Learning verb alternations in a usage-based Bayesian model

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

Abstract

One of the key debates in language acquisition involves the degree to which children's early linguistic knowledge employs abstract representations. While usage-based accounts that focus on input-driven learning have gained prominence, it remains an open question how such an approach can explain the evidence for children's apparent use of abstract syntactic generalizations. We develop a novel hierarchical Bayesian model that demonstrates how abstract knowledge can be generalized from usage-based input. We demonstrate the model on the learning of verb alternations, showing that such a usage-based model must allow for the inference of verb class structure, not simply the inference of individual constructions, in order to account for the acquisition of alternations.

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

Introduction

An important debate in language acquisition concerns the nature of children's early syntax. On one side of the debate lies a claim that children develop their syntactic knowledge in an item-based manner. This claim of usage-based learning argues that very young children associate verb argument structure with specific lexical items, only gradually abstracting syntactic knowledge after four years of age (*e.g.*, Tomasello, 2003). An alternative claim suggests that young children do indeed possess abstract syntactic representations—*i.e.*, generalizations about the structure of their language that are not necessarily tied to lexical items (*e.g.*, Fisher, 2002).

Syntactic alternation structure is often considered to be a central phenomenon in this debate. Consider the following example of the English dative alternation:

(1) I gave a toy to my dog.

(2) I gave my dog a toy.

These sentences mean roughly the same thing, but are expressed in different ways. The first, a *prepositional dative*, expresses the theme (*a toy*) as an object and the recipient (*my dog*) in a prepositional phrase. The second, a *double-object dative*, expresses both the theme and recipient as objects and reverses their order.

Verbs that allow similar alternations often have similar semantics (Levin, 1993), which suggests that alternations reflect much of our cognitive representations of verbs. Furthermore, these regularities appear to influence our language use. In word learning experiments, children as young as three years of age appear to use abstract representations of the dative alternation (Conwell & Demuth, 2007). While this is evidence of abstract syntax at a very young age, it does not necessarily invalidate the usage-based hypothesis, since the abstractions may originate from item-specific representations.

One way to bring these opposing positions together is to demonstrate, using naturalistic data, how to connect a usage-based representation of language with abstract syntactic generalizations. We argue that alternation structure can be acquired and generalized from usage patterns in the input, without a priori expectations of which alternations may or may not be acceptable in the language. We support this claim using a hierarchical Bayesian model (HBM) which is capable of making inferences about verb argument structure at multiple levels of abstraction simultaneously. We show that the information relevant to verb alternations can be acquired from observations of how verbs occur with individual arguments in the input. In this sense, we present a *competency* model showing what can be acquired, but we do not make claims regarding the specific processing mechanisms involved.

From a corpus of child-directed speech, our model acquires a wide variety of argument structure constructions over hundreds of verbs. Moreover, by forming classes of verbs with similar usage patterns, the model can generalize knowledge of alternation patterns to novel verbs. This stands in contrast to earlier models which have focused on either the acquisition of the constructions themselves, or the formation of classes over given constructions. The integration in our model of these two important aspects of verb learning has implications for current theories of language acquisition, by showing how abstract syntactic knowledge can be acquired and generalized from usage-level input.

Related work

Previous computational approaches to language acquisition have used HBMs to represent the abstract structure of verb use. Alishahi and Stevenson (2008) used an incremental Bayesian model to cluster individual verb usages (or *tokens*), simulating the acquisition of verb argument structure constructions. Using naturalistic input, the authors showed how a probabilistic representation of constructions can explain children's recovery from overgeneralization errors. In another Bayesian model of verb learning, Perfors et al. (2010) cluster verb *types* by comparing the variability of constructions for each of the verbs. The model can distinguish alternating from non-alternating dative verbs and can make appropriate generalizations when learning novel verbs.

Both of the above models show realistic patterns of generalization, but they operate at complementary levels of abstraction. The model of Alishahi and Stevenson does not capture the alternation patterns of verbs, while Perfors et al. assume that the individual constructions participating in the alternation have already been learned. Furthermore, Perfors et

al. limit their model to only consider two possible constructions (the prepositional and double-object dative), and only the verbs that participate in those constructions.

