The Onset of Syntactic Bootstrapping in Word Learning: Evidence from a Computational Study

Afra Alishahi
Computational Linguistics and Phonetics
Saarland University, Germany
afra@coli.uni-saarland.de

Pirita Pyykkönen
Computational Linguistics and Phonetics
Saarland University, Germany
pirita@coli.uni-saarland.de

Abstract

The syntactic bootstrapping hypothesis suggests that children's verb learning is guided by the structural cues that the linguistic context provides. However, the onset of syntactic bootstrapping in word learning is not well studied. To investigate the impact of linguistic information on word learning during early stages of language acquisition, we use a computational model of learning syntactic constructions from usage data, and adapt it to the task of identifying target words in novel situations. Our results show that having access to linguistic information significantly improves performance in identifying verbs (but not nouns) in later stages of learning, yet no such effect can be observed in earlier stages.

Introduction

Learning verbs is a challenging task for young children: their early vocabulary contains many more nouns than verbs, and they learn new nouns easier than new verbs of the same frequency (e.g., Imai et al., 2005; Waxman, 2006). The acquisition of nouns is mainly attributed to cross-situational evidence, or regularities across different situations in which a noun is used (Quine, 1960). In contrast, learning verbs seems to depend on the syntactic frames that they appear in. It has been suggested that children draw on syntactic cues that the linguistic context provides in verb learning, a hypothesis known as *syntactic bootstrapping* (Gleitman, 1990). According to this view, verbs are learned with a delay because the linguistic information that supports their acquisition is not available during the early stages of language acquisition.

To investigate the impact of linguistic and extralinguistic cues in identifying words, Gillette et al. (1999) proposed the *Human Simulation Paradigm* (HSP): adult participants watch videos of caregivers interacting with their toddlers, and are asked to identify target words marked by a beep. Videos are displayed without sound, and subjects are provided with different degrees of information about the linguistic context of the target verbs. Various HSP studies have shown that having access to linguistic and structural cues significantly improves the performance of adults in identifying verbs. Piccin & Waxman (2007) adopted the HSP paradigm for testing school-age children, and showed that children also rely on linguistic information for identifying verbs, but their performance is inferior to adults. These findings hint at a gradual development of syntactic bootstrapping, but it is uncertain whether the same effect can be observed in much younger children who have not mastered the syntactic structure of their language yet. A continuous picture of the developmental path of word learning is lacking.

In this paper, we propose a novel approach to studying this problem. We use an existing computational model of early verb learning which incrementally learns syntactic constructions of language from usage data. We adapt this model to the task of identifying target words in novel situations given different sets of (perceptual and linguistic) cues. Our results show that having access to linguistic information significantly facilitates identifying verbs in later stages of learning, but no such effect is observed at the earlier stages. For identifying nouns, additional linguistic information does not affect performance at all.

Time Course of Syntactic Bootstrapping

Several preferential-looking studies have shown that children are sensitive to the structural regularities of language from a very young age, and that they use these structural cues to find the referent of a novel word (e.g., Naigles & Hoff-Ginsberg, 1995; Gertner et al., 2006). In a typical setup, children are given more than one interpretation for an utterance (e.g., different activities displayed on parallel screens), and their looking behaviour reveals their preferred interpretation. However, these studies cannot compare the impact of different cues in word learning, since the same *type* of input is available to subjects in different conditions.

In contrast, HSP studies manipulate the number and type of cues that subjects receive for performing a task across conditions, and thus evaluate the impact of each set of cues. In their influential study, Gillette et al. (1999) provided their adult subjects with various combinations of visual cues (videos), a list of co-occurring words, the syntactic pattern of the sentence, and the full transcript of the narration. Their findings and those of later studies have consistently shown that the more linguistic information adult subjects receive, the more accurately they identify missing verbs.

