# Generalizing between form and meaning using learned verb classes

**Christopher Parisien** and **Suzanne Stevenson**
Department of Computer Science, University of Toronto
Toronto, ON, Canada
{chris, suzanne}@cs.toronto.edu

**Abstract**

Early language development critically depends on the ability to form abstract representations of linguistic knowledge, and to generalize that knowledge to new situations. In verb knowledge, much generalization appears to be driven by various regularities between form and meaning, but it is difficult to assess how these factors interact in a complex learning environment. We extend a hierarchical Bayesian model to acquire abstract knowledge of verbs from naturalistic child-directed speech, then generalize these abstractions to novel verbs, simulating child behaviour. We use the syntactic alternation structure of a novel verb to infer aspects of its meaning, and use the meaning of a novel verb to predict its range of acceptable syntactic forms. The model provides a useful framework to investigate the interaction of complex factors in verb learning.

**Keywords:** Verb learning; language acquisition; Bayesian modelling; computational modelling.

## Introduction

The productivity of language lies in the ability to generalize linguistic knowledge to new situations. The emergence of generalizations in language development signals important changes in children's representation of linguistic knowledge—in particular, the learning of abstract representations is at the core of these generalizations. To understand this developmental process, we must investigate both how children can find the right abstractions over their input, and how those abstractions can actually guide generalization.

Several lines of research, including recent usage-based approaches (*e.g.,* Langacker, 2000; Goldberg, 2006) as well as earlier perspectives (*e.g.,* Pinker, 1989), suggest that, in many situations, children's ability to generalize is governed by strong regularities between form and meaning. Much discussion in this area has centred on the notion of *alternations* in verb argument structure, in which verbs show different patterns in how they can express their semantic arguments in syntactic forms. For example, the English verb *break* commonly participates in the causative/inchoative alternation:

(1) John$_{\text{Agent}}$ broke the vase$_{\text{Patient}}$. / The vase$_{\text{Patient}}$ broke.

The verb *laugh* also occurs in both transitive and intransitive forms, but with differing semantic roles, such that the intransitive is much more frequent:

(2) Jane$_{\text{Agent}}$ laughed her glee$_{\text{Theme}}$. / Jane$_{\text{Agent}}$ laughed.

Such patterns are not accidental: the pattern with *break* is common with change of state verbs (*freeze, split*), while that of *laugh* is typical of expression verbs (*cry, snort*). Alternation patterns thus capture a connection between the semantics of verbs and their syntactic expression which reflects a possible class structure of verbs (Levin, 1993).

Moreover, such form–meaning regularities have been shown to influence language learning. Two-year-old children can use the alternation structure of a novel verb to predict aspects of its meaning (Naigles, 1996; Scott & Fisher, 2009), and somewhat older children can also use aspects of a verb's meaning to predict its range of acceptable syntactic structures (Ambridge et al., 2011; Kline & Demuth, 2010). This kind of inference in language acquisition appears to involve the interaction of many complex factors, including frequency, verb meaning, and animacy of the arguments. Human and computational experiments have clearly demonstrated the role of statistical regularities over such factors in guiding generalization behaviour (*e.g.,* Merlo & Stevenson, 2001; Scott & Fisher, 2009; Perfors et al., 2010). The next step is a computational model of child language acquisition that models such inferences over verb alternations in the face of noisy, real-world data.

In this work, we use a probabilistic model that has been shown to learn abstract knowledge of verb argument structure and verb classes from naturalistic child-directed speech (Parisien & Stevenson, 2010). We extend the model to capture form–meaning regularities relevant to alternation patterns. We show that the complex probabilistic abstractions acquired by the model are robust enough to capture key behaviours of children and adults in generalizing over verb alternation knowledge. We argue that this kind of probabilistic representation is critical for learning about alternations, since it gives an explicit role for input frequency and allows detailed interactions between frequency and the cooccurrence of various form and meaning features. Moreover, by using verb classes to capture general tendencies over alternations in the data, this representation alleviates the effect of noise and uncertainty inherent in real-world usages of verbs, which show individual variation in their adherence to typical alternation patterns. These properties make this a useful framework for investigating the predictions that arise from the many interacting factors in verb learning.

