting consequences for the others. Specifically, each construction is not simply a form-meaning pair (as in (1) and (2) above),

but rather maps a form to a probability distribution over the associated elements of meaning. The probabilistic nature of constructions in the model enables it to capture both statistical effects in language learning, and adaptability in language use.

Statistical patterns are widely accepted to play a role in language acquisition, and work on construction learning has taken a usage-based approach (e.g., Goldberg, 1995; Langacker, 1999; Tomasello, 2000). Our model elaborates on this view in the context of our assumption of a probabilistic form-meaning mapping. Constructions arise through an unsupervised Bayesian categorization process that groups verb usages according to probabilities over the properties of the verb and its arguments. Each group forms a construction in which the semantic primitives occurring most frequently across the group have the highest probabilistic association with the syntactic form. We demonstrate in computational experiments that such primitives are typically the more general semantic properties, modeling the ability of a child to capture argument structure regularities by inducing general constructions from individual usages.

The probabilistic nature of our view of constructions also influences the properties of language use, which we formulate 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 an observed utterance. Constructions enable the model to generalize observed patterns of association to new or low-frequency situations. This property underlies our further experimental results that mimic children’s ability to infer the basic semantics of a novel verb from its usage. This property extends to the use of a known verb in an unusual pattern (cf. (2) above): we additionally show that the model can associate a novel construction with a verb while avoiding inappropriate overgeneralizations.

Scene-Utterance Input Pair		
PUT <sub>[cause, move]</sub> (MOM <sub>(agent)</sub> , TOYS <sub>(theme)</sub> , IN <sub>[]</sub> (BOXES <sub>(goal)</sub> )) <i>Mom put toys in boxes.</i>		
Extracted Argument Structure Frame		
head verb	<i>put</i>	
verb sem. primitives	<i>&lt;cause, move&gt;</i>	
args	roles	<i>&lt;agent, theme, goal&gt;</i>
	categories <sup>a</sup>	<i>&lt;human, concrete, dest-pred&gt;</i>
syntactic pattern		arg1 verb arg2 arg3

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

Figure 1: An input pair and extracted frame.

## 2 Probabilistic Constructions

### 2.1 Argument Structure Frames

In our view, a construction is a group of individual verb usages, the latter of which we represent as argument structure frames. Each frame records the syntactic pattern of a verb usage, along with the semantic properties of the verb and its arguments in that pattern. Our model extracts the features of a frame from an input observation in the form of a scene-utterance pair: the perceived utterance (what the child hears), and a logical form representation of the relevant aspect of the observed scene (the semantics described by the utterance).<sup>2</sup> Figure 1 shows a sample input pair and the extracted frame.

### 2.2 Constructions as Groups of Frames

Each construction is a group of extracted frames which share a common syntax—i.e., a particular syntactic configuration of arguments.<sup>3</sup> Because the semantic properties of such usages may vary in their particulars, elements of meaning in a construction are probabilistically associated with the syntactic form. 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

<sup>2</sup>We assume that the (non-trivial) task of picking out the utterance semantics from the full scene representation has been performed (as in Siskind, 1996, for example).

<sup>3</sup>Though note that not all frames with the same syntax necessarily form a single construction.

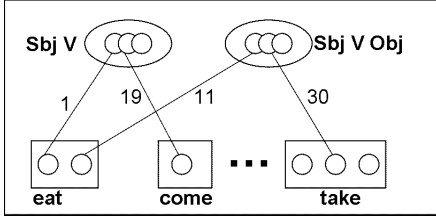


Figure 2: A portion of the lexicon showing 2 constructions. Circles represent frames.

of usages that exhibit this form, then we would find a higher probability for the primitive *act* given this construction than for the others. In this way, constructions probabilistically generalize the semantics of a set of frames.

Each verb with an observed usage 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 a compatible frame, capturing the statistical usage pattern of the verb. These frequencies are also used in the calculation of the probabilities of association between a construction and the features occurring in its observed frames. Moreover, the sum of its incoming link frequencies contributes to the overall probability of a construction. Figure 2 illustrates a portion of the acquired lexicon; in the next section, we describe how the link between a frame and a construction is established.

### 3 Acquisition of Constructions

#### 3.1 Overview

Every new frame is input to an incremental Bayesian clustering process that groups the new frame together with an existing group of frames—a construction—that probabilistically has the most similar properties to it. If none of the existing constructions has sufficiently high probability, then an entry for a new construction is created, containing only the new frame.