Piccin & Waxman (2007) used HSP for studying seven-year-olds as well as adults. Subjects in each age group were randomly assigned to either 'no linguistic information' (-LI) or 'full linguistic information' (+LI) condition. In the -LI condition, participants heard no audio other than beeps indicating the target words. In the +LI condition, participants heard all the surrounding speech as well as the beeps. After watching each clip, subjects were asked to guess the target word (a noun or a verb). Their results show a similar pattern of behaviour for adults and children: In the -LI condition, all subjects identified nouns more successfully than verbs. In the +LI condition, linguistic information significantly improved the identification of verbs (but not nouns) by children as well as adults. The performance of both age groups was comparable in the -LI condition, but adults significantly outperformed 7-year-olds in the +LI condition. That is, adult subjects were more successful in incorporating linguistic information in the task of identifying verbs.

Due to the nature of the word guessing task, HSP is not suitable for very young children. Therefore, it is yet unknown whether linguistic information can facilitate word learning in the very early stages when the acquisition of syntax is still in progress. Computational modeling is an appropriate tool for tackling this problem: it allows us to examine the time course of word learning and the contribution of linguistic input from the very beginning.

Existing Computational Models

Many computational models have demonstrated that cross-situational learning is a powerful mechanism for mapping words to their correct meanings and explaining several behavioural patterns in children (e.g., Siskind, 1996; Fazly et al., 2010). These models ignore the syntactic properties of utterances and treat them as unstructured bags of words. Probabilistic models of Yu (2006) and Alishahi & Fazly (2010) integrate syntactic categories of words into a model of cross-situational learning and show that this type of information can improve the overall performance. In these models, perfect categories are assumed to be formed prior to cross-situational learning.

There are only a few computational models that explicitly study the role of syntax in word learning. Maurits et al. (2009) investigate the joint acquisition of word meaning and word order using a batch model. This model is tested on an artificial language with a simple relational structure of word meaning, and limited built-in possibilities for word order. The Bayesian model of Niyogi (2002) simulates the syntactic and semantic bootstrapping effects in verb learning (i.e. drawing on syntax for inducing the semantics of a verb, and using semantics for narrowing down possible syntactic forms in which a verb can be expressed). This model relies on extensive prior knowledge about the associations between syntactic and semantic features, and is tested on a toy language with very limited vocabulary and syntax. None of these models investigate the time course of syntactic bootstrapping and the differences between learning verbs and nouns.

Overview of our Computational Model

In a typical language learning scenario, a child observes an event which involves a number of participants, and at the same time hears a natural language utterance describing the observed scene. Such scene-utterance pairings are the main source of input for the acquisition of word-concept mappings as well as for learning syntactic constructions. Ideally, we need a computational model of syntactic bootstrapping to draw on such usage-based information in order to acquire form-meaning associations at word and sentence levels.

We investigate the time course of using syntax in word learning through computational simulation, using the construction learning model of Alishahi & Stevenson (2010). The model uses Bayesian clustering for learning the allowable frames for each verb, and their grouping across verbs into constructions. Each frame includes the conceptual properties of an event and its participants (the cross-situational evidence), and the linguistic properties of the utterance that accompanies the observed event. A construction is a grouping of frames which share form-meaning associations; these groupings typically correspond to general constructions in the language such as intransitive, transitive, and ditransitive.

By detecting similar frames and clustering them into constructions, the model forms probabilistic associations between syntactic positions of arguments with respect to the verb, and the conceptual properties of the verb and the arguments. These associations can be used in various language tasks where the most probable value for a missing feature must be predicted based on the available features. We simulate HSP in this fashion, where the most probable values for a missing head predicate (verb) or an argument (noun) are predicted based on the (perceptual and linguistic) information cues available in the current scene, using the acquired constructions.

The following sections review the model and describe the simulation of the word identification task.

