# Modeling Verb Lexicalization Biases using Hierarchical Bayesian Models

**Catherine Havasi (havasi@media.mit.edu)**
MIT Media Lab, 20 Ames Street
Cambridge, MA 02139 USA

**Robert Speer (rspeer@mit.edu)**
MIT Media Lab, 20 Ames Street
Cambridge, MA 02139 USA

## Abstract

The expression of motion verbs differs between languages. The path of motion, such as crossing or entering, is more prominently featured in path-based languages such as Spanish than in manner-based languages such as English. Here, we revisit the data from a study on manner and path biases in verb lexicalization (Havasi & Snedeker, 2004), and create a hierarchical Baysian computational model to further explore, verify, and define these biases. With this model, we can discover the large differences in subjects' pre-existing manner and path biases that depend on the syntactic frame in which new verbs appear, as well as a difference in the learning rate between English speakers taking the experiment in English and bilingual Spanish speakers taking the experiment in Spanish. We can also use the model to predict the responses of subjects in the experiment with more accuracy than before.

**Keywords:** verb learning; bayesian modeling; hierarchical Bayes modeling; manner and path verbs

## Linguistic lexicalization biases

People have the ability to intuit the meaning of a new verb after hearing it used to describe just a single event. In the case of a novel verb, there are many potential hypotheses of the verb's meaning which may be consistent with the event witnessed. Suppose you hear a novel verb, such as "gorp", being used to describe an event in which Jesse throws a frisbee across a field to , her dog. The verb could refer to Jesse throwing the frisbee, the frisbee's motion as it glides across the field, the frisbee's traverse of the field, or Edison's act of catching the frisbee. To understand which aspect of the action the verb refers to, you must use situational clues and background knowledge.

When one encounters a new object noun, one encounters the same ambiguity in meaning. In practice, languages systematically favor a few different characteristics such as common ancestry or base level category (Nelson, 1973) for noun meanings which is often indicated by shape. However, event categorization tends to be flexible across languages and even with a language (Talmy, 1975). A motion verb, for example, could easily refer to the manner, cause, or path of the motion with no universal preference across languages (Aske, 1989; Berman & Slobin, 1994; Jackendoff, 1990).

Given the plethora of possible referents for a novel verb, how do children learn verb meanings? One solution would be to observe, over several examples, that certain semantic features seem to always be present and are thus associated with the verb's meaning. However, this would require too much data to match the way that children learn words; children can often determine the relevant aspect of a word's meaning from a single example (Gentner & Boroditsky, 2001), and they can even learn words for events they are unable to observe (Landau & Gleitman, 1985).

Two faster and more noise-resistant strategies have been hypothesized by researchers. One is syntactic bootstrapping (Gleitman, 1990). In this theory, the syntactic frame of the verb is used to constrain hypotheses to those which makes sense in the given frame and are similar to known verbs with similar frames. In the manner/path example given earlier, you would be more likely to think the meaning of the novel verb was related to its motion if you had heard the semantically rich fame "Jesse gorped the frisbee to Edison."

Another hypothesis is that we are able to quickly learn words from few examples because we rely on our learned lexicalization biases about the meanings of words (Gentner & Boroditsky, 2001). Learners select word meanings that align with the features that are dominant in the learner's native language (Naigles, 1990), indicating that language learners observe general features of the meanings of other words and apply them to new words as well.

Modern evidence suggests that children use a combination of these strategies (Papafragou & Selimis, 2010). But how are these biases learned and regulated? In this paper, we explore the possibility that biases for certain components of meaning are associated with language and semantic frame. These biases represent examples of Bayesian overhypotheses about what a word is likely to mean, and these overhypotheses can themselves be learned from examples (Kemp, Perfors, & Tenenbaum, 2007). The overhypotheses can depend on observable features such as whether the referent is animate (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002), the syntactic patterns in which the word appears (Cifuentes-Férez & Gentner, 2006), or known lexical relations to other words (Pustejovsky, 1998).