In this work, we address both levels of abstraction of the above models. We cluster individual verb usages to learn argument structure constuctions and their patterns of use across many verbs, and we also cluster verb types to learn alternation behaviour, generalizing that behaviour to novel verbs. Moreover, we use representative corpora of child-directed speech to model the acquisition of verb alternation behaviour in the context of many constructions, verbs, and alternations.

Vlachos et al. (2009) used a Dirichlet Process mixture model to cluster verb types by their subcategorization preferences, but did not address learning the argument structures themselves. Other work has modelled different aspects of the dative alternation, such as how discourse features affect the expression of dative constructions (de Marneffe et al., submitted), yet did not consider how these preferences are learned.

Model description

We discuss the feature representation of a verb usage and develop two contrasting models to show how alternation classes contribute to generalization in verb learning. Model 1 is an adaptation of an existing probabilistic topic model, the Hierarchical Dirichlet Process (HDP; Teh et al., 2006), to the problem of learning verb argument structure. Model 2, a novel extension to the HDP, addresses the limitations of Model 1 by learning verb alternation classes, allowing regularities in construction use to be transferred to novel verbs.

Verb features

Following from existing approaches (as in Joanis, Stevenson, and James (2008)), we use syntactic "slot" features to encode basic argument information about a verb usage. Table 1 presents the 14 features used in our representation. The first 12 (up through "PP") are binary features denoting the presence or absence of the stated syntactic slot, such as an object (OBJ) or a prepositional phrase (PP); the slots are indicated by labels used by the CHILDES dependency parser (Sagae et al., 2007).[1] When a PP is present, the nominal feature PREP denotes the preposition used. Such syntactic slot features are easier to extract than full subcategorization frames. We make the assumption that children at this developmental stage can distinguish various syntactic arguments in the input, but may not yet recognize recurring patterns such as transitive and double-object constructions. The following examples show this representation used with a double-object dative and a prepositional dative, respectively:

(3) I sent my mother a letter.
⟨ OBJ, OBJ2, PREP = null, NSLOTS = 2 ⟩

(4) I sent a letter to my mother.
⟨ OBJ, PP, PREP = to, NSLOTS = 2 ⟩

[1]We consider only the slots internal to the verb phrase, for now ignoring syntactic subjects. We also do not attempt to distinguish true arguments from adjuncts, a very difficult distinction to make.

Features	Description
OBJ, OBJ2	Objects
COMP, XCOMP	Clausal complements
PRED, CPRED, XPRED	Predicate complements
LOC	Locatives
JCT, CJCT, XJCT	Adjuncts
PP	Prepositional phrases
PREP	Preposition (nominal value)
NSLOTS	Number of slots used

Table 1: Slot features.

Model 1: Argument structure constructions

Like other topic models, the HDP (Teh et al., 2006) is essentially a model of category learning: the model clusters similar items in the input to discover structure. Adopting a usage-based approach to language (*e.g.*, Goldberg, 2006), we view the acquisition of verb argument structure as a category-learning problem. In this view, structured verb knowledge translates well to the hierarchical nature of the model.

Model 1 is a straightforward adaptation of the HDP to verb argument structure, which we will use as a point of comparison for an extended model. Figure 1(a) provides an intuitive description of the hierarchical levels of inference in Model 1. At level 1, the lowest level of abstraction, individual verb usages y_i are represented by sets of features as described above.

At level 2, the model clusters similar usages together to form argument structure constructions, where a construction is represented by a set of multinomial distributions, one for each feature. Since the clustering mechanism is *nonparametric*, we need not specify the total number of constructions to learn. Each of these constructions, denoted by its multinomial parameters θ, probabilistically represents a pattern such as a simple transitive or a prepositional dative. While a construction here encodes only syntactic information, with no semantic elements, the model can be generalized to a combined syntactic/semantic input representation.