Input and Frame Extraction

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

- Head words for the main predicate (i.e., verb) and its arguments (i.e., nouns or pronouns).
- Syntactic pattern, or the word order of the utterance.
- Number of arguments that appear in the utterance.
- Basic (conceptual) characteristics of the event (or verb), e.g., {`cause`,`change`,`rotate`,...}.
- Conceptual properties of the arguments which are independent of the event that the argument participates in, e.g., {`woman`,`adult`,`person`,...}.
- Event-based properties that each argument takes on in virtue of how it participates in the event, e.g., {`moving`,`volitional`,...}.

In the Experimental Results section, we explain the selection of semantic properties in our simulations.

Learning Constructions

Each extracted frame is input to an incremental Bayesian clustering process that groups the new frame together with an existing group of frames—a construction—that probabilistically has the most similar properties to it. If none of the existing constructions has sufficiently high probability for the new frame, then a new construction is created, containing only that frame.

Adding a frame F to construction k is formulated as finding the k with the maximum probability given F:

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

where k ranges over the indices of all constructions, with index 0 representing recognition of a new construction. Using Bayes rule, and dropping $P(F)$ which is constant for all k:

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

The prior probability $P(k)$ indicates the degree of entrenchment of construction k, and is given by the relative frequency of its frames over all observed frames. The posterior probability of a frame F is expressed in terms of the individual probabilities of its features, which we assume are independent, thus yielding a simple product of feature probabilities:

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

where j is the value of the i^{th} feature of F, and $P_i(j|k)$ is the probability of displaying value j on feature i within construction k. This probability is estimated using a smoothed version of this maximum likelihood formula:

$$P_i(j|k) = \frac{\text{count}_i(j,k)}{n_k} \quad (4)$$

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

For single-valued features (head words, number of arguments, syntactic pattern), $\text{count}_i(j,k)$ is calculated by simply counting those members of construction k whose value for feature i exactly matches j. However, for features with a set value (semantic properties of the verb and the arguments), counting the number of exact matches between the sets is too strict, since even highly similar words very rarely have the exact same set of properties. We instead assume that the members of a set feature are independent of each other, and calculate the probability of displaying a set s_j on feature i in construction k as

$$P_i(s_j|k) = \frac{1}{|S(i)|} \prod_{j \in s_j} P_i(j|k) \times \prod_{j \in S(i)-s_j} P_i(\neg j|k) \quad (5)$$

$P_i(j|k)$ and $P_i(\neg j|k)$ are estimated as in Eqn. (4) by counting members of construction k whose value for feature i does or does not contains j. The product is rescaled by the length of $S(i)$, which is the superset of all the values that feature i can take.

Identifying Nouns and Verbs

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

The probability of an unobserved feature i displaying value j given other feature values in a partial frame F is estimated as

$$\begin{aligned} P_i(j|F) &= \sum_k P_i(j|k)P(k|F) \quad (6) \\ &= \sum_k P_i(j|k)P(k)P(F|k) \end{aligned}$$

The conditional probabilities $P(F|k)$ and $P_i(j|k)$ are determined as in the learning module. Ranging over the possible values j of feature i, the value of an unobserved feature can be predicted by maximizing $P_i(j|F)$:

$$\text{BestValue}_i(F) = \underset{j}{\text{argmax}}\ P_i(j|F) \quad (7)$$

Identifying a target verb as in HSP can be simulated as finding the head verb j with the highest $P_{\text{verb}}(j|F)$, or estimating $\text{BestValue}_{\text{verb}}(F)$. Here F is a partial frame which can include only the semantic features, or additional linguistic and syntactic features. Similarly, identifying a target noun which corresponds to an argument i in a scene is modeled as estimating $\text{BestValue}_{\text{noun}_i}(F)$.

Experimental Results

We use our computational model to investigate the role of linguistic input in learning verbs and nouns. Following Piccin & Waxman (2007), we pursue three main goals in our experiments. First, to simulate the task of identifying a target word in the presence of visual stimuli, and to study the impact of linguistic input on performance. Second, to investigate whether verbs benefit more than nouns from linguistic cues. Third, to examine the role of linguistic input in identifying verbs versus nouns in early stages of learning.