## Related Bayesian Models

Previous computational approaches have used probabilistic models to capture relationships between form and meaning in children's verb learning. Alishahi and Stevenson (2008) used a Bayesian model to simulate the acquisition of verb argument structure constructions, showing that a probabilistic representation of constructions can explain a variety of generalization behaviours. However, the model does not capture alternation patterns and the generalizations that depend upon them. Another Bayesian model of verb learning acquires classes of verbs based on different alternation patterns, exhibiting appropriate generalizations over those patterns when learning novel verbs (Perfors et al., 2010). However, the model is applied to a limited number of verbs and constructions, and uses idealized semantic knowledge that indicates the verb classes. It remains to be shown that semantic features realistically available in verb usages can appropriately constrain alternation behaviour, across many verbs and classes.

We recently presented a probabilistic model to acquire a broad range of verb argument structures and verb classes from large, naturalistic corpora (Parisien & Stevenson, 2010). The model operated only on syntactic features of the input and did not address semantic generalization. In this work, we extend the model to capture semantic properties of argument structure and show how this influences generalization at the level of verb alternations. We use a representative corpus of child-directed speech to model this acquisition in the context of many constructions, verbs, and alternations.

## Model description

### Representation of Verb Usages

Our representation of individual verb usages comprises both syntactic and semantic information. For the syntactic side, we use the representation from Parisien and Stevenson (2010), which includes 14 features for the number and type of syntactic arguments occurring with a verb. The arguments are recorded individually, under the assumption that children at this developmental stage can identify these various syntactic arguments in the input, without necessarily being able to keep track of full subcategorization frames (a more difficult task).

In this work, we have extended the representation to add a further 15 binary features which capture general semantic information about a verb usage. The first of these features denotes the animacy of the syntactic subject, a method previously used to help distinguish the Agent from other roles in subject position (*e.g.,* Merlo & Stevenson, 2001; Joanis et al., 2008). The next 14 features denote the presence or absence of various coarse-grained semantic properties concerning the event described by the verb. We use general features (not tied to specific verbs or classes) that capture a wide range of verb semantic characteristics, thereby enabling the model to distinguish important aspects of verb semantics discussed in the acquisition literature. While the behaviour of the model is not dependent on any specific set of features, in this work we adopt the following semantic predicates that have been used in the VerbNet verb classification (Kipper-Schuler, 2005): cause, exist, motion, direction, contact, force, has-possession, perceive, experience, expression, disappear, emit, change-state, and result.

The following examples show this representation. (Binary features with a value of 1 are listed, along with the value of non-binary features.)

(3) John broke the window.
⟨ OBJ, NUMSLOTS = 2, SUBJ = animate,
CAUSE, CONTACT, CHANGE-STATE, RESULT ⟩

(4) The window broke.
⟨ NUMSLOTS = 1, SUBJ = inanimate,
CHANGE-STATE, RESULT ⟩

### The hierarchical model of verb knowledge

Our model in Parisien and Stevenson (2010) follows on a large body of research in nonparametric Bayesian topic modelling (*e.g.,* Teh et al., 2006; Wallach, 2008), a robust method of discovering syntactic and semantic structure in very large datasets. In this section, we give an overview of the model as it relates to the interaction between verb alternation classes and verb semantics. (For mathematical details, please refer to Parisien & Stevenson, 2010.)

Adopting a usage-based approach to language (*e.g.,* Langacker, 2000; Goldberg, 2006), we view the acquisition of verb argument structure as a category-learning problem (*cf.* Alishahi & Stevenson, 2008). By grouping together similar items found in the input, the model comes to recognize common underlying structures and to efficiently represent patterns of verb use. The model consists of a hierarchy wherein each level corresponds to a different level of abstraction over such commonalities in verb knowledge. Figure 1 provides an intuitive description of these levels of inference.

At level 1, the lowest level of abstraction in the hierarchy, individual verb usages are represented by sets of syntactic and semantic features as described above. At level 2, the model probabilistically groups similar verb usages into clusters. This set of clusters captures a range of argument structure constructions, where each of these constructions is represented by a set of probability distributions over the syntactic and semantic features in the input. In this way, the model acquires probabilistic associations between form and meaning, a central notion in construction grammar and usage-based language acquisition. We need not specify the total number of constructions to learn; the model itself selects an appropriate set of constructions to represent the input.