The probability of each construction for the frame is determined by both syntactic and semantic features. Currently, a construction with a different syntactic pattern from that of the frame, or a different set of argument roles, would have a very low probability. The probability of

the semantic primitives of the verb, as well as the semantic categories of its arguments, is determined by how frequently those of the frame occur across the frames of the candidate construction. Because these probabilities may be calculated over partial information, we can simulate learning in the face of an incomplete frame.

#### 3.2 Details of the Learner

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>4</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.

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

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

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

$$P(k|F) = \frac{P(k)P(F|k)}{P(F)} \approx P(k)P(F|k) \quad (2)$$

The prior probability of  $k$  is given by:<sup>5</sup>

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

where  $n$  is the total number of observed frames;  $n_k$  is the number of frames participating in construction  $k$ , for  $k > 0$ ; and  $n_0 = 1$ . Thus, the prior for an existing construction is proportional to the frequency of its frames, and that of a new construction is inversely proportional to the number of observed frames overall. The prior probability estimation for each construction follows the intuition that it is more probable for

<sup>4</sup>In Alishahi and Stevenson (2005), we referred to these categories as ‘classes’; we use the terms ‘constructions’ and ‘classes’ interchangeably to refer to a group of similar frames.

<sup>5</sup>This is the formula used by Anderson (1991) with his “coupling probability” set to the mid value of 0.5.

a frame to come from a more entrenched construction (i.e., one with more frames), and that as the number of the observed frames increases, the probability that a frame comes from a new construction decreases.

The probability of a frame  $F$  is expressed in terms of the individual probabilities of its features (shown above in Figure 1). 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 features:

$$P(F|k) = \prod_{i \in \text{FrameFeatures}} P_i(j|k) \quad (4)$$

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.

## 4 Language Use as Prediction

### 4.1 Overview

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. If a usage of a verb is sufficiently complete and frequent, then it will have a strong influence on the determination of unknown features. On the other hand, constructions play an important role in the face of missing or low-confidence verb-based information, because a prediction based on a construction generalizes over all its frames. This enables the model to produce or understand a verb in a novel (for that verb) syntactic pattern, as long as semantically similar verb usages have been observed.

### 4.2 Details of the Prediction Process

To integrate verb-based and construction-based knowledge, we must extend the prediction aspect of the model of Anderson (1991). We begin with his prediction formula:

$$\begin{aligned} \text{BestValue}_i(F) &= \underset{j}{\operatorname{argmax}} P_i(j|F) \\ &= \underset{j}{\operatorname{argmax}} \sum_k P_i(j|k)P(k|F) \end{aligned} \quad (5)$$

where  $F$  is a partial frame,  $i$  is a missing feature,  $j$  ranges over possible values of  $i$ , and  $k$  ranges over all categories. Intuitively, the model generalizes over the items within a category 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 category given the partial frame ( $P(k|F)$ ).

However, the structure of our acquired knowledge is more complex than that of Anderson’s model. Specifically, we have two groups of items over which we might generalize: in addition to the groupings of frames into constructions (which would be indicated by the formula above), we also have the groups of frames associated with each verb.

Given the importance of verb-based knowledge in language acquisition (Tomasello, 2000), we begin by focusing only on the frames associated with the verb  $v$  in frame  $F$ :

$$P_i(j|F) = \sum_{k_v \in C(v)} P_i(j|k_v)P(k_v|F) \quad (6)$$

where  $C(v)$  is the set of constructions linked to by the frames of verb  $v$ . This formulation generalizes over the constructions associated with a verb, but ignores all other constructions. This unnecessarily restricts the model when a partial frame for a verb does not match well with any frame previously seen *for that verb*, but may be compatible with another learned construction.

To allow more general knowledge of constructions to influence the prediction process, we find the best construction for a partial frame  $F$  during prediction as if it were a newly observed frame to be learned. We apply our Bayesian learner to determine the most compatible construction  $k_F$  for the partial frame  $F$  using eqn. (1), and temporarily insert the frame into the lexical entry for  $v$  with a frequency of 1

on the link to  $k_F$ .<sup>6</sup> This ensures that the overall best construction is taken into consideration, along with the constructions associated with the verb, in predicting values for a partial frame.

