Proceedings of the 2nd European Cognitive Science Conference, 2007, Delphi, Greece.

A Computational Usage-Based Model for Learning General Properties of Semantic Roles

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

Abstract

We present a Bayesian model of early verb learning that acquires a general conception of the semantic roles of predicates based only on exposure to individual verb usages. The model forms probabilistic associations between the semantic properties of arguments, their syntactic positions, and the semantic primitives of verbs. Because of the model's Bayesian formulation, the roles naturally shift from verb-specific to highly general properties. The acquired role properties are a good intuitive match to various roles, and are useful in guiding comprehension in the face of ambiguity.

Learning and Use of Semantic Roles

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

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

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

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

Early theories suggested that roles are drawn from a pre-defined inventory of semantic symbols or relations, and that innate "linking rules" that map roles to sentence structure enable children to infer associations between role properties and syntactic positions (e.g., Pinker, 1989). However, numerous questions have been raised concerning the plausibility of innate linking rules for language acquisition (e.g., Fisher, 1996; Kako, 2006).

An alternative, usage-based view is that children acquire roles gradually from the input they receive, by generalizing over individually learned verb usages (e.g., Lieven et al., 1997; Tomasello, 2000). For instance, Tomasello (2000) claims that, initially, there are no general labels such as Agent and Theme, but rather verb-specific concepts such as 'hitter' and 'hittee,' or 'sitter' and 'thing sat upon.' Recent experimental evidence confirms that access to general notions like Agent and Theme is age-dependent (Shayan, 2006). It remains unexplained, though, precisely how verb-specific roles metamorphose to general semantic roles. Moreover, experiments with children have revealed the use of verb-specific biases in argument interpretation (Nation et al., 2003), as well as of strong associations between general roles and syntactic positions (e.g., Fisher, 1996, and related work). However, specific computational models of such processes have been lacking.

We have proposed a usage-based computational model of early verb learning that uses Bayesian clustering and prediction to model language acquisition and use. Our previous experiments demonstrated that the model learns basic syntactic constructions such as the transitive and intransitive, and exhibits patterns of errors and recovery in their use, similar to those of children (Alishahi and Stevenson, 2005a,b). A shortcoming of the model was that roles were explicit labels, such as Agent, which were assumed to be "perceptible" to the child from the scene. In this paper, we have extended our model to directly address the learning and use of semantic roles.

Our Bayesian model associates each argument of a predicate with a probability distribution over a set of semantic properties—a *semantic profile*. We show that initially the semantic profiles of an argument position yield verb-specific conceptualizations of the role associated with that position. As the model is exposed to more input, these verb-based roles gradually transform into more abstract representations that reflect the general properties of arguments across the observed verbs.

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

Related Computational Work

A number of computational approaches for learning the selectional preferences of a verb first initialize WordNet

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

concepts with their frequency of use as the particular argument of a verb, and then find the appropriate level in the WordNet hierarchy for capturing the verb's restrictions on that argument (e.g., Resnik, 1996; Clark and Weir, 2002). However, none of these models generalize their acquired verb-based knowledge to a higher level, yielding constraints on the arguments of general constructions such as the transitive or intransitive.

Many computational systems model human learning of the assignment of general roles to sentence constituents, using a multi-feature representation of the semantic properties of arguments (e.g., McClelland and Kawamoto, 1986; Allen, 1997). Others learn only verb-specific roles that are not generalized (e.g., Chang, 2004). As in our earlier work, these models require explicit labelling of the arguments that receive the same role in order to learn the association of the roles to semantic properties and/or syntactic positions. In the work presented here, we show that our extended model can learn general semantic profiles of arguments, without the need for role-annotated training data.

Our Bayesian Model

Our model learns the argument structure frames for each verb, and their grouping across verbs into constructions. An argument structure frame is the pairing of a syntactic form (a particular word order of a verb and its arguments) with the meaning of the expression (the semantic primitives of the predicate and the semantic properties of the arguments). A construction is a grouping of individual frames which probabilistically share form-meaning associations; these groupings typically correspond to general constructions in the language such as transitive, intransitive, and ditransitive.