At level 3, a multinomial distribution for each verb (π) represents the range of constructions that tend to occur with the verb. For example, in Figure 1(a), *give* (π_2) would have a high probability for the double-object dative and prepositional dative constructions (θ_2 and θ_3, respectively), but a low probability for the transitive construction, θ_1. Let y_{ij} denote feature j of usage i. Levels 1 through 3 are given by the following:

$$\begin{aligned} \pi_v &\sim \text{Dirichlet}(\alpha \cdot \beta) \\ z_i &\sim \text{Multinomial}(\pi_v) \\ \theta_{jz_i} &\sim \text{Dirichlet}(1) \\ y_{ij} &\sim \text{Multinomial}(\theta_{jz_i}) \end{aligned}$$

The indicator variable z_i selects a cluster (*i.e.*, a construction, one of the θ) for usage i. Given a verb v, this is drawn from a multinomial distribution which includes a small probability of creating a new construction.

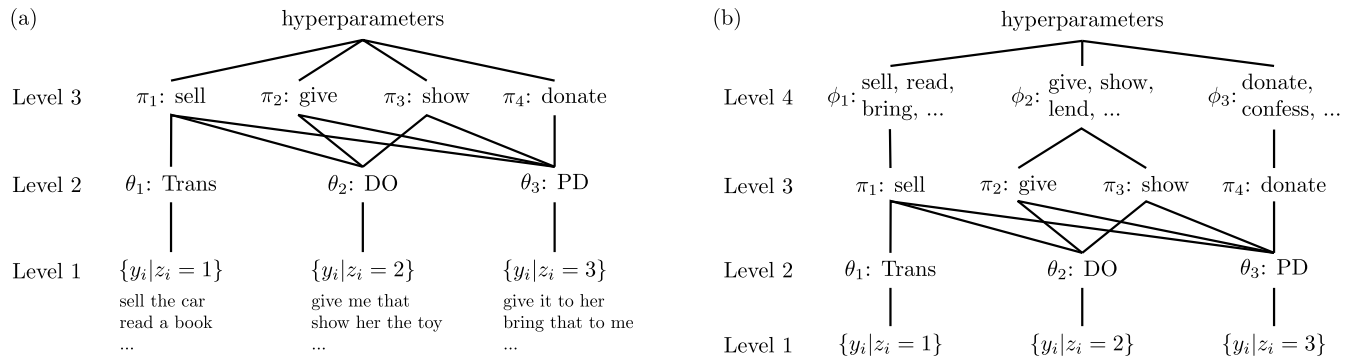

Figure 1: (a) Model 1, a Hierarchical Dirichlet Process applied to learning verb argument structure constructions. (b) Model 2, an extension of Model 1 to learn verb alternation classes.

The verb-specific distributions π_v depend on hyperparameters which encode expectations about constructions in general, across all verbs. They represent acquired knowledge about the likely total number of constructions, which constructions are more likely to occur overall, and so on:

$$\begin{aligned} \gamma &\sim \text{Exponential}(1) \\ \alpha &\sim \text{Exponential}(1) \\ \beta &\sim \text{Stick}(\gamma) \end{aligned}$$

As with lower-level parameters, these are influenced by observed structure in the input. β, drawn from a stick-breaking process (Stick), encodes how many constructions will be used and which constructions are more likely overall. α affects the variability of π_v. Large values of α push π_v closer to β, the global distribution over constructions, while smaller values encourage more variation among verbs. γ affects the total number of constructions; small values of γ correspond to fewer constructions. By drawing α and γ from an exponential distribution, we give a weak preference for verb-specific behaviour and for solutions with fewer constructions. These preferences are effectively designed into the model; they may be informed by general human category-learning behaviour. For further details of this model, see Teh et al. (2006).

Model 2: Alternation classes

Model 1 acquires argument structure constructions from individual verb usages, and learns how those constructions are used by individual verbs, but it is unable to recognize that certain *kinds* of verbs behave differently than others. Competent language speakers regularly use this kind of information. For example, if a verb occurs in a double-object dative construction, then we should infer that it is also likely to occur in a prepositional dative. We develop a novel extension of the above model to capture this phenomenon by learning clusters of similar verbs.

