A Probabilistic Model of Early Argument Structure Acquisition

Afra Alishahi and **Suzanne Stevenson**
Department of Computer Science
University of Toronto
{afra,suzanne}@cs.toronto.edu

Abstract

We present a computational model of usage-based learning of verb argument structure in young children. The model integrates Bayesian classification and prediction to learn from utterances paired with appropriate semantic representations. The model balances item-based and class-based knowledge in language use, demonstrating appropriate word order generalizations, and recovery from overgeneralizations with no negative evidence or change in learning parameters.

Argument Structure Acquisition

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. Children, however, learn to correctly use common verbs quite early. Moreover, they grasp argument structure regularities at a young age, producing novel utterances that obey the mapping of arguments to syntactic positions in their language (Demuth et al., 2002; MacWhinney, 1995). Children even "correct" experimenters' non-SVO (subject-verb-object) usage of novel verbs to fit the SVO order of their ambient language as early as age 3 (Akhtar, 1999).

The ability to generalize observed argument structure patterns to novel situations sometimes leads to *overgeneralization* (Bowerman, 1982). For example, children often use an intransitive-only verb such as *fall* in a transitive construction, as in:

Adam fall toy. [Adam 2;3, CHILDES MacWhinney (1995)]

Such usages are not arbitrary errors, but rather generalizations of the association between causative action and transitive form, to an intransitive action verb.

Thus, acquisition of argument structure exhibits a U-shaped learning curve seen in other areas of language learning (Marcus et al., 1992): correct use of a verb may be followed by a period containing incorrect (overgeneralized) usages, before convergence on adult behaviour. Moreover, negative evidence (corrective feedback) plays little to no role in this process; only additional positive evidence of correct usages is necessary for "unlearning" of overgeneralized rules (Marcus, 1993).

Learning in this domain has been suggested to rely on rich innate knowledge of argument structure regularities (Pinker, 1989). However, recent psycholinguistic evidence has questioned this assumption, and a number of usage-based proposals have argued that children learn such regularities from the input alone (Bowerman, 1982; Akhtar, 1999; Tomasello, 2000; Demuth et al., 2002). In support of this view is evidence 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 (Tomasello, 2000). Indeed, Goldberg (1999) claims that it is the form-meaning mappings, or *constructions*, of a set of high-frequency verbs that serve as the basis for generalization of their associated forms to other verbs. Moreover, experimental evidence has shown that the frequency of an individual verb influences the likelihood that children accept it in an overgeneralized usage (Theakston, 2004).

To further test these ideas, explicit usage-based models must be explored, both of the underlying learning mechanisms and of the use of the acquired knowledge. Here, we present a computational model that elaborates a specific mechanism for how children learn the argument structure of individual items, and how this knowledge is generalized to new forms in language use.

Generalization in our model is achieved through a Bayesian classifier that groups similar argument structure frames. Each frame represents a combination of semantic and syntactic properties of a verb and its arguments in a particular usage. Frames with shared syntax are grouped according to probabilities over their semantic properties. The result is that the semantic primitives most frequently used across all frames in a class have the highest probabilistic association with the syntactic form. The emergent classes capture form-meaning pairings that generalize over the fine-grained semantics of both the verb and its arguments. While the resulting classes have a similar function to the constructions of Goldberg (or the event structure templates of Rappaport Hovav and Levin (1998)), in our case, the syntactic pattern is associated with a range of verb and argument meanings probabilistically.

A key property of our computational model is the interaction between item-based and class-based information, and how it plays a role in language use—both comprehension and production. Here, we focus on simulations of the child in utterance production and demonstrate how the classes of argument structure frames enable the model to generalize over observed forms. Examples of the model experiencing a period of overgeneralization illustrate that the probabilistic balance of item-based and class-based information may shift toward class knowledge when knowledge of particular verbs is infrequent. Subsequent recovery occurs with additional positive evidence due to the increased strength of item-based

Scene-Utterance Input Pair			
$\text{TAKE}_{[cause,move]}(\text{TIM}_{\langle agent\rangle}, \text{BOOKS}_{\langle theme\rangle}, \text{TO}_{[]}(\text{SCHOOL}_{\langle goal\rangle}))$ *Tim took books to school.*			

Extracted Frame		
head verb		*take*
semantic primitives of verb		⟨*cause, move*⟩
arguments	roles	⟨*agent, theme, goal*⟩
	categories[a]	⟨*human, concrete, dest-pred*⟩
syntactic pattern		arg1 *verb* arg2 arg3

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

Figure 1: An input pair and its corresponding frame.

