Learning Syntax-Semantics Mappings to Bootstrap Word Learning

Chen Yu (chenyu@indiana.edu)
Department of Psychological and Brain Sciences, and Program in Cognitive Science
Indiana University, Bloomington, 47405, IN USA

Abstract

This paper addresses possible interactive effects between word learning and syntax learning at an early stage of development. We present a computational model that simulates how the results from a syntax learning process and a word learning process can be integrated to build syntax-semantics mappings, and how the emergence of links between syntax and word learning could facilitate subsequent word learning. The central idea of our statistical model is to categorize words into groups based on their syntactic roles and then estimate semantic meanings of those syntactic categories using lexical knowledge acquired from a concurrent word learning process. Once built, those syntax-semantics mappings can be further utilized as a syntactic constraint in statistical word learning. We applied the model to realistic data collected from child-mother picture book reading interaction. A comparative study between a statistical model and the model based on both statistical and syntactic information shows that syntactic cues can be seamlessly integrated in statistical learning and significantly improve word learning performance.

Introduction

One of the most complex learning tasks young children are faced with is to learn their native language. Language acquisition, of course, consists of several distinct tasks, such as speech perception, speech segmentation, word learning and syntax learning. Among others, word learning involves how to map a phonological form to a conceptual representation, such as associating the sound “dog” to the concept of dog. Thus, the crucial issue in word learning is to build word-to-world mappings from language and extralinguistic contexts. Syntax learning, on the other hand, is mainly about how to categorize words into grammatical categories (e.g. noun, verb, etc.) which are basic building blocks of grammar, and then how to acquire the hierarchical and context-sensitive structures that are represented by those syntactic categories. Therefore, syntax learning uses sequential symbolic data (sentences in a language, etc.) to construct a grammar.

Although acquisition of the lexicon and acquisition of the grammar seem to address totally different issues, these two learning processes might be closely related due to universal correspondences between syntax and semantics. For instance, Bloom (1994) pointed out bidirectional mappings between syntax and semantics, such as count nouns to kinds of individuals, mass nouns to kinds of portions, and Noun Phrases (NPs) to individuals. Such mappings suggest a possible bootstrapping procedure between these two learning processes – the progresses in one learning process could facilitate the other learning process. In fact, two compelling hypotheses have been proposed by theorists. The semantic bootstrapping hypothesis (Pinker, 1989) argued that word meanings can be antecedently acquired from the observation of events and then used to determine the syntactic category of each word and to deduce the word argument structures. In contrast, Gleitman (1990) proposed an alternative account called syntactic bootstrapping. She argued that children use syntactic knowledge they have developed to learn what words mean. More specifically, the semantically relevant syntactic structures surrounding a verb, such as the subcategorization frames around a verb, provide contextual cues for its meaning. These two hypotheses focus on different aspects of potential interactions between syntax and semantic learning, and both of them have been supported by empirical studies.

The present paper proposes a computational model of how syntax-semantics mappings can be learned and emerged from two language learning processes – syntax learning and word learning, and how these mappings can then facilitate word learning. Our study is quite different from previous work in several important ways. First, we propose and implement a general statistical-learning mechanism in which syntactic cues can be seamlessly integrated with already learned semantic knowledge to help the learning of new words. We suggest that syntax can act as a linguistic spotlight that facilities word learning by selecting, grouping and highlighting those words that are likely to have the same type of referents. Using the proposed learning mechanism, we demonstrate how syntactic learning could help object name learning and how the development of grammatical abilities continues to be highly linked to lexical development. Second, both the proposed learning mechanism of syntax-semantics mappings and the mechanism of utilizing the mapping knowledge in word learning are general that can be applied not only to a specific syntactic category (verb, etc.) but also to other categories. Thus, we suggest that the acquisition of syntax and the integration of syntactic cues in word learning might partially account for the explosive expansion of vocabulary as primary syntactic structures are gradually acquired. Third, we apply the model to raw data collected from everyday parent-children interaction but not to some artificial or synthesized data, and show a dynamic picture of how the learning mechanism works with realistic input.