Return momentarily to the analogous results for shape biases in nouns — that early nouns that children learn tend to be easily clustered by the shape of their referents. In order to model this bias, it was postulated that children learn a "second-order generalization" that objects are often categorized by their shape (Samuelson & Smith, 1999). Smith et al. demonstrated this generalization by teaching 17-19 month old children a precocious shape bias (Smith et al., 2002). Kemp, Perfors, and Tenenbaum explained this kind of learning using a hierarchical Bayesian model, which could learn both base meanings and overhypotheses simultaneously (Kemp et al., 2007) and cases have been made that this approach applies to many aspects of word learning (Xu, Dewar, & Perfors, 2009).

To further understand how biases shape learning, Havasi and Snedeker (Havasi & Snedeker, 2004) began to explore the lexicalization bias for motion verbs. The experiments focused on the plasticity of the English manner motion verb bias and its dependence on syntactic frame. As noted by Talmy, the relative rarity of path verbs in English leads most English speakers to think of these verbs as secondary (Talmy, 1985). Like most linguistic biases, we are clearly not born with such a bias. Through their experiments, they sought to discover if the manner bias could be reversed to a path bias, and if so, to explore the possibility that these biases are influenced by the set of verbs that a subject is exposed to. Their results pointed to a bias which was surprisingly plastic and adaptable with training, even in adult subjects.

Prior models of verb learning were Bayesian in nature, but encountered effects which could not be explained with a single-level model. Some of these effects can be accounted for by a two level model: children are learning verbs and the behavior of verbs in the language simultaneously. From here on, we will refer to knowledge about the meanings of particular verbs as "level-1 knowledge", and knowledge about verbs in the language as a whole as "level-2 knowledge".

In this paper, we design and test a hierarchical model for which matches the pattern of performance in the Havasi and Snedeker verb learning studies, with a particular focus on modeling level-2 knowledge. We are interested in building a deeper understanding of that work in the context of work, mentioned above, which models similar problems in noun learning.

## Modeling verb biases

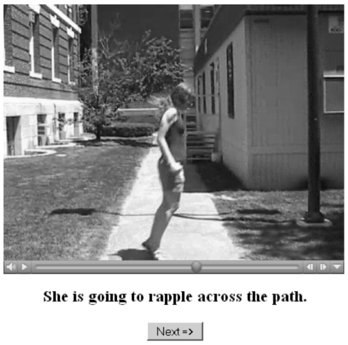
The MIT/Harvard studies on manner/path bias were constructed as a corollary to the shape bias work, showing that the learned lexical constraint biases extended beyond noun learning and remained adaptable into adulthood. During these experiments, we were trying to teach the subjects new words and adjust their manner-path biases. This adjustment would determine how plastic these biases were and in what ways they could and would change.

### Experimental setup

In the course of each experiment, adult subjects learned twelve novel motion verbs. For each verb:

1. Subjects saw a single ambiguous scene with a prominent path and manner of motion, and a sentence describing the scene using a novel verb. An example screenshot of this step can be seen in Figure .

2. Subjects were asked two questions to determine their initial interpretation of the verb (their bias). Each question asked if a new scene was an example of the verb; one of the scenes was consistent with the initial example's manner, while the other scene was consistent with the initial example's path. This step is called the "initial test".



She is going to rapple across the path.

Figure 1: A screenshot from the original Havasi-Snedeker experiment presenting the subject with a novel verb-event pair.

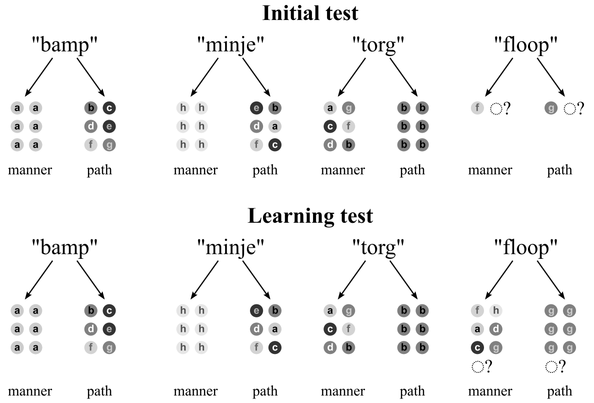


Figure 2: The information that a subject in the manner-path experiment encounters. "a" through "h" represent the different values each aspect may take.