$P(k_v|F)$  in eqn. (6) is rewritten using Bayes rule and dropping the  $P(F)$  term (cf. eqn. (2)):

$$P(k_v|F) \approx P(k_v)P(F|k_v) \quad (7)$$

$P(F|k_v)$  is determined as in eqn. (4), using a uniform probability distribution over the possible values of the missing feature. In calculating  $P(k_v)$ , the frequency of each construction (its number of frames) is weighted by the frequency of  $v$ 's frame which links to it, balancing the overall likelihood of the construction with the likelihood that it is a construction for  $v$ :

$$w_{k_v} = \frac{n_{k_v}}{\sum_{k_{v'} \in C(v)} n_{k_{v'}}} \times \text{freq}(v, k_v) \quad (8)$$

where  $n_{k_v}$  is the frequency of the construction  $k_v$ ,  $C(v)$  is the set of constructions linked to by the frames of verb  $v$ ,  $\text{freq}(v, k_v)$  is the frequency of  $v$ 's frame which links to construction  $k_v$ , and  $\text{freq}(v, k_F) = 1$ . The prior probability  $P(k_v)$  is then calculated by normalizing the weight  $w_{k_v}$  from eqn. (8):

$$P(k_v) = \frac{w_{k_v}}{\sum_{k_{v'} \in C(v)} w_{k_{v'}}} \quad (9)$$

A particularly interesting situation arises when the best overall construction,  $k_F$ , is previously unseen for  $v$ . The calculation for  $P(k_v)$  entails that  $k_F$  generally has less influence the more often the verb has been seen overall; that is, greater weight from  $v$ 's observed frames to their constructions will outweigh the influence of the single partial frame to its “new” construction. This factor is responsible for recovery from overgeneralization errors (Alishahi and Stevenson, 2005). However, this effect is modulated by the other factor,  $P(F|k_v)$ . If the partial frame  $F$  is more compatible with the “new” construction than with an existing construction for  $v$ , the use of  $k_F$  may increase in probability even for a frequent verb. This can be the situation in productive generalizations (the use of “unusual”

<sup>6</sup>Note that  $k_F$  may or may not be a construction already linked to by a frame of  $v$ .

constructions), as we demonstrate in our experimental results below.

## 5 Experimental Materials

For our computational experiments, we automatically create input corpora—sequences of scene-utterance pairs—using early-acquired verbs in distributional patterns of child-directed speech. The input-generation lexicon includes the 13 most frequent verbs in mother’s speech across three children in CHILDES (age 1;6 to 5;1), along with their argument structure frames and associated frequencies, determined through manual analysis of the children’s conversations. Additional words in the lexicon include prepositions used in the same conversations, as well as a small number of nouns.

We enumerated a set of 9 primitives for describing the coarse-level semantics of a verb (*act, cause, move, become, etc.*), along with features as needed to capture finer-grained meaning distinctions (such as *consume* and *rest*). The semantic categories of the nouns were selected to reflect a simplified early ontology of a child. We assume that at the stage of acquisition being modeled, these features, along with the semantic roles of the arguments, can be largely determined by the child from the observed scene. The corresponding syntactic forms are simplified to indicate the order of arguments, with words used only in their root form.

For each simulation, a random sequence of input pairs is produced from the input-generation lexicon, using the frequencies to determine the probabilities of selecting a particular verb and argument structure for each input. Arguments which are predicates (such as prepositional phrases) are constructed recursively. To simulate noise, every third input pair in every generated corpus has one of its features randomly removed. During a simulation, each missing feature is replaced with the most probable value predicted at that point in learning, corresponding to a child learning from her own inferred knowledge. The resulting input data is noisy, especially in the initial stages of learning.

## 6 Experimental Results

We report computational experiments that demonstrate the ability of our model to learn constructions representing the knowledge of argument structure regularities, and to generalize this knowledge to novel situations in language use. All reported results are averaged over 10 simulations using different randomly generated input corpora; all simulations use the same system parameter settings.

We first show how general constructions emerge as the model is exposed to verb usages over time. We then turn to use of the constructions in novel situations. We demonstrate that the model induces the coarse meaning of an unknown verb based on its syntactic usage, and deals appropriately with cases where a verb appears in a usage that is unusual for it.

### 6.1 The Emergence of Constructions

Goldberg (1995) suggests that an argument structure construction is a grouping of similar verbs around a *light verb*, a semantically simple and highly frequent verb such as *go*, *make*, or *give*. The syntactic form and semantic properties of the core light verb determine the form-meaning mapping of the construction.

We propose an alternative view, in which constructions arise regardless of the generality of any particular verb’s semantic properties. (Though constructions may be more likely to form around light verbs due to their frequency.) Even if none of its individual verb usages is semantically simple, the more-basic verb semantic primitives associated with a construction will, over time, increase in probability. While an increasing number of semantic primitives may be associated with a construction given exposure to a greater variety of verbs, the more-basic semantic primitives will become entrenched because they will be observed across many more frames.