Related Work

There are a number of existing models that account for different aspects of word learning. Plunkett, Sinha, Miller, and Strandsby (1992) built a connectionist model of word learning in which a process termed autoassociation mapped preprocessed images with linguistic labels. The linguistic behavior of the network exhibited non-linear vocabulary growth (vocabulary spurt) that was similar to the pattern observed in young children. Colunga and Smith (2005) presented a connectionist model showing that regularities among object and substances categories were learnable and generalizable enabling the system to become, after training, a more rapid learner of new object and substance names. Siskind (1996) developed a mathematical model based on cross-situational learning and the principle of contrast, which learned word-meaning associations when presented with paired sequences of pre-segmented tokens and semantic representations. Regier (2005) suggested that attention to relevant aspects of form and meaning could account for developmental changes without a change in associative mechanism. Tenenbaum and Xu (2000) developed a computational model based on Bayesian inference which could infer meanings from one or a few examples without encoding the constraint of mutual exclusion. Li, Farkas, and MacWhinney (2004) proposed a SOM-based developmental model that learned topographically organized representations for linguistic categories over time. However, the role of syntax in word learning has not been systematically studied in cognitive development (but also see Siskind, 1992 using syntactic constraints to help the acquisition of semantics based on a logic-inference mechanism). In addition, artificial or synthesized data are used to demonstrate a model's performance and illustrate the key ideas in a model. In contrast, this work applies the data collected from natural parent-child interaction to our proposed model and the results show what a learning mechanism could achieve from realistic data.

Data

Six 18-month-old children and their parents participated in data collection. Each parent was asked to narrate one picture book. In total, six books for 1-3 year old children were used. Parents were also instructed to act naturally without any constraint about what they had to say or what they had to do. Picture book narration is one of common parent-child activities in everyday life from which children learn the names of objects shown in the picture books. Therefore, the data collected from this setting is realistic and representative of everyday word learning. The data used in this simulation study were our descriptions of video clips. More specifically, our description of the audio input – what we feed into the statistical simulated learner – is the entire list of spoken words. Our description of the video stream, again what we feed into the statistical learner, is the list of all the (basic-level) objects in picture books that a narrator was attending to from moment to moment when spoken utterances were produced. Table 1 shows several examples wherein each row represents one learning situation (defined by speech silence) consisting of multiple words and multiple objects.

Table 1: Examples of training data

speech	visual context
is that a little baby	boy, flowers, bird
and what is the little baby holding	boy, flowers, bird
that is right flowers	boy, flowers, bird
......	......
that is a pumpkin and look	boy, pumpkin, leave
what is this back there	boy, pumpkin, leave
......	......
look what he is doing now	boy, hat, bird, wall
......	......

The statistics of the data set (the sum over six subjects) are described in Table 2. The learning environment is rather highly ambiguous wherein on average more than 8 words and 3 objects co-occur in a single learning moment without any information about which word goes to which object from a trial itself. Only $3.13\%(132/4223)$ of co-occurring pairs are correct, which shows the difficulty of the word-learning task. As shown in Table 1, every word in a learning situation can be potentially associated with any co-occurring object. Thus, the learning task for both young children and simulated learners is to find a very few correct lexical items from a huge amount of irrelevant co-occurring words and objects.

Table 2: Statistics of training data

# of words	# of unique words	words per situation
3571	581	>8
# of objects	# of unique objects	objects per situation
1230	113	>3
# of pairs	# of unique pairs	# of correct pairs
12173	4223	132

The Model

Our model consists of three components: statistical word learning without syntax, syntax learning and the integration of syntactic knowledge in word learning. The central idea is that a syntax learning process can learn structural information in unsupervised mode based on statistical regularities in speech. Meanwhile, a word learning process uses co-occurrence regularities between language and extralinguistic contexts to build word-referent mappings. Importantly, the results in these two learning processes can be merged to generate new kinds of statistical regularities. More specifically, the syntax learning process categorizes words into several groups based on their linguistic roles. Semantic meanings of the words in a syntactic category can be acquired through the word learning process and jointly determine the semantic meaning of that syntactic category. Thus, the integration of the results in these two learning processes generates the mappings between syntax and semantics, which in turn can facilitate lexical acquisition by considering the syntactic role of a new word in word learning. The following subsections will describe this learning mechanism in detail.