3. Subjects saw five additional instances of the new verb which clarified the meaning, which would be consistent with the initial scene in either the manner or the path.

4. Subjects answered two more questions to ensure that they had learned the novel verb. They are again shown two scenes, one consistent with the initial ambiguous scene in manner and one in path. This is the learning test step.

A graphical depiction of the steps of the experiment can be seen in Figure 2.

When discussing how a verb's meaning is inferred from examples, we will say that an aspect of a verb is *consistent* if it is the same in all examples. In Figure 2, the consistent aspect is the one with six examples that are the same color.

It is possible for a verb to be consistent in neither its manner nor its path – perhaps it is defined by some other aspect. We thus classify the possible beliefs about a verb's meaning as "manner", "path", and "neither".

An important part of this experiment was to vary the proportion of path and manner verbs across groups of subjects. Some subjects learned only manner verbs, some learned only path verbs, and others saw different proportions of both types

( 0%, 25%, 50%, 75% and 100%). The subjects' answers to the first initial test – after only one scene had been shown, and therefore before any level-2 learning could take place – were taken to represent their initial manner-path bias.

During the first experiment, a "rich frame" was used to present the novel verbs. In these experiments, the sentences that were used contained a preposition as well as a ground element, as in "she is glipping around the tree" where "tree" serves as the ground. In English, this type of frame is more frequently used with manner verbs.

The impact of syntactic frame has on the initial assumption that a verb is a manner or path verb based on a single scene and utterance pair is also a matter of interest. In the second experiment, we used a different "poor" syntactic frame, lacking a preposition, to express the novel verb. The ground element appeared instead as the direct object, as in "she is glipping the tree". This poor frame is more commonly used with path verbs in English than the rich frame. Results for both experiments were originally published in (Havasi & Snedeker, 2004).

Another version of the experiment was run on native Spanish speakers (who generally spoke English as well), with the scenes described by Spanish sentences that were similar to the English rich frame. Another experiment evaluated the manner-path bias in children, but its results are not directly comparable, because it required a different experimental setup (Havasi, Snedeker, & Malik, 2005).

### Early work modeling manner and path biases

In her master's thesis, Catherine Havasi (Havasi, 2004) used the results of these experiments in the development of a computer model to explain the human responses during the experiments. This model, which progressed through the training and evaluation steps in the same manner as the subjects, used a multinomial distribution to represent the likelihood of seeing a manner or path verb next. This distribution was updated at each time step based on its previous value and the most recently seen example, and its prior distribution was trained to provide the closest match between the model and human responses.

However, this initial model did not seem to adequately fit the observed behavior, because subjects' manner and path biases would change significantly during the course of the experiment. The subjects were acquiring level-2 knowledge about the language of the experiment.

To account for this, she added a "memory effect", which would incrementally change the prior of the manner-path distribution based on an average of the last several observations. Adding the memory effect improved the ability of the model to match the experimental data.

### Flexible bias as an overhypothesis

The understanding of hierarchical Bayesian models and their application to cognitive science has grown significantly since these results. We now know a mathematical way to describe this "memory effect": it is an overhypothesis. This overhypothesis itself has a prior, representing the subjects' pre-existing bias toward manner or path verbs. What the early model accomplished with a changing prior, we can now do in a more principled way. We replace the prior with an overhypothesis that is informed by the level-2 data, which itself has a prior that is fixed for the course of the experiment.

This follows a program of research that has been successful in describing many aspects of language learning. One benefit of hierarchical Bayesian models is that learning can take place on multiple levels at once; (Perfors & Tenenbaum, 2009) models how people learn categories at the same time that they "learn to learn" categories. Another benefit is that it accounts for how people learn in the absence of counterexamples, as in the syntax-learning model of (Perfors, Tenenbaum, & Wonnacott, 2010). In this experiment, too, people show that they gain knowledge of verb meanings at multiple levels, and can correctly answer that a stimulus is "not gorping" even when they have seen only positive examples of "gorping".