In level 3, for each verb in the input, we estimate a distribution over the range of possible argument structure constructions. This gives a general pattern of usage for each verb in the lexicon. For example, in Figure 1, *break* would have a high probability for at least two constructions: the transitive change-of-state construction (*John broke the window*) and an intransitive form (*The window broke*). A key benefit of this kind of representation is that it can distinguish alternative constructions by their degree of entrenchment. While it is possible to use a verb like *laugh* transitively (*Jane laughed her glee*), it is far more likely to be used as an intransitive. The intransitive form of *laugh* should be more entrenched in the lexicon, and should have a greater effect on generalization patterns for *laugh* and other verbs of expression.

Level 4 of the hierarchy allows the model to acquire classes of syntactically and semantically similar verbs. The model groups together verbs with similar patterns of argument structure use—precisely the probability distributions acquired in level 3. Each one of the verb classes in level 4 is represented by another distribution over argument structure constructions,

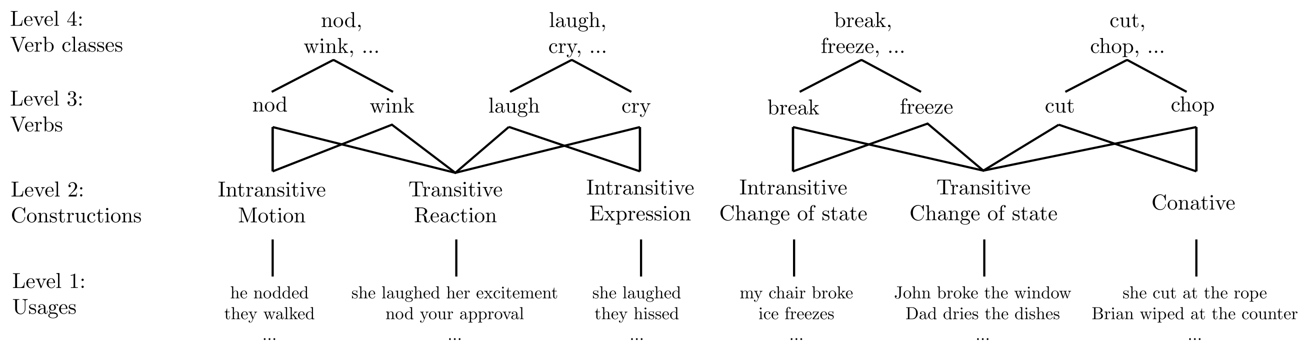


Figure 1: Structure of the model. Each level in the hierarchy corresponds to a distinct level of abstraction in verb knowledge.

but this time accounting for the patterns of *all* of the verbs in the class as a group.

These levels in the model—of abstractions over verb usages—are central to its ability to generalize verb knowledge beyond the data explicitly seen in the input. Each level in the hierarchy provides a more general form of knowledge that can be used to make predictions about the level below it, so that all levels play a role in generalization. In this way, we can predict the usage patterns of a relatively infrequent verb like *rend* using knowledge of similar verbs like *break*, *split*, and *crack*. As we discuss below, these generalizations allow the model to predict syntactic and semantic aspects of novel verbs, capturing important aspects of child behaviour.

## Experimental set-up

We use the Thomas corpus, a longitudinal study of a British English-speaking boy from 2 to 5 years of age (Lieven et al., 2009), part of the CHILDES database (MacWhinney, 2000). Our input includes all child-directed utterances from this corpus that have at least one verb, using every second sentence for development data and the rest for evaluation. The evaluation dataset contains 170,076 verb usages and 1,393 verb types. All reported results are obtained from evaluation data.

The 14 syntactic features for each verb usage are extracted using the parser of Sagae et al. (2007). We manually annotate as animate or inanimate all 4,213 noun phrases that occur as subjects in the input. We estimate the 14 event semantic features for each usage using VerbNet (Kipper-Schuler, 2005): We look up all the argument frames in VerbNet (over all senses/classes of the verb) that are compatible with the syntactic frame of the current usage, and extract all the semantic primitives associated with each such frame. Then, each semantic feature for that usage is marked as True if it is contained in the extracted set. This procedure results in a very noisy representation of the semantics of a verb usage. In particular, because the features for a usage are drawn from all possible senses of the verb in that frame (and not just its intended sense in the usage), the semantics includes features from VerbNet classes that are irrelevant to that usage. Thus, while we use VerbNet to enable us to automatically determine the semantic features, this process does not simply build perfect information about the VerbNet classes of the verb usages into our input. Moreover, the automatic extraction of both the syntactic and semantic features yields a noisy representation that is reasonable given the capabilities of young children in determining such properties.