Statistical Word Mapping Without Syntactic Cues

In early word learning without syntactic cues, children have to start by pairing spoken words with co-occurring contexts, collecting multiple such pairs, and then figuring out the common elements. Although no one doubts this process, there has been few modeling studies (but also see Siskind, 1996). Yu, Ballard, and Aslin (2005) introduce a formal model of statistical word learning which provides a probabilistic framework for encoding multiple sources of information. Given multiple scenes paired with spoken words collected from natural interactions between caregivers and children, the model is able to compute the association probabilities of all the possible word-meaning pairs.

The general setting is as follows: suppose we have a word set $X = \{w_1, w_2, ..., w_N\}$ and a meaning set $Y = \{m_1, m_2, ..., m_M\}$, where N is the number of words and M is the number of meanings (basic-level objects, etc.). Let S be the number of spoken utterances. All word data are in a set $\chi = \{(S_w^{(s)}, S_m^{(s)}), 1 \leq s \leq S\}$, where each spoken utterance $S_w^{(s)}$ consists of r words $w_{u(1)}, w_{u(2)}, ..., w_{u(r)}$, and $u(i)$ can be selected from 1 to N. Similarly, the corresponding contextual information $S_m^{(s)}$ include l possible meanings

$m_{v(1)}, m_{v(2)}, ..., m_{v(l)}$ and the value of $v(j)$ is from 1 to M. Assume that every word w_n can be associated with a meaning m_m. Given a data set χ, We use the machine translation method proposed by Brown, Pietra, Pietra, and Mercer (1994) to maximize the likelihood of generating the meaning strings given English descriptions:

$$\begin{aligned} & P(S_m^{(1)}, S_m^{(2)}, ..., S_m^{(S)} | S_w^{(1)}, S_w^{(2)}, ..., S_w^{(S)}) \\ = & \prod_{s=1}^{S} \sum_a p(S_m^{(s)}, a | S_w^{(s)}) \\ = & \prod_{s=1}^{S} \frac{\epsilon}{(r+1)^l} \prod_{j=1}^{l} \sum_{i=0}^{r} p(m_{v(j)} | w_{u(i)}) \end{aligned}$$

where the alignment a indicates which word is aligned with which meaning. $p(m_{v(j)}|w_{u(i)})$ is the association probability for a word-meaning pair:

$$p(m_{v(j)}|w_{u(i)}) = p(m_{v(j)}|w_{u(i)}, g_{u(i)}) p(g_{u(i)}|w_{u(i)})$$

where $p(g_{u(i)}|w_{u(i)})$ is the probability that a word has a referent and $p(m_{v(j)}|w_{u(i)}, g_{u(i)})$ is the association probability of a word-referent pair given that the word has a referent. Without any linguistic knowledge and starting from scratch, the model assumes that every word could have a referent. Therefore, $p(g_n|w_n)$ is set to be 1 for every word. The estimate of $p(m_m|w_n, g_n)$ can be found in Yu et al. (2005). The upper right figure in Figure 1 shows the results of word-referent association probabilities $p(m_{v(j)}|w_{u(i)})$.

Early Syntax Learning

We used to the learning algorithm developed by Solan, Horn, Ruppin, and Edelman (2005) to extract linguistic structures. The method represents sentences as paths on a graph and words as vertices on the paths. It aligns and identifies those sentences that share some words and extracts both common and variant words in those sentences. The approach then progressively infers linguistic structures from the accrued statistical knowledge. For instance, given two simple sentences "this is a cat" and "here is a dog", the method can extract the pattern "**x** is a **y**" while **x** can be replaced by {this, here} and **y** stands for {cat, dog}. Technical details can be founded in Solan et al. (2005).