## The hierarchical model of verb aspect learning

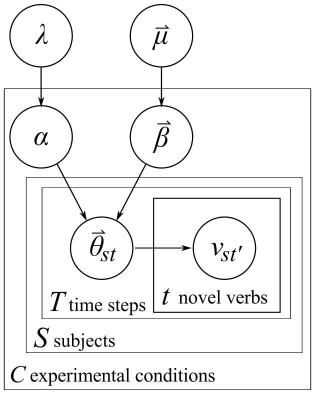
Building on Kemp, Perfors and Tenenbaum's result (Kemp et al., 2007) in which they modeled the shape bias with a hierarchical Bayesian model, we have created a higher-level hierarchical Bayesian model that describes how these biases change among different subjects in multiple experimental conditions. We are looking for differences in the way people learn, and particularly for differences in subjects' initial biases, when subjects are presented with text in different languages or different syntax.

We start by assuming (based on the cited previous work) that a hierarchical Bayesian model adequately describes how a subject acquires level-2 knowledge, and that the "initial test" in our experiment reveals information about each subject's level-2 knowledge at each step of the experiment. When we make use of this assumption and observe changes in subjects' level-2 knowledge over the course of the experiment, we can discover information that was not directly revealed by the questions in the experiment.

For example, subjects' answers to the initial test for the first verb tell us something about their relative manner and path biases at the start of the experiment. But to infer how *strong* those biases are, or inversely how prone they are to change, we need to observe how these biases change during the experiment across many subjects. In other words, we are using this new hierarchical model not to model how a single person learns, but to design an experiment that helps us learn how people learn.

The different experimental conditions will cause subjects to reveal different biases – not because the subject knows they are in one of a set of experimental conditions, but because the conditions naturally draw on different sets of prior knowledge. We describe these differences in biases with an overhypothesis that varies according to the experimental condition, representing subjects' initial level-2 knowledge before the experiment teaches them any new verb meanings.

We do not extend this model to include level-1 learning,



$$\begin{aligned}
\alpha &\sim \text{Exponential}(\lambda) \\
\vec{\beta} &\sim \text{Dirichlet}(\mu) \\
\vec{\theta}_{st} &\sim \text{Dirichlet}(\alpha\vec{\beta}), s \in [1,S], t \in [1,T] \\
v_{t'} &\sim \text{Multinomial}(\vec{\theta}_{st}), t' \in [1,t-1]
\end{aligned}$$

Figure 3: A hierarchical model of the manner-path experiment.

because we have already isolated level-2 learning from level-1 learning in the experimental design. As we have no reason to believe that people will learn differently at level 1 in different conditions, starting from the level-2 data gives us a simpler model with less room for unnecessary variation.

Figure 3 shows this model, using the traditional "plate notation" to show variables that are sampled many times. To describe the role of the different variables, let us begin at the innermost plate.

$v_{st'}$ represents a subject's knowledge or hypothesis about which aspect defines a particular verb. It can take one of three values in our model: *manner*, *path*, and *neither*. This is the knowledge that is probed by the initial test, taught by the five training examples, and confirmed by the learning test. $t'$ is enumerated from 1 to $t$ because the subjects only have information about the verbs they have seen so far; their set of knowledge grows as the experiment proceeds.

Note that $v$ contains two different kinds of information. For all $t' < t$, $v_{st'}$ represents what subject $s$ has learned about a previous verb. The case where $t' = t$ is different: it represents the subject's prediction of what the current verb means, after they have seen only one example. This comes from the subject's response to the initial test.

We assume here that subjects learn the meaning of the previous verbs correctly, given that they have seen six examples of each. We cannot be sure of what the subjects actually learned, but on average the subjects answer the learning test question correctly 89% of the time. A model of level-1 learning instead of level-2 learning may be able to predict when some of these errors occur.