Algorithm 1 Learning an argument structure frame

1. Extract frame F from a scene-utterance pair (see Figure 1), and access lexical entry L of the head verb of F.
2. If F is incomplete, then use both the frame structure of L and the information stored in the classes to predict the best value for the missing part(s) of F.
3. Update the lexicon:
   (a) If L already contains a matching frame F', then increase the frequency of both F' and its parent class;
   (b) Else add F to L, and find the best class for F (creating a new class if needed).

knowledge in the probability formulas (*not* due to explicit negative evidence, nor any change in the learning parameters).

Overview of the Computational Model

Learning Argument Structure Knowledge

We assume that the input to the argument structure acquisition process consists of pairs of representations, one for the perceived utterance (what the child hears), and one for the relevant aspect of the observed scene (the semantics described by the utterance).[1] The first step for our learning model is to extract the corresponding argument structure frame for the main verb of the utterance.[2] Each frame contains a set of features drawn from the scene-utterance pair, as shown in Figure 1.

Each observed frame is stored in the lexical entry of the verb. If the current extracted frame has been previously observed with the verb, then the frequency of the stored frame is increased, otherwise a new frame is inserted. Some of the frames of a verb may merge to form a more general one (e.g., if two frames are identical except that the semantic types of the arguments of one frame are more general than the other).

Any new frame (whether newly observed or the output of a merging process) is input to the incremental Bayesian classifier, which groups the new frame together with an existing class of frames that probabilistically has the most similar properties to it. If none of the existing classes has sufficiently high probability, then a new class is created. The probability of each class is determined by both syntactic and semantic features. Currently, a class 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 semantic primitives is determined by how frequently those of the frame occur across the frames of the candidate class. The probability of semantic categories of arguments is calculated similarly, taking into account the relationship of categories in the ontology.

The overall learning process of the model is summarized in Algorithm 1.

[1]Picking out the specific semantic descriptor that matches an utterance from the full scene representation is itself a challenge for the child. We assume that this "word-meaning mapping" has been performed, e.g., as in Siskind (1996).

[2]Here, we focus on *verb* argument structure, but our model applies to other predicates as well.

Using the Acquired Knowledge

As mentioned, both item-based and more general class-based information are updated with the processing of each input. A key property of our model is how these two sources of information interact in the use of language during the course of acquisition. We formalize a number of tasks in language use as different versions of a prediction problem. For example, sentence production is seen as predicting the syntactic pattern of a frame based on its semantic components; in comprehension, argument roles may be assigned, or partial meaning of a verb or noun induced, based on the participants in the scene and the syntax of the utterance. Hand in hand with our Bayesian classifier is a Bayesian prediction process for use in such language tasks.

The prediction process integrates item-based and class-based information to make probabilistic predictions about argument structure information based on available frame features. If the child's knowledge of a particular use of a verb is sufficiently complete and frequent to have become *entrenched*, then that information will be used in both comprehension and production. On the other hand, by generalizing over known frames, classes enable predictions about missing or low-confidence item-based information. For example, in production, such predictions allow the child to use a verb in a novel (for that verb) syntactic pattern, as long as semantically similar verbs have been observed in that usage.

Bayesian Classification and Prediction

The Bayesian Classifier

The model we use for classification is an adaptation of a model of human categorization proposed by Anderson (1991).[3] This model has desirable properties for our domain. Its use of probabilities over observed information captures the emphasis on item frequencies in child language acquisition. Also, the classification model is incremental, enabling us to classify frames as they are observed.

Classification of a frame F is formalized as a process of maximizing the probability of class k given the frame:

$$\textbf{BestClass}(F) = \underset{k}{\operatorname{argmax}}\ P(k|F) \qquad (1)$$

[3]Our classes are not pre-defined, as is often intended by the term in machine learning. Our classes/classification process could as well be termed clusters/clustering.

where k ranges over the indices of all classes, including an index of 0 to represent creation of a new class. Using Bayes rule, this probability can be rewritten as:

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

The prior probability of a class k is given by:[4]

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

where n is the total number of observed frames; n_k is the number of frames in class k, for $k > 0$; and $n_0 = 1$. Thus, the estimation of the prior probability of an existing class is proportional to the frequency of frames in that class, and the probability of a new class is inversely proportional to the number of observed frames overall.