As described in Parisien and Stevenson (2010), the model is defined by the parameters of a set of probability distributions representing each level of abstraction. To estimate these parameters, we use Gibbs sampling, a Markov Chain Monte Carlo (MCMC) method (Teh et al., 2006). This is an iterative process that results in a large number of samples from the posterior distribution—*i.e.*, the model parameters given the observed data. On development data, the parameters always converge within 3,000 iterations. We perform 10 randomly initialized MCMC simulations on the evaluation data, running each simulation for 5,550 iterations, discarding the first 3,050 as burn-in. We record a sample of the model parameters on every 25th iteration after the burn-in, giving 100 samples per simulation, for 1,000 in total. In the experiments, we average over this set of samples to estimate what the model has learned about the input.

In the simulations, the model acquires approximately 100 argument structure constructions and 90–100 verb classes. Particularly in the smaller classes, low frequency verbs tend to be placed in several different classes over different parameter samples, which is a reflection of the uncertainty in classifying infrequent verbs.

## Experiments

Using its abstract knowledge, the model exhibits two important forms of syntactic and semantic generalization. Firstly, we show how the model can use distributional cues in the alternation structure of a novel verb to infer previously unobserved aspects of its meaning. Secondly, we demonstrate that the model uses the semantic class of a novel verb to appropriately constrain its expected alternation behaviour.

### From alternations to verb meaning

Two-year-old children have been shown to use the alternation structure of a novel verb to infer aspects of the verb's meaning (Naigles, 1996; Scott & Fisher, 2009). For example, in Scott and Fisher (2009), children first heard a dialogue (audio-only) containing a novel verb used with one of two different alternation patterns—i.e., two combinations of transitive and intransitive usages with varying animacy of the ar-

| Experimental condition | Transitive Anim. | Transitive Inanim. | Intransitive Anim. | Intransitive Inanim. |
|---|---|---|---|---|
| Alt-AnimCS | 9 | 3 | 6 | 6 |
| Alt-AnimEX | 9 | 3 | 12 | 0 |
| Intrans-AnimCS | 0 | 0 | 12 | 12 |
| Intrans-AnimEX | 0 | 0 | 24 | 0 |

Table 1: Training conditions for the novel verb in Exp. 1.

guments. They were then shown two videos with two different events and asked to find the event matching the just-heard novel verb. Although the children were not shown a depiction of the novel verb when they heard the dialogue, they were able to map the verb to the semantically-appropriate visual scene based solely on its alternation pattern.

**Experimental Design.** We test our model's ability to generalize in this way from alternation patterns to verb semantics, as follows. We present a novel verb to the model in a particular alternation pattern, but without any event semantics (i.e., the 14 semantic features corresponding to general verb semantics are left blank). We then compare the likelihood of two possible events paired with the verb, one much more compatible with a verb class displaying that alternation, and one much less. The event that is deemed more likely by the model should be the one with the semantic features that match those expected for a verb with the given alternation behaviour. In other words, the model should use the alternation pattern of a novel verb to choose a scene with appropriately matching event semantics.

We use novel verbs comparable to two English verb classes that differ in overall alternation patterns: change of state (*e.g.*, *break*, *freeze*, *dry*) and nonverbal expression (*e.g.*, *laugh*, *giggle*, *cry*). Both types of verbs occur in both transitive and intransitive usages (see Examples 1 and 2), but with differences in two important aspects. First, they differ in the relative frequency of occurrences in these frames. The change of state verbs overall occur equally in each frame, while the expression verbs occur predominantly in the intransitive. Second, because of the differing roles taken by their subjects, they have different patterns of subject animacy (since Agents tend to be animate more than Patients). Change of state verbs have animate subjects about 70% of the time in the transitive and 50% in the intransitive, while expression verbs have animate subjects about 80% of the time in both frames.