The values of $v$ are selected from the multinomial distribution $\vec{\theta}$, representing the subject's current beliefs about which semantic aspects typically define verbs in their language. This is a vector that is specific to the subject and changes over time. If someone's $\vec{\theta}$ is [0.6, 0.3, 0.1], for example, that means they believe there is a 60% chance that a verb they have seen one example of will be defined by its manner, a 30% chance it will be defined by its path, and a 10% chance that it will be defined by neither.

Determining where the values in $\vec{\theta}$ come from, and how they tend to change over time, is the goal of the experiment. We suppose that there are two hyperparameters, $\alpha$ and $\vec{\beta}$, which are similar among a population of people who speak the same language. These hyperparameters represent whether a population of verb learners expects, in general, to learn manner verbs or path verbs.

$\vec{\beta}$ represents a person's initial bias toward manner or path verbs, as a vector of probabilities. $\alpha$ represents the strength of this bias: a low $\alpha$ can be easily overridden by evidence in the experiment, while a high $\alpha$ represents a bias that is hard to change.

These hyperparameters can still vary according to the language the person is using, and according to other information such as the syntactic frame that the novel verb appears in. There may be hyper-hyperparameters that determine the way that people learn verbs in general, but we do not have nearly enough data to study what they are. We call these parameters $\lambda$ and $\vec{\mu}$, and assume neutral and uninformative priors on them: $\lambda$ yields $\alpha$ from an exponential distribution with mean 1, and $\vec{\mu}$ yields $\beta$ from the flat Dirichlet distribution with parameter [1,1,1].

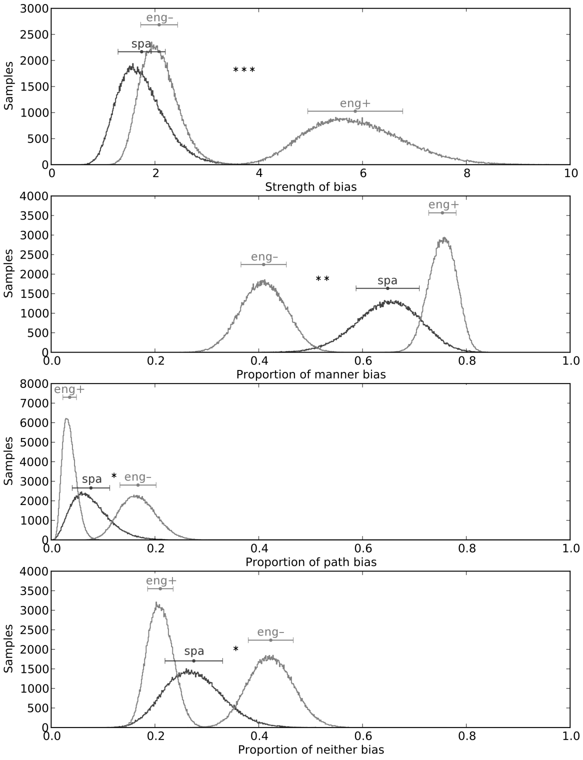
Given the data collected in the experiment, we can discover likely values of $\alpha$ and $\vec{\beta}$ for each condition, and therefore learn what people's manner and path biases are and how they depend on the language and its syntax. Then, we can use this hierarchical model, initialized with the given values, to model people's performance in the experiment.

## Sampling to find $\alpha$ and $\vec{\beta}$

The goal of this experiment has been to discover subjects' biases in verb learning. We have taken into account the fact that the biases can change over the course of the experiment as subjects adapt to what they are being taught. But the result of the change is not what we are interested in – that mostly reflects how well the subjects ultimately learn about the class of novel verbs in the experiment, verbs that they will never need to use in their life. What we are interested in is what the experiment reveals about what the subjects' biases were *before* the experiment, at time 0.