Factors and conditions. We model different factors in the study of Piccin & Waxman (2007) as follows:

- *Word category*: we simulate the identification of a target verb and a target noun as estimating $\text{BestValue}_{\text{verb}}(F)$ and $\text{BestValue}_{\text{noun}_i}(F)$ respectively, based on Eqn. (7).
- *No (-LI) vs. full (+LI) linguistic information*: the information cues available to subjects are reflected by the included features in partial frame F in equations (6) and (7). In the -LI condition, included features are the properties of the event and the conceptual and event-based properties of the arguments (observable from a

muted clip). In the +LI condition, the following features are also included: number of arguments, the head words of the main verb and the arguments (except for the target word), and the syntactic pattern (available from the narration of the clip).

- *Age groups*: we train our model on a set of scene-utterance pairs before evaluating it on a word identification task. The age of the model is determined by its exposure to input data prior to performing the task. We simulate different age groups by varying the size of the training data.

Evaluation. In evaluating the model when identifying target words in a test set, we use the following criteria:

- *Absolute accuracy*: the number of test items for which $\text{BestValue}_i(F)$ in Eqn. (7) returns the correct value for the target word i.
- *Probabilistic accuracy*: the sum of the probabilities $P_i(\textit{target}|F)$ for each *target* (the correct answer) in the test set, where i is the focus feature and $P_i(\textit{target}|F)$ is calculated using Eqn. (6). This probability reflects the *confidence* of the model in predicting *target*.
- *Improvement*: the gain achieved by using our prediction model instead of a simple frequency baseline: $(P_i(\textit{target}|F) - \text{baseline}(\textit{target}))/P_i(\textit{target}|F)$. The baseline only relies on the relative frequency of each head word ($\text{freq}(w)/\sum_{w'}\text{freq}(w')$, where $\text{freq}(w)$ is the frequency of w in the input).

Data. We used the Brown corpus of the CHILDES database (MacWhinney, 2000) for constructing the input to our model. We extracted the 20 most frequent verbs in mother's speech to each of Adam, Eve, and Sarah, and selected 13 verbs from those in common across these three lists. We constructed an input-generation lexicon based on these 13 verbs, including their total frequency among the three children. We also assigned each verb a set of possible argument structure frames and their relative frequencies, which were manually compiled by the examination of 100 randomly sampled uses of a verb from all conversations of the same three children. Finally, from the sample verb usages, we extracted a list of head words (total 259) that appeared in each argument position of each frame, and added these to the lexicon.

For each noun in the lexicon, we extracted a set of lexical properties from WordNet (Miller, 1990) as follows. We extracted all the hypernyms for the first sense of each word, and added one member from each hypernym synset to the list of its properties. For each verb frame, we manually compiled a set of semantic primitives for the event as well as a set of event-based properties for each of the arguments. We chose these properties from what we assumed to be known to the child at the stage of learning being modeled.

Due to the limited lexicon, the scale of the learning problem in our experiments is much smaller than the real-world experience of children. However, our dataset has the same statistical characteristics as the input data that children receive, and therefore we hope that the general patterns observed in our experiments are representative of the developmental patterns in children.

Linguistic Input and Identifying Words

In the experiments reported below, we simulate the identification of nouns and verbs in the -LI and +LI conditions across different simulations, where each simulation represents a subject. For each simulation, we randomly generate a set of 500 training items and 30 test items, using the input generation lexicon discussed above. We monitor the performance of the model in intervals: after processing every 10 training items, we independently predict the head verb and a randomly selected argument noun for each of the test items, and evaluate the predictions using the evaluation measures mentioned above.