We present the model with usages of a novel verb in four different conditions (independently), each having 24 usages in one of four alternation patterns, shown in Table 1. The alternation patterns are a combination of varying proportion of transitive/intransitive usage, and varying proportion of animate subjects, with each variation reflecting idealized usages of the two types of verbs. This allows us to examine a possible interaction between the syntactic frame patterns and subject animacy patterns. The *Alt* conditions correspond to a typical frame pattern of change of state verbs, which alternates freely between a transitive and intransitive usage. The *Intrans* conditions correspond to the frame pattern of a typical expression verb, which is predominantly intransitive. Within these conditions, we manipulate the animacy of subjects in intransitive frames, reflecting idealized proportions of a change of state verb (*AnimCS*) or an expression verb (*AnimEX*). In all conditions, all of the event semantic features of the presented usages are left unspecified, corresponding to the child hearing a dialogue with a particular alternation pattern and with no accompanying depiction of the verb.

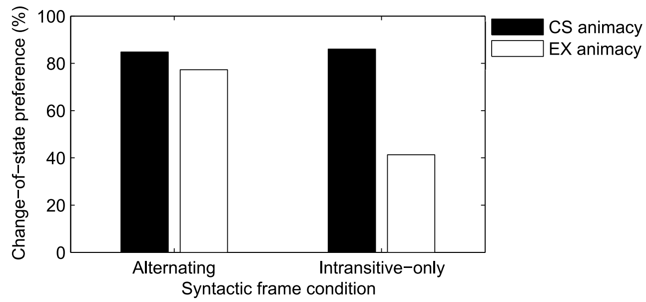


Figure 2: Using the alternation and animacy patterns of a novel verb to infer meaning. The plot shows the percentage preference for the scene with change of state semantics.

For each of these four conditions, we present the model with two test frames. Both are intransitive with an animate subject (consistent with a novel verb of either semantic class), and one has the semantics of a change of state verb, while the other has the semantics of an expression verb, as follows:

(5) ⟨ SUBJ = animate, CHANGE-STATE, RESULT ⟩

(6) ⟨ SUBJ = animate, EXPRESSION ⟩

We then compare the preference in the model for each of these two frames, to see whether the model can infer appropriate semantics from an alternation pattern.

**Experimental Results.** Given the observed usages of the novel verb, $\mathbf{Y}_{nov}$, in one of the four conditions, *Alt-AnimCS*, *Alt-AnimEX*, *Intrans-AnimCS*, *Intrans-AnimEX*, we estimate the likelihood of each test frame $\mathbf{y}_{test}$ using the acquired argument structure constructions $k$ and verb classes $c$:

$$P(\mathbf{y}_{test}|\mathbf{Y}_{nov}) = \sum_k \sum_c P(\mathbf{y}_{test}|k)P(k|c)P(c|\mathbf{Y}_{nov}) \quad (1)$$

This estimate considers how likely the test frame would be if the novel verb happened to be a member of each class $c$, weighted by the probability that the observed pattern $\mathbf{Y}_{nov}$ could have been generated by a verb in that class. We normalize these likelihoods over the two test frames, and average the preference over all 1,000 samples from the simulation.

Figure 2 shows the percentage preference in the model for the test frame with the change of state semantics. When the input alternates between transitive and intransitive frames, there is a strong preference for the change of state scene. There is a small effect of animacy here, such that the animacy pattern of an expression verb reduces this preference slightly. When the input consists entirely of intransitive usages, we see preferences in line with the predictions of the animacy feature: there is a preference for the change of state scene in the *AnimCS* animacy condition, and for the expression scene given *AnimEX* animacy.

There is thus a strong interaction between the alternation cues and the animacy cues. When animacy reflects a change of state verb (the black bars in Figure 2), the alternation has no effect on the preference. When animacy reflects an expression verb (the white bars in Figure 2), the alternation pattern has a strong effect. Verb usage patterns in the corpus provide a strong bias for a change of state interpretation of a novel verb, and the model requires two strong cues (frequent intransitives as well as frequent animate subjects) in order to pull its interpretation in favour of an expression verb. These results show that the model can use the distributional information carried over multiple syntactic frames to help infer the meaning of a novel verb. Moreover, this shows how two distinct features interact to guide generalization behaviour.