These pre-existing biases are represented by $\vec{\beta}$, and now that we have constructed a Bayesian model of the experiment, we can sample the model to find their likely values. The rigidity of these biases during the experiment is represented by $\alpha$. When a subject's biases change significantly during the experiment, this is reflected by a low value of $\alpha$.

To find a distribution on $\alpha$ and $\vec{\beta}$ given the hyperparameters and the experimental data, we use a implement the Markov Chain Monte Carlo process using the Metropolis-Hastings



**eng+** = English rich frame; **eng–** = English poor frame;
**spa** = Spanish

Figure 4: The distributions of model parameters that are likely given the data. From top to bottom, the parameters indicated are $\alpha$ (the strength of the bias), $\beta_1$ (the manner bias), $\beta_2$ (the path bias), and $\beta_3$ (the "neither" bias).

algorithm. This tells us about the distribution of parameter values that explain the data, by giving us a number of samples from that distribution.[1]

We ran this sampling process separately on the rich frame and poor frame conditions of the English experiment, as well as the data from the Spanish experiment (which only had one verb frame, with similar content to the English rich frame), for 200,000 steps in each case. Figure 4 shows the distributions of the initial manner bias, initial path bias, and the weight of the bias ($\alpha$) for each condition.

## Evaluating the predictive accuracy

We evaluate our model by comparing it to the way people actually predict the meanings of new verbs in the experiment. When we run the model forwards using the mean of the parameters we found earlier, then for each time step in each condition, we get a prediction of the probability that a person will predict a path verb, a manner verb, or neither. We compare this to the proportion of subjects who actually predicted each option, and calculate the error as the average Euclidean distance between these vectors of probabilities.

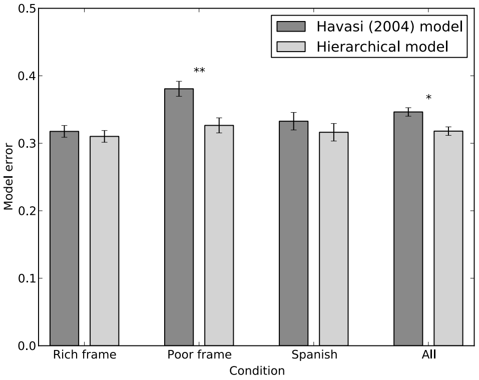


Figure 5: The prediction error for the model of (Havasi, 2004) and the hierarchical model presented here.

We can evaluate the model of (Havasi, 2004) with the same methodology, and therefore show that this model is an improvement over the previous model. We run both models on the same data and compare them with a paired $t$-test, showing that the new hierarchical model is more accurate when averaged over all data ($t = 2.25$, $df = 287$, $p < .05$), and particularly on the poor-frame data ($t = 2.43$, $df = 120$, $p < .01$), where people began with very uncertain biases that changed quickly. These results are plotted in Figure 5.

## Analysis of results

The results in figure 4 show very different initial biases for the different experimental conditions. The clearest difference is, in fact, not due to language but due to the syntactic frame used in the experiment. In the rich-frame English experiment, the typical subject has a prior manner bias of approximately 0.77 (they consider an unknown verb to have a 77% probability of being a manner verb), and a very small path bias, around 0.02. Compare this to the poor syntactic frame, in which subjects reveal a manner bias around 0.41 and a path bias around 0.16 (leaving 43% of the probability for neither manner nor path verbs).

There is also a striking difference in the strength of the biases, as indicated by $\alpha$. The strong manner bias in the English rich frame is also difficult to overcome, having an $\alpha$ value that averages around 5.9. The poor frame induces a much weaker bias, with $\alpha \approx 2.0$.

We can evaluate the significance of these differences in values by sampling from the distributions and establishing whether one value is larger 95% (or 99%, or 99.9%) of the time. (The standard error of the mean does not apply, because these samples are not independent, and they come from distributions that *already* reflect our uncertainty about the parameter values.) For $\alpha$ and all entries in $\vec{\beta}$, the difference between the English rich frame and poor frame is significant at the $p < .001$ level. We conclude that the syntax in which a verb originally appears has a strong effect on a learner's hypothesis about its meaning.