Simple intransitive and transitive constructions emerge consistently in all our simulations. Table 1 shows the probabilistic association of verb semantic primitives with the intransitive usage (“Subj V”), after 100 and after 1000 input pairs (averaged over 10 simulations). The

Verb Sem. Primitives	# Input Pairs	
	100	1000
<i>act</i>	.47	.50
<i>move</i>	.37	.34
<i>manner</i>	.09	.11
<i>consume</i>	.01	.03
<i>rest</i>	.01	.01
<i>possess</i>	.01	10 <sup>-4</sup>
<i>cause</i>	.01	10 <sup>-4</sup>
<i>become</i>	.01	10 <sup>-4</sup>
<i>change-state</i>	.01	10 <sup>-4</sup>
<i>perceive</i>	.01	10 <sup>-4</sup>
<i>contact</i>	.01	10 <sup>-4</sup>

Table 1: Probabilities of semantic primitives associated with an intransitive construction.

probabilities indicate that the construction has a very strong bias toward the general primitives *act* and *move*. Although a few primitives that are variously associated with some intransitive verb usages increase slightly in probability (*manner*, *consume*, and *rest*), the remaining primitives drop off to negligible values. An examination of the transitive construction shows a similar pattern, with primitives *act*, *possess*, and *cause* strongly entrenched after 1000 inputs.

### 6.2 Inferring Verb Meaning

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 known as syntactic bootstrapping (Gleitman, 1990). For example, Naigles (1990) showed that children who heard an intransitive utterance with a novel verb ( $U_I$ : *The bunny and duck are blicking*) were more likely to look at a picture of two characters independently performing an action, while those who heard a similar transitive form ( $U_T$ : *The bunny is blicking the duck*) were more likely to look at a picture of one character (*the bunny*) performing an action on the other, when told to “find” the novel action in the pair of scenes. 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.

We demonstrate this ability in our model as

follows. We create scene representations corresponding to the pictures shown to the children:

$S_{CA}$  : BLICK<sub>[cause,act]</sub>(DUCK<sub>(agent)</sub>, BUNNY<sub>(theme)</sub>)

$S_A$  : BLICK<sub>[act]</sub>(AND<sub>∅</sub>(DUCK, BUNNY)<sub>(agent)</sub>)

In one 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 former input pair corresponds to the appropriate selection of the scene to go along with the perceived utterance, and the latter to the inappropriate selection. (We form analogous pairings with the intransitive form,  $U_I$ .) We then (separately) input each pair to our model to have it extract a corresponding frame  $F$  and determine the best construction  $k$  for it. The model records the value from eqn. (1),  $P(k|F)$ , as its response to each input.

If the input pair yields a frame that matches an existing construction—that is, if the scene-utterance combination corresponds to a reliable association in the model—then the value of  $P(k|F)$  will be higher than if no such construction exists. (In the latter case, the best construction is a new one, with low prior probability.) Thus, when comparing the values recorded in response to the appropriate and inappropriate pairing for each utterance, a higher value of  $P(k|F)$  corresponds to the child “recognizing” the appropriate scene for the utterance.

Table 2 shows the value of  $\log P(k|F)$  across the conditions, after varying amounts of learning (10, 100, and 1000 input pairs; averaged over 10 simulations). 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.

### 6.3 The Use of Unusual Constructions

Like children, the model mistakenly overgeneralizes, but recovers from these errors only by receiving additional positive evidence (Alishahi and Stevenson, 2005). However, this ability to converge on appropriate argument structures for each verb should not prevent the language learner from making productive generalizations

Utterance	Scene	# Input Pairs		
		10	100	1000
$U_T$	$S_{CA}$ (matched)	-5	-5	-5
	$S_A$ (unmatched)	-9	-10	-11
$U_I$	$S_{CA}$ (unmatched)	-12	-13	-14
	$S_A$ (matched)	-4	-3	-3

Table 2:  $\log P(k|F)$  for matched and unmatched scene-utterance pairs.

such as *The fly buzzed into the room* (cf. example (2) in Section 1).

We test our model with a verb appearing in an unusual (for that verb) construction, to see whether the model can determine appropriate semantic properties. We add a new verb *dance* to the input generation lexicon, with one frame:

head verb	<i>dance</i>	
verb sem. primitives	$\langle act, manner \rangle$	
args	roles	$\langle agent \rangle$
	categories	$\langle animate \rangle$
syntactic pattern	arg1 <i>verb</i>	

After training on 1000 input pairs, in which *dance* appears only intransitively, we present the model with the following scene-utterance pair:

DANCE<sub>[???]</sub>(KITTY<sub>(?)</sub>, DOGGY<sub>(?)</sub>, UNDER<sub>∅</sub>(TABLE)<sub>(?)</sub>)  
*kitty dance doggy under table*

The scene representation has been modified to remove the semantic primitives of the verb and the roles of the arguments. Given this partial input, the model must predict multiple missing semantic features, based on the utterance.