Table 3 shows examples of applying the syntax learning method on our data. Induced syntactic structures are represented in two forms: patterns and equivalence classes. A pattern represents a set of full (or part of) sentences or phrases that share common symbols. Those symbols can be either a word or a group of words termed an equivalence class that can be replaced in the pattern to form different sentences or phrases. For instance, the pattern P587 can represent different phrases by selecting different members in the equivalence classes E582 and E588, such as "and the rabbit", "and the rooster", "at the rabbit", "below the rooster" and so on. Thus, the members in an equivalence class of a pattern play the same syntactic role and are replaceable in context of that syntactic pattern. We notice that most equivalence classes are not necessarily identical to grammatical categories. Although all the members in E588 are object names, E586 contains not only object names (apples, flowers,etc.) but also pronouns (they,these,etc.). This is because the unsupervised method here can induce only partial and not-precise knowledge from a limited amount of data. From a developmental perspective, this result might be quite in line with the learning performance of young children who also learn syntactic knowledge in an unsupervised manner and would not learn a grammar overnight but instead gradually acquire and accumulate syntactic knowledge. In this way, equivalence classes can be treated as temporary results toward grammatical categories. The next section will show how this partial linguistic knowledge can be utilized in word learning.

Table 3: Examples of the results of syntax learning

E582	{ and at below hope so in by maybe except... }
E584	{ big little }
E586	{ apples here there these they flowers }
E588	{ rabbit rooster bear bunny cat mom dog duck }
E590	{ front case }
P583	[a E584]
P585	[E586 are]
P587	[E582 the E588]
P589	[in E590 of]

Learning Syntax-Semantics Mappings

The general idea is like this: the syntax learning process categorizes words into groups based on their syntactic roles. Meanwhile, the word learning process builds the association probabilities between words and objects. Building syntax-semantics mappings can then be accomplished by integrating the results of these two learning processes. The present model explores how two specific mappings could emerge from the integration: (1) those words with high association probabilities with objects (thus, with similar semantic properties) are also likely to be in the same syntactic groups built by the syntax learning process, and (2) the words that are less likely to refer to objects are also grouped together based on their syntactic roles. In this way, all the words in a syntactic class can jointly define the semantic role of that class. Next this link between syntax and semantics can be used to guide subsequent word learning. For example, if most words in a syntactic class are associated with some object kinds and therefore they are likely to be object names, other words in the same class should also be likely to associate with object kinds because of the inherent relationship between count nouns and object names. Thus, when the model considers whether a new word refers to an object, it is based on not only the association probability between these two (co-occurrence statistical regularities between word-to-world mappings) but also what syntactic class this word is in. By doing so, syntax-semantics mappings may improve the performance of statistical word learning.

Without any linguistic cues, the previous model assumes that every word (even function words) could be associated with an object referent and therefore estimates the association probabilities between any co-occurring word-object pairs. Now association probabilities are estimated by two parts:

$$\begin{aligned} p(m|w) &= p(g|w)p(m|w,g) \\ &= \sum_C p(C|w)p(g|C) \times p(m|w,g) \end{aligned}$$

where a new variable C represents the syntactic class of a word. The model uses both the syntactic class of a word $p(C|w)$ and the semantic property of the syntactic class $p(g|C)$ to estimate the probability that this word refers to