A simple linear model of probabilistic accuracy with input type and age as predictors reached a significant interaction between the predictors for both nouns ($F = 3.983$, $p = 0.003$) and verbs ($F = 7.375$, $p = 0.000$). The same was observed in the improvement data (nouns: $F = 6.376$, $p = 0.000$; verbs: $F = 10.249$, $p = 0.000$). That is, different age groups behave differently when processing linguistic input. In order to further examine the effect of linguistic information throughout development, we conducted separate analyses for each word category across all age groups, as reported below.[1]

Verbs. Figure 1 shows the absolute accuracy and improvement of identifying verbs for 30 test items, in intervals of 10 over a total of 500 input items, averaged over 50 simulations. (The improvement and probabilistic accuracy plots show a very similar trend.) As can be seen from the top panel, a target verb can be identified more accurately when the model has access to linguistic information about the co-occurring words and syntactic pattern of the utterance. The bottom panel shows the same pattern, and further emphasizes the benefit of using perceptual and linguistic features in our model compared to predicting verbs based on their frequency of observation in the input data. Performance is boosted as early as processing 100 training items (A100), a stage at which relatively robust constructions are formed by the model. The gap between the accuracy and improvement curves

[1] The probabilistic accuracy and improvement were analyzed with linear mixed models with the condition (+LI, -LI) as a fixed factor and subjects and items as a crossed-random factor in order to allow by-subject and by-item variation in one model. Estimates (*Est*) report the regression coefficients for the fixed effect, and p-values for estimates were obtained using Markov-Chain Monte Carlo (MCMC) sampling with 10000 replications. The absolute accuracy was analyzed with a logistic mixed model with the condition as a fixed factor and subjects and items as a crossed-random factor. P-values for estimates were obtained from z-statistics.

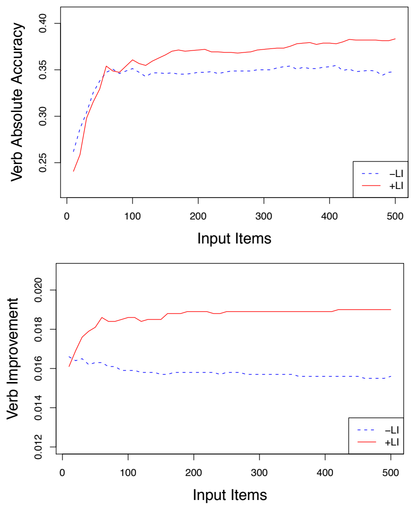

Figure 1: Absolute accuracy and improvement of identifying verbs, averaged over 50 simulations.

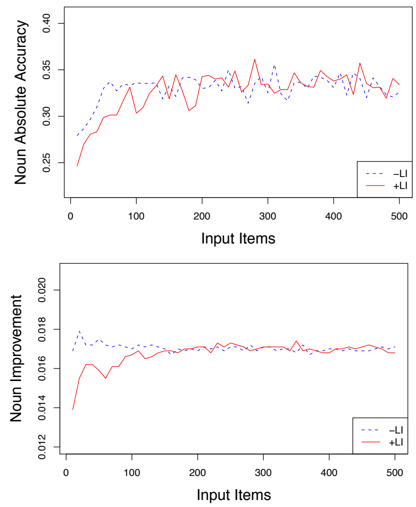

Figure 2: Absolute accuracy and improvement of identifying nouns, averaged over 50 simulations.

in the -LI and +LI conditions widens slowly but consistently as the model processes more input items; that is, the model can use linguistic input more efficiently as it ages. In both age groups, the probabilistic accuracy was positively affected by the presence of linguistic context (A100: $Est = 0.041$, $p_{\mathrm{MCMC}} = 0.0001$; A500: $Est = 0.053$, $p_{\mathrm{MCMC}} = 0.0001$). Same effect was also found in improvement measurements (A100: $Est = 0.041$, $p_{\mathrm{MCMC}} = 0.000$; A500: $Est = 0.053$, $p_{\mathrm{MCMC}} = 0.0001$). The effect on absolute accuracy was significant only in the last age group ($Est = 0.157$, $p = 0.040$) and marginally significant for age group A200 ($Est = 0.108$, $p = 0.161$).