Most importantly for this paper, the model forms probabilistic associations between syntactic positions of arguments, their semantic properties, and the semantic primitives of the predicate. These associations are generalized (through the constructions) to form more abstract notions of role semantics, dependent on argument position and verb primitives. The following sections review basic properties of the model from Alishahi and Stevenson (2005a,b), and introduce extensions that support the learning and use of semantic profiles.

The Input and Frame Extraction

The input to the learning process is a set of scene-utterance pairs that link a relevant aspect of an observed scene (what the child perceives) to the utterance that describes it (what the child hears). From each input pair, our model extracts the corresponding argument structure frame, which is a set of form and meaning features.

Figure 1 shows that we use a simple logical form for representing the semantics of an observed scene, while an utterance simply specifies a sequence of words in root form. In the extracted frame, verbs and prepositions are represented as predicates (e.g., **Make**, **On**) that can take a number of arguments. Each predicate has a set of semantic primitives which describes the event characteristics (e.g., [*cause*, *become*]). Each argument can be an entity (e.g., TIM, CAKE) or a predicate structure itself (e.g., **On**(TABLE)). The syntactic pattern in the frame indicates word order of the predicate and arguments.

Scene-Utterance Input Pair
Scene: $\mathbf{Make}_{[cause, become]}$(TIM, CAKE)
Utterance: *Tim made cake*

Extracted Frame	
head verb	*make*
semantic primitives of verb	[*cause, become*]
arguments	⟨*Tim, cake*⟩
syntactic pattern	arg1 *verb* arg2

Figure 1: An input pair and its corresponding frame.

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

Figure 2: Semantic properties for *cake* from WordNet

In a frame, each word for an entity has a link to the lexical entry that contains its semantic properties, which are extracted from WordNet (version 2.0) as follows. We hand-pick the intended sense of the word, extract all the hypernyms (ancestors) for that sense, and add all the words in the hypernym synsets to the list of the semantic properties. Figure 2 shows an example of the hypernyms for *cake*, and its resulting set of semantic properties.[2]

Learning as Bayesian Clustering

Each extracted 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 for the new frame, then a new construction is created, containing only that frame. We use a modified version of Alishahi and Stevenson's (2005a,b) probabilistic model, which is itself an adaptation of a Bayesian model of human categorization proposed by Anderson (1991). It is important to note that the categories (i.e., constructions) are not predefined, but rather are created according to the patterns of similarity over observed frames.

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

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

where k ranges over the indices of all constructions, with index 0 representing recognition of a new construction.

Using Bayes rule, and dropping $P(F)$ which is constant for all k:

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

The prior probability, $P(k)$, indicates the degree of entrenchment of construction k, and is given by the relative frequency of its frames over all observed frames. The posterior probability of a frame F is expressed in terms of

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

the individual probabilities of its features, which we assume are independent, thus yielding a simple product of feature probabilities:

$$P(F|k) = \prod_{i \in FrameFeatures} P_i(j|k) \quad (3)$$

where j is the value of the i^{th} feature of F, and $P_i(j|k)$ is the probability of displaying value j on feature i within construction k. Given the focus here on semantic profiles, we return to the calculation of the probabilities of semantic properties below.

Language Use as Bayesian Prediction

In our model, language use (production and comprehension) is a prediction process in which unobserved features in a frame are set to the most probable values given the observed features. For example, sentence production predicts the most likely syntactic pattern for expressing intended meaning components, which may include semantic properties of the arguments and/or semantic primitives of the predicate. In comprehension, these semantic elements may be inferred from a word sequence.

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

$$\begin{aligned}\mathbf{BestValue}_i(F) &= \operatorname*{argmax}_j P_i(j|F) \quad (4)\\ &= \operatorname*{argmax}_j \sum_k P_i(j|k)P(k|F)\\ &= \operatorname*{argmax}_j \sum_k P_i(j|k)P(k)P(F|k)\end{aligned}$$

Here, F is a partial frame, i is an unobserved (missing) feature, j ranges over the possible values of feature i, and k ranges over all constructions. The conditional probabilities, $P(F|k)$ and $P_i(j|k)$, are determined as in our learning module. The prior probability of a construction, $P(k)$, takes into account two important factors: its relative entrenchment, and the (smoothed) frequency with which the verb of F participates in it.