The pre-existing biases for Spanish actually appear to fall

[1] The Python script that implements this process, along with the experimental data, are available at `http://github.com/rspeer/verb-aspect-learning`.

between the two English cases. Spanish-speaking subjects may have slightly less of a manner bias and slightly more of a path bias than the English rich-frame subjects, but to an extent that is not statistically significant over our data. On the other hand, they have a significantly *stronger* manner bias than the English poor-frame subjects. This difference in manner biases is significant at the $p < .01$ level, and the differences in other biases are significant at the $p < .05$ level.

The very significant difference between the Spanish data and the English rich-frame data occurs in $\alpha$, the strength of the subjects' pre-existing biases. With an average $\alpha$ around 1.9, the Spanish-speaking subjects adapt to the distribution of meanings in the experiment as quickly as the English poor-frame subjects. Their bias is weaker than the English rich frame at a significance level of $p < .001$.

An explanation we propose for the Spanish data is that we are observing, in addition to the language difference, the fact that all the Spanish native speakers were bilingual. They have learned English, along with its lexicalization biases, so they are apt to learn unfamiliar words in the same way that they learn English words. On the other hand, their bilingualism has given them practice at adapting their overhypotheses about the meanings of words, so they adapt to the "language" of the experiment more quickly.

Examining the Spanish result further, and determining whether their initial biases differ from the English rich frame, would require more experimental data. To further explore this phenomenon, it would be quite useful to re-run the Spanish experiment with an equivalent to the poor frame. Additionally, it would be informative to run the experiment with monolingual Spanish speakers, in order to isolate the possible effect of bilingualism on adaptiveness in verb learning.

## Acknowledgements

We wish to thank Jesse Snedeker and Josh Tenenbaum for their helpful discussions and feedback about this model.

## References

Aske, J. (1989). Path predicates in English and Spanish: A closer look. In *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society* (p. 1-14). Berkeley, CA, USA: Berkeley Linguistics Society.

Berman, R., & Slobin, D. (1994). *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cifuentes-Férez, P., & Gentner, D. (2006). Naming motion events in Spanish and English. *Cognitive Linguistics*, *17*, 443–462.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & Levinson (Eds.), *Language acquisition and conceptual development.* Cambridge, England: Cambridge University Press.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Havasi, C. (2004). *Bayesian modeling of manner and path psychological data*. Masters thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge.

Havasi, C., & Snedeker, J. (2004). The adaptability of language specific verb lexicalization biases. In *Proceedings from the Annual Meeting of the Cognitive Science Society* (Vol. 26). Mahwah,NJ: Erlbaum.

Havasi, C., Snedeker, J., & Malik, M. (2005). The adaptability of lexicalization biases in English speaking five year olds. In *Proceedings of the Tenth Annual International Congress for the Study of Child Language.* Berlin, Germany.

Jackendoff, R. S. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Kemp, C., Perfors, A., & Tenenbaum, J. (2007). Learning overhypotheses with hierarchical bayesian models. *Developmental Science*, *10*, 307 – 321.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *117*, 357374.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*.

Papafragou, A., & Selimis, S. (2010). Lexical and structural biases in the acquisition of motion verbs. *Language Learning and Development*, *6*, 87–115.

Perfors, A. F., & Tenenbaum, J. B. (2009). Learning to learn categories. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

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

Pustejovsky, J. (1998). *The generative lexicon*. Cambridge, MA: MIT Press.

Samuelson, L., & Smith, L. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *71*, 1–33.

Smith, L., Jones, S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13-19.

Talmy, L. (1975). Semantics and syntax of motion. In J. Kimball (Ed.), *Syntax and semantics, volume 4.* New York, New York: Academic Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and lexical description: Grammatical categories and the lexicon.* Cambridge, England: Cambridge University Press.

Xu, F., Dewar, K., & Perfors, A. (2009). Induction, overhypotheses, and the shape bias: Some arguments and evidence for rational constructivism. In B. M. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 263–284). Oxford University Press.