The probability of a frame F is expressed in terms of the individual probabilities of its features (shown above in Figure 1). Although these features in reality interact (e.g., the role of an argument and the syntactic position it occurs in are interrelated), we adopt an independence assumption to make the calculation feasible. Thus, the conditonal probability of a frame F is the product of the conditional probabilities of its features:

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

where j is the value of the ith feature of F, and $P_i(j|k)$ is the probability of displaying value j on feature i within class k. This probability is estimated using a smoothed maximum likelihood formulation.

The Corresponding Prediction Model

Once the child learns the information above—i.e., the individual frames stored with each verb, along with the class structure over them—we must consider how this knowledge is used. An important aspect of our model is that essentially the same Bayesian framework can be employed in *using* the knowledge as that which acquires it. We formulate a language task as a prediction problem, in which missing feature values in a frame are filled in based on the most probable values given the available features. Following Anderson (1991):

$$\mathbf{BestValue}_i(F) = \underset{j}{\operatorname{argmax}}\ P_i(j|F) \quad (5)$$

where F is the partial frame, i is the missing feature, and j ranges over the possible values of i.

In Anderson's model, the classes are used in the calculation of $P_i(j|F)$ to determine the most probable value for the missing feature. However, the structure of our acquired knowledge is more complex than that of his category structures. In addition to the groupings of frames into classes, we also have the groups of frames associated with each particular verb. Thus there are two potential levels of generalization, rather than one—over the frames associated with a single verb (item-based), and over all frames, through the full class structure (class-based).

[4]This formula is the same as that used by Anderson (1991) with his "coupling probability" set to the mid value of 0.5.

One option is to limit generalization to only the frames associated with the verb:

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

where *classes*(v) is the set of indices of the classes of the frames learned for verb v, such that k_v is a class containing a frame of verb v. Thus we use the classes associated with a verb to probabilistically predict missing values ($P_i(j|k_v)$), and weight those predictions by the likelihood of the class given the partial frame ($P(k_v|F)$).

However, this formulation ignores the information embedded in the class structure more generally, unnecessarily restricting the child when a partial frame for a verb does not match well with any frames seen previously *for that verb*. We require a model in which prediction of missing features takes into account both the knowledge of likely values across the frames associated with the given verb, as well as the knowledge of likely values associated with *any* class compatible with the partial frame. That is, we must balance item-based knowledge with class-based knowledge.

Interestingly, we can achieve the item- and class-based integration by incorporating the classification process into our probability model for prediction. Essentially, we classify a partial frame F using equation (1) as if it were a new frame, and treat it during prediction as if it were inserted into the lexical entry for v with a frequency of 1. The best class k_F (across all classes) for F may or may not be a class already linked to by a frame of v, so we must modify equation (6) by extending (if necessary) *classes*(v), over which k_v takes its values, to include k_F. This ensures that both the overall best class, as well as the classes associated with the verb, are taken into consideration in predicting values for a partial frame.

The probability $P(k_v|F)$ in equation (6) is rewritten using Bayes rule (cf. equation (2)):

$$P(k_v|F) = \frac{P(k_v)P(F|k_v)}{P(F)} = \frac{P(k_v)P(F|k_v)}{\sum_{k'_v} P(k'_v)P(F|k'_v)} \quad (7)$$

The factor $P(F|k_v)$ is calculated as in equation (4), using a uniform probability distribution over the possible values of the missing feature. The prior probability of the class, $P(k_v)$, is calculated taking into account only the classes in *classes*(v) (including k_F), not the full class structure. In calculating $P(k_v)$, the frequency of each class (its total number of frames) is weighted by the frequency of the frame for v which points to it, balancing the overall likelihood of the class with the likelihood that it is a class for v.

If k_F was not previously a class for v, then the weight from the frame frequency is only 1. Thus, a "new" class (new to v) for F has more influence the less often the verb has been seen overall; if the verb has been seen frequently, then the weight from its observed frames to their classes will outweigh the influence of the single partial frame to its "new" class. (Cf. the influence of a new class achieved by equation (3).) Thus, class information outside the verb is always a factor in prediction, but will have decreased influence with increased item-based frequency.

Experimental Set-Up

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. We assume that a small number of nouns have been previously learned by the child, forming a simple ontology that indicates their semantic category. As illustrated in Figure 1, we use a simple logical form for representing an observed scene. We assume that at the stage of learning being modelled, the child has reliable hypotheses about the assignment of roles to arguments.

In the current implementation, a syntactic pattern is limited to the order of the arguments with respect to the predicate term. Also, for now we do not address learning of morphology; all words appear only in their root forms.