Scott and Fisher (2009) discuss possible mechanisms children might use in making this generalization. They consider a *category-mediated* process, similar in principle to our model, as well as a *direct inference* process, by which children directly employ distributional cues to interpret the novel verb, without recourse to a previously learned class. Using the estimated model parameters, we repeat the above experiment using one possible method of direct inference. Rather than measuring the scene preference by comparing the novel verb to each verb class, as in Equation 1, we instead compare the novel verb directly against each of the known verbs from the input. By doing so, in all four training conditions, we observe a 96-98% preference for the change of state scene, with no clear effect of syntactic frame or animacy use. This is a result of drawing inferences over 1,393 verbs, where noise in the data is compounded over such a large number of comparisons. By using verb classes to capture general tendencies in the data, a category-mediated model helps to alleviate the effect of noise, providing better inference in generalization.

## From verb meaning to alternations

The previous experiment considered cases where the alternation structure of a novel verb can help determine the verb's meaning. The reverse can also be true: information about a novel verb's semantic class constrains adults' and children's expectations concerning the syntactic structures that can be used with the verb (Ambridge et al., 2011; Kline & Demuth, 2010). For example, in the Ambridge et al. experiments, subjects were taught a novel verb that was used only in intransitive frames, then asked to rate the verb in a transitive usage. Subjects were more likely to rate the transitive use of the verb as acceptable if its semantics matched a class of verbs which display a transitive/intransitive alternation, than if the class was predominantly intransitive. That is, the semantic class of the novel verb constrains its generalization to a previously unobserved syntactic usage. Here, we show how verb semantics and entrenchment can similarly be used to constrain generalization in our model.

**Experimental Design.** We simulate this experiment by presenting our model with novel verbs comparable in meaning to five different semantic classes, the same classes used by Ambridge et al. Verbs in the first three classes occur freely in the intransitive, but are much less likely to be used in the transitive: disappearance (*e.g., disappear, die*), directed motion (*fall, tumble*), and nonverbal expression (*laugh, giggle*). The other two classes are likely to alternate between transitive and intransitive forms: manner of motion (*roll, spin*) and change of state (*break, split*).

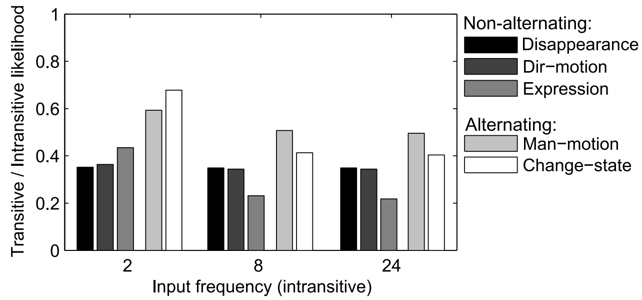


Figure 3: Using the meaning of a novel verb, shown only as an intransitive, to constrain the likelihood of the transitive.

We present the model with a set of either 2, 8, or 24 intransitive frames of a novel verb, coupled with semantic values from one of the following five verb class conditions:

| Non-alternating classes | |
|---|---|
| Disappearance: | ⟨ DISAPPEAR ⟩ |
| Directed motion: | ⟨ MOTION, DIRECTION ⟩ |
| Nonverbal expression: | ⟨ EXPRESSION ⟩ |
| **Alternating classes** | |
| Manner of motion: | ⟨ MOTION ⟩ |
| Change of state: | ⟨ CHANGE-STATE, RESULT ⟩ |

As with the previous experiment, we set the proportion of frames with animate versus inanimate subjects in accord with the proportions for these classes in our corpus data.

**Experimental Results.** Given the training conditions of the novel verb—i.e., a set of intransitive usages with each of 5 possible semantics (as indicated above), at a given frequency level (2, 8, 24)—we use Equation 1 to measure the likelihood of an intransitive and a transitive test frame. (The test frames each have an animate subject, since those are more frequent overall.) Since the test frame likelihoods produced by this method cannot be directly compared with acceptability ratings (as in the human experiments), we instead report the likelihood of the transitive frame relative to that of the intransitive. That is, we divide the transitive likelihood by the intransitive likelihood and report this ratio in Figure 3.