All predictions of the model are mediated by construction knowledge. For a well-entrenched usage of a verb, predictions are guided by the construction that the usage is a member of. For a novel verb, or a novel use of a known verb, predictions arise from constructions that are the best match to the observed partial frame.

Probabilities of Semantic Properties

In both learning and prediction, the probability of value j for feature i in construction k is estimated using a smoothed version of this maximum likelihood formula:

$$P_i(j|k) = \frac{\mathbf{count}_i^k(j)}{n_k} \quad (5)$$

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

For most features, $\mathbf{count}_i^k(j)$ is calculated by simply counting those members of construction k whose value for feature i exactly matches j. However, for the semantic properties of words, counting only the number of exact matches between the sets is too strict, since even highly similar words very rarely have the exact same set of properties. We instead compare the set of semantic properties of a particular argument in the observed frame, S_1, and the set of semantic properties of the same argument in a member frame of a construction, S_2, using the Jaccard similarity score:[3]

$$\mathbf{sem\_score}(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} \quad (6)$$

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

cake:	{baked goods,food,solid,substance,matter,entity}
water:	{liquid,fluid,food,nutrient,substance,matter,entity}

The intersection of the two sets is {food, substance, matter, entity}, therefore the **sem_score** for these sets is $\frac{4}{9}$.

In general, to calculate the conditional probability for the set of semantic properties, we set $\mathbf{count}_i^k(j)$ in equation (5) to the sum of the **sem_score**'s for the new frame and every member of construction k, and normalize the resulting probability over all possible sets of semantic properties in our lexicon.

Representation of Semantic Profiles

Recall that a semantic profile is a probability distribution over the semantic properties of an argument position. This requires looking at the probability of all the individual properties, j_p, rather than the probability of the full set, j. We use a modified version of equation (5) in which $\mathbf{count}_i^k(j_p)$ is the number of frames in construction k that include property j_p for the argument whose set of semantic properties is the i^{th} feature of the frame. The resulting probabilities over all j_p form the semantic profile of that argument.

A semantic profile contains all the semantic properties ever observed in an argument position. As learning proceeds, a profile may include a large number of properties with very low probability. In order to display the profiles we obtain in our results section below, we create truncated profiles which list the properties with the highest probabilities, in decreasing order of probability value. To avoid an arbitrary threshold, we cut the ordered list of properties at the widest gap between two consecutive probabilities across the entire list.

Experimental Results

The Input Corpora

Large scale corpora of utterances paired with meaning representations, such as required by our model, do not currently exist. The corpora for our experimental simulations are generated from an extended version of the input-generation lexicon used in our earlier work. The lexicon is created to reflect distributional characteristics of the data children are exposed to. We extracted 13 of the most frequent verbs in mother's speech to each of

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

Adam (2;3–4;10), Eve (1;6–2;3), and Sarah (2;3–5;1) in the CHILDES database (MacWhinney, 1995). We entered each verb into the lexicon along with its total frequency across the three children, as well as its (manually compiled) set of possible argument structure frames and their associated frequencies. We also randomly selected 100 sentences containing each verb; from these, we extracted a list of head nouns and prepositions that appear in each argument position of each frame, and added these to the lexicon.

For each simulation in our set of experiments, an input corpus of scene-utterance pairs is automatically randomly generated using the frequencies in the input-generation lexicon to determine the probability of selecting a particular verb and argument structure. Arguments of verbs are also probabilistically generated based on the word usage information for the selected frame of the verb. Our generated corpora are further processed to simulate both incomplete and noisy data: 20% of the input pairs have a missing (syntactic or semantic) feature; another 20% of the pairs have a feature removed and replaced by the value predicted for the feature at that point in learning. The latter input pairs are noisy, especially in the initial stages of learning. Other types of noise (such as incomplete sentences) are not currently modelled in the input.