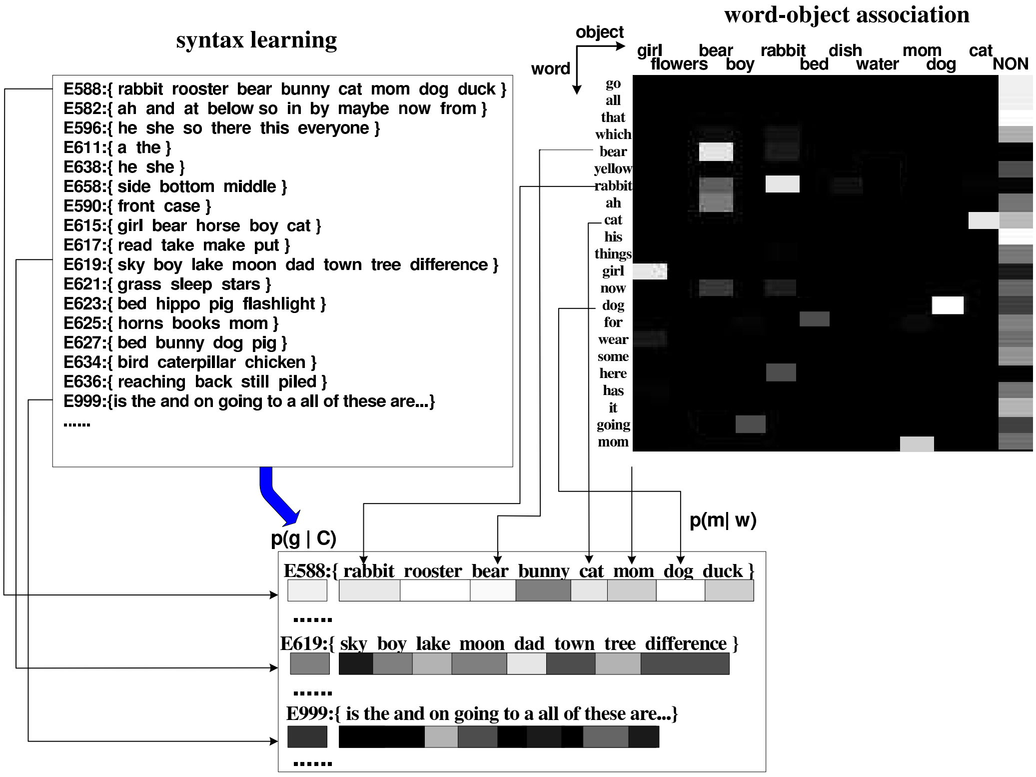

Figure 1: **Learning syntax-semantics mappings.** Upper left: the syntax learning process categorizes words into several syntactic groups. Upper right: the word-learning process estimates the association probabilities of any co-occurring word-object pairs represented by cells in the figure. White color means high association probabilities and dark color means low association probabilities. Bottom: the integration of semantic and syntactic knowledge results in $p(g|C)$ – the semantic property of each syntactic class. $p(g|C)$ can then be used to improve the estimates of word-object association probabilities by considering semantic properties of other words in the same groups.

an object kind. In practice, a word could appear in more than one equivalence (syntactic) classes, each of which is associated with a linguistic pattern. Therefore, $p(C|w)$ is the probability that a word w belongs to a syntactic class C and $\sum p(C|w) = 1$. The probability that a word is in a specific class can be estimated based on normalizing the occurrences of the word in that pattern:

$$p(C_i|w_j) = \frac{\# < C_i, w_j >}{\# \ w_j}$$

Meanwhile, $p(g|C)$ is the probability that the words in a syntactic class C refer to object-kind categories, which is jointly determined by the association probabilities of all the words in this class:

$$p(g_i|C_i) = \frac{1}{|C_i|} \sum_{w \in C_i} \max_{m \in Y; m \neq NON} p(m|w, g)$$

Figure 1 illustrates the syntax-semantics mapping mechanism with examples. If a word is in a syntactic class wherein other words have high association probabilities to objects, the probability that this word also associates with an object kind would increase. Similarly, if a word is syntactically grouped with other words without semantic mappings (such as function words), it is less likely that the word refers to an object kind. Overall, $p(m|w)$ is determined not only by $p(m|w, g)$ but also the semantic properties of all the syntactic classes that this word belongs to.