Averaged across 10 simulations, the model predicts novel semantic primitives for the verb *dance* in this usage with a probability of .49 for *cause*, .43 for *move*, and negligible probabilities for other primitives. The roles predicted for the arguments, with associated probabilities, are *agent* (.90), *theme* (.94), and *goal* (.99). The model has generalized the feature values of a construction corresponding to the usage of a verb such as *put*, shown above in Figure 1.

We test this ability in sentence production as well, presenting the model with the full scene representation and predicting the most probable syntactic pattern. The sentence *kitty dance doggy under table* is produced as expected.

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).

## 7 Related Computational Models

Some recent usage-based models of language acquisition handle syntax/semantics interaction, and generalization to novel situations (Allen, 1997; Niyogi, 2002). For example, Niyogi (2002) proposes a Bayesian model that shows how syntactic and semantic features of verbs interact to support learning. In contrast to our model, the structure of the verb classes and their probabilities, as well as the probabilities of verbs showing particular features, are all fixed. The connectionist model of Allen (1997) is able to make interesting generalizations over argument structure syntax and semantics. However, learning of general constructions is implicit, and the acquired knowledge cannot be used in any language task other than limited comprehension.

Only a few computational models directly address learning of argument structure constructions. Chang (2004) presents a computational model which learns lexical constructions as a mapping between graphical representations of form (typically word order) and meaning (typically role-filler bindings) from annotated child data. Unlike our model, this approach relies on noise-free input and extensive prior knowledge, and constructions are not generalized across verbs. The model of Dominey (2003) learns constructions from narrated video images. The model successfully assigns semantic roles in familiar data, and allows limited generalization of this ability over new verbs. However, in contrast to our approach, learning is highly dependent on the unrealistic assumption of having each form

uniquely identify the associated meaning (i.e., forms and meanings are in a one-to-one mapping).

## 8 Discussion

We have described a Bayesian model for the representation, acquisition and use of argument structure constructions in a usage-based framework. The results of computational experiments with the model demonstrate the feasibility of learning general constructions from individual examples of verb usage, even in the presence of noisy or incomplete input data. We also show that the model can use its acquired construction-based knowledge to generalize to new or low-frequency verbs, or verbs appearing in unusual constructions. These results stem from our novel view of constructions as a mapping of a form to a probability distribution over semantic features, and the corresponding formulation of language learning and use as a Bayesian categorization and prediction process.

Psycholinguistic evidence suggests that children are aware of verb-independent regularities in comprehension much earlier than they can use them in production. Children may begin by learning *weak* constructions which enable only certain kinds of linguistic operations, but become more robust over time (Tomasello and Abbot-Smith, 2002). Our model follows this trend: Constructions emerge early, and can be used in recognizing appropriate scene-utterance inputs as demonstrated here; however, in production, the model generally exhibits an initial strict imitative phase, before a construction is entrenched enough to generalize (Alishahi and Stevenson, 2005).

An important result of our model is that it can account for recovery from overgeneralization, while at the same time allowing for productive generalization of “unusual” constructions. A question that arises is, what are the limitations of using a verb in a novel construction? Although many innovative uses of verbs are acceptable, many others are not. The distinction seems to come from the fundamental semantic properties of the verb; for example, one can say



*I hammered the metal flat but not I played the game finished.* We are currently considering a more sophisticated, fine-grained semantic representation for verbs and scenes, to enable the model to capture these effects.

## References

- Alishahi, A. and Stevenson, S. (2005). A probabilistic model of early argument structure acquisition. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. to appear.
- Allen, J. (1997). Probabilistic constraints in acquisition. In Sorace, A., Heycock, C., and Shillcock, R., editors, *Proc. of GALA97*, pages 300–305.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429.
- Bencini, G. M. L. and Goldberg, A. E. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language*, 43:640–651.
- Chang, N. (2004). Putting meaning into grammar learning. In *Proc. of the First Workshop on Psycho-computational Models of Human Language Acquisition*.
- 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*.
- Fillmore, C., Kay, P., and O'Connor, M. K. (1988). Regularity and idiomatcity 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.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1:135–176.
- 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.
- 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.
- 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*.
- Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. MIT Press.
- Rappaport Hovav, M. and Levin, B. (1998). Building verb meanings. In Butt and Geuder, editors, *The Projection of Arguments: Lexical and Computational Factors*, pages 97–134. CSLI Publications.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74:209–253.
- Tomasello, M. and Abbot-Smith, K. (2002). A tale of two theories: response to Fisher. *Cognition*, 83:207–214.