Formation of Semantic Profiles for Roles

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

We train our model on 200 randomly generated input pairs,[4] and then present it with a test input pair containing a novel verb *gorp* appearing in a familiar construction, with unknown nouns appearing as its arguments. As an example, a test pair for a novel verb appearing in a transitive construction looks as follows:

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

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

Our model learns semantic profiles for argument positions in a range of constructions. Here, due to lack of

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

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

Figure 3: Semantic profiles of argument positions.

space, we focus only on a few such profiles corresponding to roles that have received much attention in the literature. Figure 3 shows the predicted semantic profiles for the arguments in the subject and object positions of a transitive construction (corresponding to X and Y in the *gorp* test input above), and the subject position in an intransitive construction. Even though we use semantic properties from WordNet, which lack any event-specific features, the emerging semantic profile for each argument position demonstrates the intuitive properties that the role received by that argument should possess.

For example, the semantic profile for an argument that appears in the subject position in a transitive construction (the left box of Figure 3) demonstrates the properties of an animate entity, most likely a human. In contrast, the semantic profile for an argument in the object position (the middle box of Figure 3) most likely corresponds to a physical entity. These results are consistent with Kako (2006), who finds that, given unknown verbs and nouns in the transitive, adults attribute more Agent-like semantic properties to subjects, and more Patient-like properties to objects (specifically, a variation on Dowty's (1991) proto-properties). Kako (2006) also reports similar results using a *known* verb in an incompatible construction. To simulate this, we gave our model a transitive input like the *gorp* pair above, but with an intransitive-only verb *dance*. We found a similar profile for the noun in object position as in Figure 3. Since *dance* has never been seen with an object, this profile must come from the model's learned associations over existing verbs in the transitive construction.

The right box of Figure 3 shows the semantic profile for the subject of an intransitive. This argument can receive an Agent role (*John went*) or a Theme role (*the ball fell*). We think that the semantic profile represents more Agent-like characteristics because, in our input data, the probability of an Agent in that position is much higher than a Theme. We explore this issue next.

Multiple Possible Roles for a Position

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

To test this, we created an input corpus with an artificially increased frequency of *fall*, the only intransitive verb in our lexicon that can have a Theme (rather than an Agent) in the subject position, so that the model would be given sufficient examples of such a usage. We then tested the model with two kinds of novel verbs: one with semantic primitives [*act,move*] (associated with agentive intransitives like *come* and *go*), and one with semantic primitives [*move,direction,vertical*] (like the Theme-assigning verb *fall*). In response to the former (*go*-type) input, the model still predicts a semantic profile very similar to the one shown in the right box of Figure 3. In contrast, for the latter (*fall*-type) input, the predicted semantic profile contains {*artifact, artefact, whole, whole thing, unit*} in addition to the Agent-like properties, yielding a profile that overlaps that of the transitive object, shown in the middle box of Figure 3.

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

Verb-Based vs. General Semantic Profiles

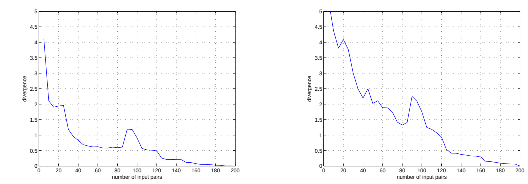
Tomasello (2000) (among others) has proposed that children initially learn verb-specific roles such as 'hitter' and 'hittee,' and only later move to more general roles. Moreover, Shayan (2006) shows that general notions like Agent and Patient develop over time. Our model illustrates how such role generalization might come about. Although the semantic profiles of our model reflect the general properties that a particular role-bearing argument must have, they are formed from input that contains only argument properties for specific verb usages. The generalization occurs as more and more semantic properties are associated with an argument position, and only the most general ones are seen sufficiently frequently to have high probability.