Experimental Results

We applied the same data set to two learning approaches. One is purely based on statistical learning as described in the above section. The other approach uses syntax-semantics mappings to facilitate word learning. In both approaches, a lexical item $L(m_j \Leftrightarrow w_i)$ is discovered based on both association probabilities of a word-object pair and the number of their co-occurring times:

$$L(m_j \Leftrightarrow w_i) = p(m_j|w_i) \times (\# < m_i, w_j >)$$

Both models can then set up a threshold to select a set of word-object pairs from all the co-occurring ones in the association matrix. Two metrics are used to evaluate the word-learning performances for these two approaches: (1) word-learning accuracy measures the proportion of selected pairs that are actually correct and (2) word-learning completeness measures the proportion of correct pairs in the data that a model successfully selects. The choice of different thresholds leads to different values in accuracy and completeness. To compare these two approaches, we make one metric constant and measure the difference in the other metric. In one measure as shown on the left of Figure 2, the completeness percentage is fixed and we show that encoding syntactic cues improve accuracy. Similarly, the completeness in the statistical and syntactic model is better than that of the purely statistical model when the accuracy is fixed.

Table 4 shows the top 25 word-object pairs selected by two models as a comparison. 15 word-object pairs are selected by both models and all of them are correct except one pair "climbs"–*boy*. For the pairs (marked by *) selected by the statistical model only, 2 out 10 are correct. One crucial reason that most of those pairs are incorrect is that some function

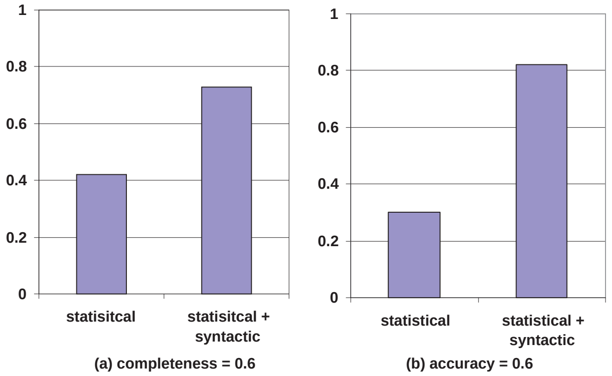

Figure 2: **A comparison of statistical learning and statistical learning with syntactic cues.** Left: With a fixed value of completeness, the accuracy of our proposed model is much better than that of the statistical model. Right: With a fixed accuracy, the completeness of our model is significantly better than that of the statistical model.

words happen to frequently co-occur with some specific objects (but not other objects). Therefore, both their association probabilities and their numbers of co-occurrences are relatively high. In contrast, the statistical and syntactic approach improved the performance by removing those pairs from the top items in its list based on the syntactic roles of those function words. As shown in Table 4, 8 out 10 pairs (marked by ○) selected by our model are correct. Overall, these two lists provide a concrete example of the differences between these two approaches.

General Discussion

Three statistical learning mechanisms are introduced and implemented in the model: statistical word learning to build word-to-world mappings, statistical syntax learning to acquire linguistic patterns, and word learning with syntax-semantics mappings. This section discusses relevant experimental studies and findings that support the cognitive plausibility of those learning mechanisms.

Statistical word learning One of the most important findings in language acquisition is that humans are sensitive to statistical regularities in language and are able to acquire linguistic knowledge based on statistical learning. Saffran, Aslin, and Newport (1996) demonstrated that 8-month-old infants are able to find word boundaries in an artificial language based only on statistical regularities. Can statistical learning also account for word acquisition? The kind of statistical learning requested in word-to-world mappings would be quite different from statistical speech segmentation – not simply count the frequency or condition probabilities of word or syllables in a speech stream, but compute co-occurring statistical regularities across language and extralinguistic contexts. Nonetheless, our recent findings (Yu & Smith, submitted) show that both adult and children are sensitive to statistical regularities. When presented with multiple trials, each containing multiple pictures and names with no information about which picture is paired with which name, both adults and even 12-month old babies are able to build correct picture-name pairings. The computational model of statistical word learning in this work demonstrates how such learning mechanism works.