Firstly, Ambridge et al. (2011) observed that when the meaning of the novel verb matched a class of alternating verbs, participants rated transitive uses as more acceptable than if the meaning matched a non-alternating class. In our results, transitive verb usages are more acceptable in the manner-of-motion and change-of-state conditions than in the other three cases. That is, when the novel verb has a meaning similar to a class of alternating verbs, it is more expected to alternate, despite only ever being seen in the intransitive form. The model uses information about the semantic class of the novel verb to appropriately constrain generalization patterns.

Secondly, Ambridge et al. expected to find an effect of input frequency on the generalization, such that the acceptability of a transitive frame would be lower for a high-frequency novel verb than for a low-frequency novel verb. That is, as the intransitive use becomes more entrenched, it would more strongly constrain the use of the transitive. While the authors did not observe such an effect with novel verbs, they did find this effect with known verbs (which they also had subjects rate in various usages). We do see the effect in our results on novel verbs, in some conditions. Specifically, for both of the alternating classes and for one of the non-alternating classes, the likelihood of the transitive decreases as the input frequency increases. The novel verb becomes increasingly entrenched as an intransitive-only verb, even though this may conflict with the semantic cues (*i.e.*, in the case of the novel verbs from the alternating classes). These results show how the model can be used to investigate the interaction of multiple factors in verb learning: semantic cues still have an effect at higher frequencies, but the effect is tempered by the increasing frequency of the observed frame.

## Conclusions

In this paper, we show how abstract knowledge of verb argument structure and verb alternation classes contributes to syntactic and semantic generalization in verb learning. We extend a recent hierarchical Bayesian model of verb learning to capture form–meaning regularities in argument structure, and show that the complex probabilistic abstractions captured by the model are robust enough to drive realistic generalizations of verb alternation knowledge.

Our model is capable of using distributional information carried over multiple syntactic frames to infer aspects of the meaning of a novel verb, a generalization effect observed in children (Naigles, 1996; Scott & Fisher, 2009). We show that by capturing general tendencies of verb use, probabilistic representations of verb classes help to alleviate the effect of noise characteristic of real-world data.

The model is also capable of using the meaning of a novel verb to constrain alternation patterns, an effect discussed by Pinker (1989) and demonstrated recently by Ambridge et al. (2011) and Kline and Demuth (2010). Moreover, we show that as the frequency of the novel verb increases in training, the entrenchment of the observed pattern further constrains generalization, an important factor in usage-based approaches to language acquisition.

To our knowledge, this is the first computational model of verb learning from real-world data that demonstrates the use of acquired class-level knowledge to show both syntactic and semantic generalization effects. The probabilistic nature of the representation is robust to the noise and uncertainty inherent in child-directed speech. This model provides a useful framework to investigate the interaction of multiple factors in verb learning in a complex environment.

**Acknowledgments** We are very grateful to NSERC and to the University of Toronto for financial support.

## References

Alishahi, A., & Stevenson, S. (2008). A probabilistic model of early argument structure acquisition. *Cognitive Science*, *32*(5), 789-834.

Ambridge, B., Pine, J. M., & Rowland, C. F. (2011). Children use verb semantics to retreat from overgeneralization behaviours: A novel verb grammaticality judgment study. *Cognitive Linguistics*, *22*(2), 303–323.

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

Joanis, E., Stevenson, S., & James, D. (2008). A general feature space for automatic verb classification. *Natural Language Engineering*, *14*(3), 337-367.

Kipper-Schuler, K. (2005). *VerbNet: a broad-coverage, comprehensive verb lexicon*. Doctoral dissertation, UPenn.

Kline, M. E., & Demuth, K. (2010). *Syntactic generalization with novel intransitive verbs.* MIT. Manuscript.

Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language.* CSLI.

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

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cog. Ling.*, *20*(3), 481–507.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed., Vol. 2: The Database). Erlbaum.

Merlo, P., & Stevenson, S. (2001). Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, *27*(3), 373–408.

Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, *58*(2), 221-251.

Parisien, C., & Stevenson, S. (2010). Learning verb alternations in a usage-based Bayesian model. In *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society.*

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *J. Child Lang.*, *37*(3), 607–642.

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

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

Scott, R., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and Cognitive Processes*, *24*, 777–803.

Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical dirichlet processes. *Journal of the American Statistical Association*, *101*(476), 1566-1581.

Wallach, H. M. (2008). *Structured topic models for language*. Doctoral dissertation, University of Cambridge.