Recall that we represent a verb by a probability distribution over the constructions in which it may occur. In the example shown in Figure 1(a), *give* and *show* both tend to occur with a double-object dative and a prepositional dative, but are less likely to occur as simple transitives. By recognizing the similarity of π_2 and π_3, we can create a cluster containing *give*, *show*, and other similar verbs. Figure 1(b) presents this intuition in Model 2. We extend Model 1 by introducing a fourth level of abstraction, where we represent clusters of similar verbs. For each verb cluster c, we use ϕ_c to represent the range of constructions that tend to occur with any of the verbs in that cluster. By serving as a prior on the verb-level parameters π_v, ϕ_c directly influences each verb in the cluster.

The lower levels of this model are the same as in Model 1. In addition, the verb representations, π_v, depend on the alternation classes in level 4:

$$\begin{aligned} \phi_{c_v} &\sim \text{Dirichlet}(\alpha_0 \cdot \beta_0) \\ \pi_v &\sim \text{Dirichlet}(\alpha_1 \cdot \phi_{c_v}) \\ z_i &\sim \text{Multinomial}(\pi_v) \\ \theta_{jz_i} &\sim \text{Dirichlet}(1) \\ y_{ij} &\sim \text{Multinomial}(\theta_{jz_i}) \end{aligned}$$

Each verb v belongs to a cluster of verbs, denoted c_v. Now, π_v depends on ϕ_{c_v}, which gives a distribution over constructions for all the verbs in the same cluster.

As before, these parameters themselves depend on top-level hyperparameters:

$$\begin{aligned} \gamma_0 &\sim \text{Exponential}(1) \\ \alpha_{0,1} &\sim \text{Exponential}(1) \\ \beta_0 &\sim \text{Stick}(\gamma_0) \end{aligned}$$

These hyperparameters serve similar roles to those in Model 1. β_0 gives a global distribution over all the constructions in use. γ_0 affects the total number of constructions overall. α_1 affects the variability of a verb compared with its class, and α_0 affects the variability of verb classes.

To group verbs into alternation classes, we use a mechanism similar to the way we group individual verb usages into constructions. Recall that c_v acts as an indicator variable, selecting a class for verb v from the available classes in level 4. This is drawn from a multinomial distribution σ which includes a small probability of creating a new verb class:

$$\begin{aligned}\gamma_1 &\sim \text{Exponential}(1)\\ \sigma &\sim \text{Stick}(\gamma_1)\\ c_v &\sim \text{Multinomial}(\sigma)\end{aligned}$$

As with earlier uses of the stick-breaking construction, γ_1 affects the expected total number of verb classes. This method of clustering verb types is similar to Wallach (2008).

Parameter estimation

Models 1 and 2, as written, each specify a prior distribution over the complete set of possible parameters to the models (*i.e.*, all possible values for θ, $\mathbf{z}$, ϕ, and so on). We update these distributions using the observed verb usage data, thus obtaining posterior distributions over parameters.

We estimate the posterior distributions using Gibbs sampling, a Markov Chain Monte Carlo (MCMC) method (Teh et al., 2006). Model parameters are initially set randomly, then iteratively adjusted according to the observed data. We randomly set each z_i to one of 10 initial constructions, and each c_v to one of 10 verb classes (if applicable). We set the remaining parameters to random values drawn from the distributions specified in the model descriptions. We then iteratively update each model parameter *individually* by drawing it from a posterior distribution conditioned on the data and all the *other* parameters in the model. As we iterate through the parameters many times, we collect samples of their values. Over time, this set of samples converges on the posterior distribution—*i.e.*, the model parameters given the observed data. In the experiments, we average over this set of samples to estimate what each model has learned about the input.

Experimental set-up

We use child-directed speech from the Manchester corpus (Theakston et al., 2001), part of the CHILDES database (MacWhinney, 2000). The corpus covers 12 British English-speaking children between the ages of approximately 2 and 3 years. Using CLAN, we extract all child-directed utterances containing at least one verb. We parse the utterances with the MEGRASP dependency parser (Sagae et al., 2007), then reserve every second usage for an evaluation dataset, using the remainder for development. As described above, we extract 14 slot features for each verb usage. The datasets corresponding to each child contain between 4,400 and 10,700 usages and between 239 and 479 verb types. All reported results are obtained using the evaluation data.