Statistical syntax learning Gomez and Gerken (1999) have shown that after less than 2-min exposure to one of

Table 4: The top 25 lexical items learned in these two methods. The marked items are word-object pairs that are selected by one model but not the other. The bold words are incorrect.

	statistical		statistical+ syntactic
	“tree” – *tree*		“dog” – *dog*
	“bear” – *bear*	○	“pig” – *pig*
	“dog” – *dog*		“tree” – *tree*
	“bird” – *bird*	○	“sheep” – *sheep*
*	**“make”** – *boy*		“bear” – *bear*
	“flower” – *flower*	○	“book” – *book*
	“bowl” – *bowl*		“flower” – *flower*
	“horse” – *horse*		“bed” – *bed*
	“climbs” – *boy*	○	“crib” – *crib*
*	“pond” – *pond*		“bowl” – *bowl*
	“rooster” – *rooster*	○	“hat” – *hat*
*	**“finally”** – *bed*		**“climbs”** – *boy*
	“sun” – *sun*		“rooster” – *rooster*
	“snake” – *snake*		“horse” – *horse*
*	**“getting”** – *book*	○	“chicken” – *chicken*
*	“umbrella” – *umbrella*		“bird” – *bird*
	“girl” – *girl*	○	“rattle” – *rattle*
	“dad” – *dad*		“snake” – *snake*
*	**“all”** – *cat*	○	**“picnic”** – *basket*
*	**“at”** – *crib*	○	“duck” – *duck*
	“bed” – *bed*		“sun” – *sun*
*	**“these”** – *hen*		“boat” – *boat*
*	**“going”** – *duck*		“girl” – *girl*
	“boat” – *boat*		“dad” – *dad*
*	**“see”** – *boy*	○	**“blanket” –** *grass*

two grammars in an artificial language, 12-month-olds could discriminate new strings from the two grammars, suggesting that statistical learning might play a role also in acquiring rudimentary syntax (the ordering of words, etc.). Meanwhile, recent computational studies show that structural knowledge can be acquired through relatively simple computational mechanisms (Redington, Chater, & Finch, 1998; Mintz, Newport, & Bever, 2002; Solan et al., 2005). For instance, Mintz et al. (2002) showed that grammatical categories of nouns and verbs can be acquired through calculating distributions over words. Specifically, a distributional analysis was developed in which nouns and verbs were successfully categorized based on their co-occurrence patterns with surrounding words. The syntax learning mechanism we applied and integrated in our model is another example of how grammatical structures can be deduced in unsupervised mode (Solan et al., 2005). Putting together, statistical syntax learning is also likely to be a fundamental mechanism in language acquisition.

Interaction between word and syntax learning processes Recent empirical studies have suggested that syntactic cues could play a crucial role in the course of lexical development. Gleitman (1990) demonstrated that learners use evidence from the syntactic structure in which verb occurs to help verb learning. The role of syntactic cues is particularly useful when an extralinguistic scene is insufficient for discovering the meaning of a verb. For instance, there are paired verbs that most often share the same extralinguistic context, such as give and receive, the situation that statistical regularities in cross-situational observation cannot help at all to disentangle the meanings of these two verbs. More recently, Snedeker and Gleitman (2004) showed that the knowledge of known nouns co-occurred in construction with a verb can also significantly improve verb learning. To sum up, behavioral studies have shown that syntactic cues can facilitate lexical learning. The open questions, however, are how the correspondences between syntax and semantics can be learned and what kind of learning mechanisms can support the use of syntax-semantics mappings to bootstrap word learning.