The behaviour of the model at the A100 stage is similar to that of the 7-year-old subject group in Piccin & Waxman (2007), whereas the A500 stage is more similar to their adult subject group. Note that due to the small number of verbs in our lexicon and their relatively restricted syntactic behaviour, our model learns much more efficiently from a small training corpus.

Nouns. Figure 2 shows the absolute accuracy and improvement of identifying nouns, averaged over (the same) 50 simulations. These results show a different pattern than those of verbs: the model outperforms the frequency baseline, but there is no clear advantage of using linguistic input. In the oldest age group (A500), neither probabilistic accuracy nor improvement were affected by +LI and -LI manipulation ($p > 0.2$). Interestingly, in the age group A100, both probabilistic accuracy ($Est = -0.019$, $p_{\mathrm{MCMC}} = 0.0294$) and improvement ($Est = -0.011$, $p_{\mathrm{MCMC}} = 0.040$) were negatively affected by the linguistic context. The absolute accuracy showed no effects for the oldest age group ($p > 0.6$) and a marginal effect in A100 ($Est = -0.151$, $p = 0.054$).

Our results are in line with the findings of Piccin & Waxman (2007): older subjects (A500) perform better than younger ones (A100) in identifying verbs and nouns. More importantly, exploiting linguistic input significantly facilitates identifying verbs, and older subjects can use this information more efficiently than younger ones. However, identifying nouns does not benefit from additional linguistic input.[2] The gradual improvement of verb identification in the +LI condition brings us back to our original question: when does syntax begin to play a role in verb identification? We address this issue next.

Onset of Syntactic Bootstrapping

In order to investigate the contribution of linguistic input in identifying words in the earlier stages of learning, we zoom in on the performance of the model during processing the first 100 training items. Figure 3 shows the absolute accuracy of predicting verbs and nouns for 30 test items, in intervals of 5 over the course of processing 100 input items, averaged over 50 simulations. (The improvement plots are not included due to lack of space.) The curves show an interesting trend: for both verbs and nouns, linguistic information does not help at first. For verbs, the positive effect of +LI was absent in the earliest age group for both probabilistic accuracy and improvement (A10: $p > 0.6$), and for absolute accuracy as well ($p > 0.1$). However, the accuracy curve in the +LI condition takes over the -LI condition around A50, and shows significant influence of linguistic input in both probabilistic accuracy ($Est = 0.028$, $p_{\mathrm{MCMC}} = 0.001$) and improvement ($Est = 0.028$, $p_{\mathrm{MCMC}} = 0.0001$), but not in absolute accuracy ($p > 0.6$).

The results for nouns are more surprising: there is a significant negative effect of +LI for the A10 and A50 age groups in both probabilistic accuracy (A10: $Est = -0.036$, $p_{\mathrm{MCMC}} = 0.0008$; A50: $Est = -0.028$, $p_{\mathrm{MCMC}} = 0.0018$) and improvement (A10: $Est = -0.031$, $p_{\mathrm{MCMC}} = 0.0001$; A50: $Est = -0.021$, $p_{\mathrm{MCMC}} = 0.0008$). Absolute accuracy

[2] Due to the incomparable number of verbs and nouns in our lexicon and different methods of identifying them, we cannot directly compare performance in predicting verbs and nouns.