We tracked the generalization process for each semantic profile, to see how it moves from a verb-based profile to a more general one. Figure 4 (left box) shows the semantic profile for the argument in the object position right after the first transitive usage. In this particular simulation, the first transitive verb in the corpus is *eat*, and its second argument in that input pair is *pie*. The semantic profile thus reflects the properties of a pie, and not the general properties of that argument position. The profile becomes more general after processing 50 and 100 input pairs, shown in the middle and right boxes of Figure 4, respectively. (Recall that Figure 3 shows a profile for transitive object after 200 inputs.)

To observe the trend of moving from a more specific to a more general semantic profile for each argument position, we used relative entropy to measure the divergence between the semantic profile for an argument position at a given point in learning, and the profile for that position that the model eventually converges to at the end of each simulation.[5] We measured the profile divergence for subject and object positions of a transitive construction after every 5 input pairs over a total of 200 pairs, averaged over 5 simulations. Figure 5 shows that the profile for the subject position (i.e., the Agent) is learned faster than the profile for the object position (i.e., the Theme), which is a much less constrained role. The curves show that the model stabilizes on the final profiles at around 150 input pairs, when receiving more inputs ceases to make any significant difference in the profiles.

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

Figure 4: The evolution of the transitive object role.



Transitive Subject Transitive Object

Figure 5: Learning curves for semantic profiles. The x-axis is time (#inputs), and the y-axis is divergence from the profile that the model eventually converges to.

Using Semantic Profiles in Comprehension

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

We consider cases using the verb *give*, in which an utterance beginning *you give* $\langle noun \rangle$ can continue with

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

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

either a second object (*you give* $\langle noun \rangle$ *something*) or a prepositional phrase (*you give* $\langle noun \rangle$ *to someone*). In the first case, $\langle noun \rangle$ is the Recipient of the action, and in the second case, it is the Theme (the thing given). We vary $\langle noun \rangle$ to be either *her*, which is a likely Recipient, or *this*, which is a likely Theme. We then observe in each case which interpretation our model prefers.

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

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

We give each of these inputs to the model, have it extract the corresponding frame F, and then have it calculate $match\_score(F) = \max_k P(k)P(F|k)$, from the prediction model. The *match_score* corresponds to how well the model thinks the given input pair matches an existing construction. Having a higher *match_score* for the first pair means that the model has recognized *her* as the Recipient, while a higher score for the second pair means that the model has recognized *her* as the Theme.

We also compared analogous inputs pairs, using *this* instead of *her*. Since the only difference in these two sets of inputs is the use of *her* vs. *this*, differences in interpretation of the first object noun (as Recipient or Theme) depend only on the match of the noun's semantic properties to the semantic profiles of that argument position.

Table 1 shows the results after processing 20 and 200 input pairs, averaged over 5 simulations; a higher number (lower absolute value) indicates the preferred interpretation. After processing 20 pairs, the model displays a strong preference towards treating the first object as the Theme, for both *her* and *this*. This occurs because *give* is the only verb in our lexicon that appears in a ditransitive (Recipient-first) construction, whereas many high frequency verbs (e.g., *put* and *get*) appear in the competing (Theme-first) prepositional construction. Thus, at the early stages of learning, the ditransitive construction is relatively weak. However, after processing 200 input pairs, the model shows a preference for the "correct" interpretation in both cases. (The difference between the two frames for *you gave this x* is small, but consistently indicates a Theme preference across all simulations.)

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

Conclusions

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

Table 1: log(*match_score*) for Recipient-first and Theme-first frames after processing 20 and 200 input pairs.

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

References

Alishahi, A. and Stevenson, S. (2005a). The acquisition and use of argument structure constructions: A Bayesian model. In *Proc. of the 2nd Wkshp on Psycho-computational Models of Human Lang. Acqn.*, pages 82–90.

Alishahi, A. and Stevenson, S. (2005b). A probabilistic model of early argument structure acquisition. In *Proceedings of the 27th Annual Conf. of the Cognitive Science Society*.

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

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

Chang, N. (2004). A computational model of comprehension-based construction acquisition. Child Lang. Res. Forum.

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

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

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

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

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

MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum.

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

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

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

Pinker, S. (1989). *Learnability and Cognition*. MIT Press.

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

Shayan, S. (2006). How do children learn Agent and Patient roles? Submitted.

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

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