We suggest that since statistical word learning and syntax learning may function simultaneously during human development, the integration of the results from these two concurrent learning processes may lead to the emergence of syntax-semantics mappings, which in turn may generate new statistical regularities. A statistical mechanism acquiring new knowledge from the new regularities can then apply this new knowledge to bootstrap both syntax and lexical learning processes. The model we implemented demonstrates such learning system. If word learners could discover both the grammatical categories of words through a syntax learning process and the semantic categories from a word learning process, they might then be able to build mappings between syntactic and semantic categories. The mappings can then be directly utilized to facilitate word learning because the model is able to estimate semantic properties of a word based on both its syntactic and semantic roles. Our simulation results support this hypothesis by showing that syntactic cues can significantly improve word learning. But is the proposed learning mechanism at all cognitively plausible? A new finding by Saffran and Wilson (2003) demonstrated that 12-month-old infants were able to perform two kinds of statistical computations in the same sound sequence simultaneously when they were exposed to multiword utterances: one is to segment the speech into individual words and the other is to find the orderings of those words. The finding suggests that different learning processes may work concurrently and therefore there may be interactive effects between those learning processes. The present study demonstrates how two key learning processes in language acquisition – syntax learning and word learning – may bootstrap each other through universal syntax-semantics mappings.

Conclusion

This paper presents a computational model to simulate the emergence and acquisition of syntax-semantics mappings and the effects of the mappings on word learning. Using realistic data collected from natural child-parent interaction (picture book reading, etc.), we demonstrate that syntactic cues can be seamlessly integrated in and bootstrap statistical word learning. As first steps towards understanding the learning mechanisms that support syntax-semantics interfaces, the present study focuses on learning a specific kind of words – object names. Nonetheless, the proposed learning mechanism doesn't take advantage of the properties of specific grammatical or semantic categories, but is purely based on statistical regularities within language, across language and extralinguistic contexts, and interactive effects of these two. Therefore, this general learning mechanism has a potential to extend to more grammatical and semantic categories without any significant changes of underlying statistical machinery.

References

Bloom, P. (1994). Possible names: The role of syntax-semantics mappings in the acquisition of nominals. In L. Gleitman & B. Landau (Eds.), *The acquisition of the lexicon* (p. 297-329). Cambridge, MA: MIT Press.

Brown, P. F., Pietra, S., Pietra, V., & Mercer, R. L. (1994). The mathematics of statistical machine translation:parameter estimation. *Computational Linguistics*, *19*(2), 263-311.

Colunga, E., & Smith, L. (2005). A connectionist account of the object-substance distinction. *Psychological Review*, *112*, 347-382.

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

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109-135.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural networks. *Neural Networks*, *17*, 1345-1362.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in pseech to young children. *Cognitive Science*, *26*, 393-424.

Pinker, S. (1989). *Learnability and cognition.* Cambridge, MA: MIT Press.

Plunkett, K., Sinha, C., Miller, M., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? vocabulary growth in children and a connectionist net. *Connection Science*, *4*, 293-312.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*(4), 425-469.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *cognitive science*, *29*(6), 819-865.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month old infants. *Science*, *274*, 1926-1928.

Saffran, J., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273-284.

Siskind, J. M. (1992). *Naive physics, event perception, lexical semantics and language acquisition.* Unpublished doctoral dissertation, MIT.

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

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In Hall & S. Waxman (Eds.), *Weaving a lexicon* (p. 257-294). Cambridge,MA: MIT Press.

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005). Unsupervised learning of natural languages. *in Proc. Natl. Acad. Sci*, *102*(33), 11629-11634.

Tenenbaum, J., & Xu, F. (2000). Word learning as bayesian inference. In L. Gleitman & A. Joshi (Eds.), *Proceeding 22nd annual conference of cognitive science society* (p. 517-522). Mahwah,NJ: ErlBaum.

Yu, C., Ballard, D. H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, *29*(6), 961-1005.

Yu, C., & Smith, L. B. (submitted). Statistical cross-situational learning to build word-to-world mapping.