Due to flaws in the automatic part-of-speech tagging and parsing, the data contains many errors, particularly in ditransitive constructions. We manually correct the portion of the input related to the dative alternation. For each verb in the development set that occurs with at least one prepositional or double-object dative (as given by the automatic parsing), we draw a sample of up to 50 usages. We repair any cases of incorrectly parsed dative constructions, then duplicate the corrected samples as necessary. Since manual annotation is so labour-intensive, we use this same sample to correct the data for corresponding verbs in the evaluation set. We assume that the proportions of various usages are identical for these verbs across the development and evaluation sets.

We implement both learning models using an adaptation of the NPBayes package (Release 1).[2] For each of the 12 children in the input, we run 10 randomly initialized simulations. The parameters appear to converge within 3,000 iterations, so we run each simulation for 5,800 iterations, discarding the first 3,300 as burn-in. We record a sample of the model parameters on every 25th iteration after the burn-in, giving 100 samples per simulation, 1,000 per child. By averaging over these samples, we can examine the models' behaviour.

Experiments

We compare the ability of our two models to acquire knowledge about the usage patterns of verbs in the input and generalize that knowledge to new verbs. Firstly, we examine construction preferences in two related classes of verbs. Secondly, we test whether the models use an abstract representation of the dative alternation to help learn new verbs.

Verb argument preferences

We examine how our models acquire the usage patterns of verbs in the input by looking at verbs that participate in two different alternation patterns. Earlier, we demonstrated the dative alternation in examples (3) and (4). The benefactive alternation is a related pattern, in which verbs alternate between a double-object form and a *prepositional benefactive* form, as in the following examples:

(5) John made his friend a sandwich.
⟨ OBJ, OBJ2, PREP = null, NSLOTS = 2 ⟩

(6) John made a sandwich for his friend.
⟨ OBJ, PP, PREP = for, NSLOTS = 2 ⟩

We consider all verbs involved in the dative and benefactive alternations, as listed by Levin (1993, Sections 2.1 and 2.2). We test three constructions: the prepositional dative (PD); the double-object construction (DO), whether dative or benefactive; and the prepositional benefactive (PB). Using the samples of the model parameters, we estimate the posterior predictive likelihood of each of these frames for each of the verbs in the given classes. For a given test frame $\mathbf{y}_0$, using verb v, and the observed data $\mathbf{Y}$,

$$\begin{aligned}P(\mathbf{y}_0|\mathbf{Y}) &= \sum_k P(\mathbf{y}_o|k,\mathbf{Y})P(k|v,\mathbf{Y})\\ &= \sum_k \prod_j P(y_{0j}|\theta_{jk})P(k|\pi_v) \qquad (1)\end{aligned}$$

This likelihood is averaged over all 1,000 samples per child.

Figure 2 shows the behaviour of both models. We average the likelihoods over all 12 children, and over all verbs in the following cases: (a) verbs listed as dative but not benefactive, (b) verbs listed as benefactive but not dative, and (c) verbs in both classes. In both models, both dative and benefactive verbs show a high likelihood for the DO frame, and a somewhat higher likelihood for the appropriate prepositional frame

[2] http://www.gatsby.ucl.ac.uk/~ywteh/research/software.html

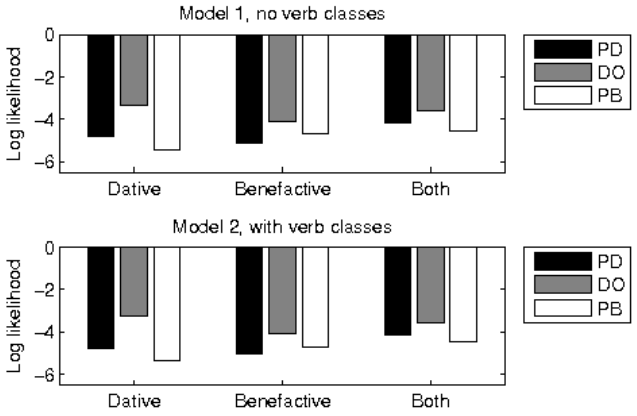

Figure 2: Argument preferences for known dative and benefactive verbs in Models 1 and 2. Shorter bars indicate higher likelihood. The two models show similar behaviour.