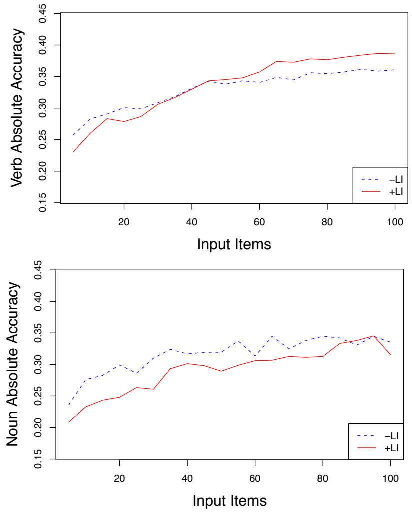

Figure 3: Average absolute accuracy of identifying verbs and nouns in early stages of learning.

showed a significant negative effect of +LI in A10 ($Est = -0.172$, $p = 0.039$) and a marginal effects in A50 ($Est = -0.148$, $p = 0.062$).

Discussion

The results of our computational simulations replicate the experimental findings of Piccin & Waxman (2007) that syntactic information boosts the identification of verbs by adults and young children. Our results also suggest that the boosting effect comes into play with a delay, and only after enough input data is processed and a relatively stable knowledge of syntactic constructions is formed. Our computational approach allows us to investigate word identification throughout various stages of development, and examine syntactic bootstrapping for age groups which cannot be easily studied in experimental settings. Specifically, our model predicts that very young children's verb learning might not be modulated by linguistic information, even though a significant impact can be found in the later stages of development. This prediction is in line with previous suggestions that generalization of syntactic information takes time to manifest (e.g., Gillette et al., 1999). Importantly, this prediction is not inconsistent with findings on the sensitivity of very young children to syntax during *comprehension* (see Alishahi & Stevenson (2010) for simulating such effects using the same computational model).

Our results make another (somehow surprising) prediction: linguistic context might have a negative effect on identifying nouns during the early developmental stage. The performance of our model in guessing nouns for the younger age groups was poorer when the linguistic information was provided, and no effect on performance by linguistic information was observed in later age groups. This might be due to the fact that most early nouns refer to observable concepts, and are less dependent on the structure of their linguistic context than verbs. Our training corpus might also play a role: it contained many more nouns (259) than verbs (13), and most verbs were not restrictive. Therefore, the same nouns can appear as arguments of different verbs, or many different nouns can be potential candidates for a verb argument, yielding several correct answers for a noun-guessing task.

It should be noted that the cross-situational scenario in the setup of our model is not realistic as there is no referential uncertainty in our data (i.e., there are no referents in the scene which are not mentioned in the utterance), an issue we plan to address in the future. But it only highlights our point that syntactic bootstrapping can facilitate verb learning even in low-ambiguity situations, given that the learner has been exposed to enough input to form a reliable knowledge of the structure of language.

References

Alishahi, A., & Fazly, A. (2010). Integrating syntactic knowledge into a model of cross-situational word learning. In *Proceedings of CogSci'10.*

Alishahi, A., & Stevenson, S. (2010). A computational model of learning semantic roles from child-directed language. *Language and Cognitive Processes*, *25*(1), 50–93.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psych. Science*, *17*(8), 684–691.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 135–176.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? *Child Development*, *76*(2), 340–355.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Lawrence Erlbaum Associates Inc, US.

Maurits, L., Perfors, A. F., & Navarro, D. J. (2009). Joint acquisition of word order and word reference. In *Proceedings of CogSci'09.*

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

Naigles, L., & Hoff-Ginsberg, E. (1995). Input to verb learning: evidence for the plausibility of syntactic bootstrapping. *Developmental Psychology*, *31*(5), 827–37.

Niyogi, S. (2002). Bayesian learning at the syntax-semantics interface. In *Proceedings of CogSci'02.*

Piccin, T., & Waxman, S. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the Human Simulation Paradigm. *Language Learning and Development*, *3*(4), 295–323.

Quine, W. (1960). *Word and object.* Cambridge: MIT Press.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–91.

Waxman, S. R. (2006). Early word learning. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology.* Wiley.

Yu, C. (2006). Learning syntax–semantics mappings to bootstrap word learning. In *Proceedings of CogSci'06.*