The Input Corpora

We create input lists of scene-utterance pairs that conform to the distributional characteristics of the data children receive from their parents. We extracted from CHILDES 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). The 13 verbs in common across these three lists were added to an input-generation lexicon, along with their frequency and a unique semantic symbol. We also assign each verb a set of possible argument structure frames and associated frequencies, which are manually compiled by examination of all uses of a verb in all conversations of the same three children. Prepositions used in these conversations were also added to the lexicon.

For each simulation in our experiments, an input corpus of scene-utterance pairs is automatically created from this input-generation lexicon, using the frequencies 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.

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.

Experimental Results

We focus on results of our model on the sentence production task, for comparison to actual child data on verb use. The simulations we report all use the same parameter settings; only the randomly generated input corpus differs. We first describe the learning curves displayed by our model for some example verbs. We then examine some interesting stages in the generally observed U-shaped curve in more detail: imitation, generalization, overgeneralization and recovery.

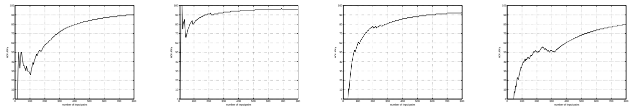

Figure 2: Sample learning curves for different verbs. X-axis is time (# of inputs); y-axis is cumulative accuracy.

The Learning Curve

As an item-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. For each verb, we ran eight separate simulations of our model over 800 input pairs: 200 sequences of 4 complete input pairs followed by a 5th test input (using the target verb) in which the utterance was removed. Our prediction model was used to find the syntactic pattern with the highest probability for the resulting partial frame. After each test input, we measured the cumulative accuracy of the model by counting the total number of times the predicted pattern was exactly the same as that used in the removed utterance.

Figure 2 shows a sample learning curve for each of the verbs *go*, *come*, *eat* and *fall*. Since the input corpora are randomly generated, the performance of the model varies across simulations in the early stages. For frequent verbs with a variety of argument structures, such as *go* and *come*, a U-shaped curve is often observed. A verb with fewer types of frames, such as *eat*, is less often overgeneralized. The learning curve for *fall*, which is less frequent, shows a delay compared to more frequent verbs.[5]

Imitation and Generalization

To study generalization in our model, we examine its behaviour when presented with a novel verb for which it must produce a sentence. After training the model, we present it with only a scene representation, whose main predicate corresponds to a verb, *gorp*, which has not been seen in an utterance:

$$\text{GORP}_{[cause,act]}(\text{KITTY}_{\langle agent\rangle}, \text{DOGGY}_{\langle theme\rangle})$$

Since this semantics resembles the typical transitive construction, we expect the model to predict the transitive pattern, "arg1 *verb* arg2", despite its lack of knowledge of this verb. To observe the pattern of responses over the course of acquisition, we test the model after varying amounts of training data. Averaging over 100 simulations on different input corpora, the model predicts this pattern with 68% probability after 5 input pairs, but with 99% probability after processing 50 input pairs.

More interesting is the varying influence of item-based and class-based knowledge 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

[5] In addition to classical overgeneralization errors, other errors reflected in the learning curves include incorrect learning from noisy input, or from the model simply not having learned the required usage.

		# Input Pairs		
		50	100	150
Generated Pattern	observed (SOV)	0.98	0.53	0.11
	"corrected" (SVO)	0.02	0.47	0.89

Table 1: Average probability of the predicted patterns.

(1999), in which English-speaking children aged 2 to 4 were taught novel verbs used in non-standard (SOV or VSO) orders. In productions of these verbs, 2- and 3-year-olds matched the observed patterns roughly half the time and "corrected" the order to SVO roughly half the time. The 4-year-olds rarely matched the observed order, almost always correcting to SVO order.

We trained our model using different numbers of input pairs to simulate differing amounts of exposure to the language. We then provided to the model a training input pair with a novel verb used in a non-standard (SOV) order (*kitty doggy gorp* for the scene representation above). Next, we presented the same scene representation to the model, without the utterance, and recorded the syntactic pattern predicted by the model. We performed 100 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 1, the model, like children, shows a shift from imitation, where it repeats even an unusual form, to generalization, where it relies on the ubiquitous patterns, the more exposure it has to the language.

Overgeneralization and Recovery

We noted that use of general knowledge sometimes leads children to *overgeneralize*, but they eventually recover with only additional positive evidence. A typical overgeneralization is when a non-causative verb is used as causative, e.g., *Don't you fall me down* (Bowerman, 1982). We tracked the usage of *fall* by our model to see if we can detect a similar pattern 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). Each use of *fall* in these simulations includes a causal agent in the scene description with a 0.5 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 pattern with *fall* in a scene with a causal agent.