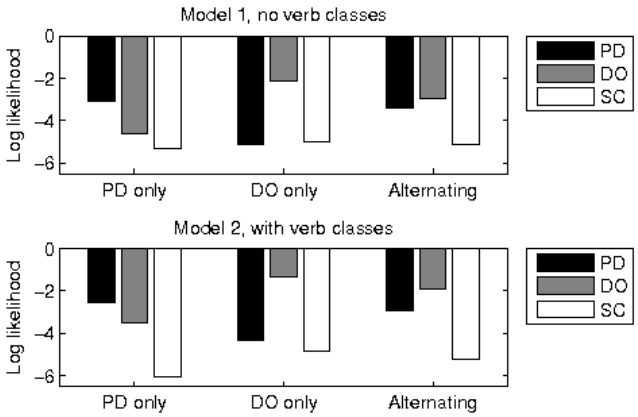

Figure 3: Generalization of novel dative verbs in Models 1 and 2, under various training conditions. Shorter bars indicate higher likelihood.

(PD and PB, respectively) than for the inappropriate one (PB and PD, respectively). Verbs that occur in both classes show closer likelihoods for all three frames.

These results suggest that both models can acquire the argument structure preferences of verbs in the input. In this case, the ability of Model 2 to acquire verb alternation classes is not necessary. Both models are able to cluster verb usages into a range of constructions and acquire appropriate usage patterns over a range of verbs. Both models acquire approximately 20 different constructions. Model 2 acquires 35-40 verb classes, depending on the child.

Novel verb generalization

Children as young as three years of age have been shown to use abstract representations of the dative alternation (Conwell & Demuth, 2007). When young children hear a sentence like *I gorped Charlie the duck*, they appear to know that the same meaning can be expressed by saying *I gorped the duck to Charlie*. We test this generalization in our models by presenting a novel verb in one form of the dative and measuring the likelihood of the alternating form.

We test each model by independently presenting it with a novel verb in three different situations: (a) two instances of the prepositional dative, (b) two instances of the double-object dative, or (c) one instance of each. Only in case (c) is the verb explicitly seen to be alternating. We test the ability to generalize alternation behaviour by comparing the likelihood of the unseen *alternating* form with an unseen form unrelated to the alternation. The non-alternating frame is the sentential complement (SC) frame, which occurs in 1-1.5% of the input, approximately the same overall frequency as either of the two dative frames. For example, if we train the novel verb using only the PD, yet the DO frame shows a higher likelihood than the unrelated SC frame, then we can say that the model has generalized the dative alternation.

Since the novel verbs are *not* in the observed data, we must further iterate the Gibbs sampler, using the new data, to obtain the appropriate samples of the verb-level distribution π_v. For each of the 1,000 parameter samples per child we obtained from the original simulations, we re-initialize the model with the parameters from the sample, add in the novel data for case (a), (b), or (c), then do a further 350 iterations, recording 10 new samples of the model parameters. This gives 10,000 new samples per test case, per child. Using equation (1) and the new samples, we estimate the posterior predictive likelihood of each of the three constructions. This gives an estimate of the relative preferences for a verb's usage and is a direct measure of the acquired lexicon. Translating this estimate to production, as seen by Conwell and Demuth (2007), would require a model of how discourse and other factors influence dative production (*e.g.*, de Marneffe et al., submitted). This is beyond the scope of this paper.

Figure 3 shows how the ability to acquire verb classes aids generalization. In Model 1, without verb classes, only the frames already seen with the novel verb are highly likely. This means that Model 1 is unable to generalize beyond observed data. In contrast, Model 2 shows appropriate generalization for the dative alternation. When the novel verb is trained with the prepositional dative, the double-object dative shows a much higher likelihood than the unrelated SC frame. A similar effect occurs with DO-only training: the PD frame is now more likely than the SC frame, although only slightly. Compared with Model 1, *both* dative frames obtain a higher likelihood across all three training cases, while the SC likelihood remains low. The ability to acquire alternation classes improves the ability to learn *both* alternating constructions.

One aspect of our results differs from the behaviour observed in children. Our verb-clustering model is more likely to generalize to the double-object form when trained only on a prepositional form, than the other way around (i.e., generalizing from a DO to a PD). However, three-year-old children seem to be biased to the prepositional form, the opposite effect (Conwell & Demuth, 2007). We suggest that this is a result of our small corpora. High-frequency dative verbs

tend to be biased toward the double-object form (Campbell & Tomasello, 2001). However, Gries and Stefanowitsch (2004) show that out of 40 alternating verbs in the larger ICE-GB corpus, 19 are prepositional-biased. This strongly suggests that more low-frequency verbs are prepositional-biased than otherwise. A small corpus will likely over-represent a double-object bias because of undersampling of low-frequency verbs. By applying Model 2 to larger corpora of child-directed speech in future work, we hope to correct this issue.

Conclusions

In this paper, we show how verb alternation classes contribute to generalization in verb learning. We develop a hierarchical Bayesian model, Model 2, that is capable of acquiring knowledge of verb argument structure at multiple levels of inference simultaneously. We demonstrate this using the wide range of verbs and constructions contained in a corpus of naturalistic child-directed speech.

By clustering individual verb usages, both of our models acquire a variety of argument structure constructions and learn their patterns of use over hundreds of verbs. Furthermore, Model 2 learns groups of verbs that occur with similar usage patterns. Using the dative alternation as a key example, we demonstrate how this knowledge of alternation classes can be generalized to novel verbs, as observed in the behaviour of children and adults. This verb class model can acquire and apply this knowledge without any prior expectation of which constructions and alternations may or may not be relevant.

In contrast to previous analyses of the dative alternation (Perfors et al., 2010; de Marneffe et al., submitted), we demonstrate its acquisition in the context of many other constructions, verbs, and alternations. Despite the low frequency of the participating constructions, our model successfully acquires the dative alternation. This is a strong endorsement of hierarchical Bayesian models of language acquisition.

This approach offers a potential bridge between differing theoretical positions in language acquisition. By simultaneously learning at multiple levels of abstraction, our model connects a usage-based representation of language, as proposed by Tomasello (2003), with weak abstract representations similar to those championed by Fisher (2002). Other usage-based Bayesian models, such as that of Alishahi and Stevenson (2008), offer a similar opportunity, although our model develops higher-level abstractions regarding the structured knowledge of verbs.

One of the key features of usage-based constructions is that they couple form to meaning (Goldberg, 2006). Moreover, Fisher argues that abstract syntactic representations influence semantics in verb learning, and vice-versa. By augmenting our model's input with semantic properties, we will examine the interaction of syntax and semantics in verb alternations. We will investigate how an argument alternation may convey semantic information, as in Scott & Fisher's (2009) demonstration of 28-month-old children inferring causation in transitivity-alternating verbs.

Acknowledgments

We thank Yee Whye Teh for valuable discussions, and NSERC and 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.

Campbell, A. L., & Tomasello, M. (2001). The acquisition of English dative constructions. *Applied Psycholinguistics*, *22*(02), 253–267.

Conwell, E., & Demuth, K. (2007). Early syntactic productivity: Evidence from dative shift. *Cognition*, *103*(2), 163–179.

de Marneffe, M. C., Grimm, S., Arnon, I., Kirby, S., & Bresnan, J. (submitted). A statistical model of the grammatical choices in child production of dative sentences.

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: a reply to Tomasello (2000). *Cognition*, *82*(3), 259-278.

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

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on alternations. *Intl. J. Corpus Linguistics*, *9*, 97–129.

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

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

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

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

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2007). High-accuracy annotation and parsing of CHILDES transcripts. In *Proc. 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.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *J. Child Language*, *28*, 127-152.

Tomasello, M. (2003). *Constructing a language: A Usage-Based theory of language acquisition*. Harvard U. Press.

Vlachos, A., Korhonen, A., & Ghahramani, Z. (2009). Unsupervised and constrained Dirichlet process mixture models for verb clustering. In *Proceedings of the EACL Workshop on Geometrical Models of Natural Language Semantics.*

Wallach, H. M. (2008). *Structured topic models for language*. Unpublished doctoral dissertation, Univ. of Cambridge.