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 semantic representation with *fall* as its main predicate, and the prediction model chose the best syntactic pattern for it. (Any place-holders for arguments that were not present in the scene were left blank.) Over 10 such simulations, the probability of receiving a training pair containing *fall* varied from 0–0.02, and the probability of receiving an instance of the transitive construction was between 0.14–0.21. In 7 out of 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 saw more examples of *fall*.

The first 8 uses of *fall* for Adam in CHILDES (at 27 months) are given below, with the first 8 sentences generated by our model in one of the simulations, illustrating the mix of the two patterns at this stage:

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

By the age of 31 months, the causative (transitive) uses of *fall* gradually disappear from Adam's conversations. Over the 7 simulations in our model showing the pattern of overgeneralization and recovery, causative sentences were no longer output after processing an average of 136 training inputs.

Summary of Results

Taken together, these experiments show an impressive match between the general behaviour of the model and that of children concerning the interplay between item-based and class-based knowledge in acquisition of argument structure. Imitation of observed forms occurs early in acquisition, but as evidence of general patterns increases, so does the tendency to generalize. This tendency can even overwhelm infrequent verbs used in less common constructions, such that a period of overgeneralization may set in. However, simply receiving additional examples of the verb in its correct usage can guide the model to recovery from overgeneralization. The model achieves this range of behaviour across the course of acquisition with no explicit negative evidence, nor even changes in the learning parameters, which are held constant. The results are simply the consequence of the Bayesian classification model and the unique interaction of class-based and item-based knowledge in the corresponding Bayesian prediction process.

Related Work

A number of recent computational models take an item-based approach to language acquisition. Niyogi (2002) proposes a Bayesian model that shows how syntactic and semantic features of verbs interact to support learning. However, 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. Chang's (2004) model successfully learns multi-word constructions (form-meaning pairs) from child data annotated with scene representations, but relies on noise-free input and extensive prior knowledge, and constructions are not generalized across verbs. 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.

Existing models also deal with other aspects of the problem we discuss here. The connectionist model of Desai (2002) learns a miniature language from a set of scene-sentence pairs, but prediction is limited to a component of the meaning of a novel word based on its syntactic context. Buttery (2003) is focused on learning syntax and addresses the learning of meaning only at the word level. Onnis et al. (2002) present evidence that child-directed input has statistical properties that enable the learner to recover from overgeneralization of argument structure patterns, given the selection of the "simplest" grammar, but they do not develop an actual model of grammar learning.

Conclusions and Future Work

Our computational model demonstrates the feasibility of learning argument structure regularities from examples of verb usage, and suggests acquisition mechanisms underlying this process in children. The model exploits item frequencies within a Bayesian classification process; the explicit use of classes allows the model to capture relevant generalizations without having to consider all of the lexical frames learned to that point. A novel formulation views language use as a Bayesian prediction process; a single probability formula smoothly integrates item-based and class-based information in predicting needed argument structure properties. This probabilistic approach makes the model robust against noisy input and low-confidence information. Furthermore, the model can apply its acquired knowledge across a variety of language tasks. Here we have focused on the task of sentence production. Our simulations of the model over the course of acquisition show a promising match with child data in terms of the observed stages of imitation, generalization, possible overgeneralization, and eventual convergence on correct argument structure usage.

The model in its current form makes simplifying assumptions that must be addressed in future work. For example, the input includes the semantic primitives for the coarse-grained semantics of the verbs, as well as the assignment of roles to arguments. A full usage-based account of argument structure acquisition must show how these are learnable from the input—indeed, a more likely scenario is that acquisition of such aspects is interleaved with the learning of the properties discussed here. We believe that using a distributed representation of roles, e.g., as in Allen (1997), will begin to address this issue by enabling the model to assign roles probabilistically. We also need to further develop our acquisition mechanism to account for learning of collocations, idiomatic phrases and fine-grained selectional preferences. One approach might begin by maintaining a probability distribution over the words that participate in each argument position of a frame, raising additional interesting issues in generalization of knowledge.

References

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

Allen, J. (1997). Probabilistic constraints in acquisition. In Sorace, A., Heycock, C., and Shillcock, R., editors, *Proc. of GALA97*, 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.

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

Chang, N. (2004). Putting meaning into grammar learning. In *Proc. of the 1st Workshop on Psychocomputational Models of Human Language Acquisition*.

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

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

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.

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

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.

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

Rappaport Hovav, M. and Levin, B. (1998). *Building verb meanings*, 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.

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.