The Acquisition of Word Meaning through Global Lexical Co-occurrences

PING LI *University of Richmond*
CURT BURGESS AND KEVIN LUND *University of California, Riverside*

1 Introduction

The acquisition of word meaning has been extensively studied for the last thirty years in the field of language acquisition. However, the question of how children acquire word meaning remains highly controversial today. Recently, a number of computational studies have examined the emergence of lexical representations in connectionist networks or similar statistical systems, suggesting that word meaning can be acquired by the computation of statistical regularities inherent in the input data. In particular, Elman (1990, 1998) showed that categories of nouns and verbs, and subcategories of animates versus inanimates (within nouns), and transitives versus intransitives (within verbs), can emerge from the network's computing of the lexical co-occurrence properties in the input. Redington, Chater, and Finch (1998) also demonstrated that the use of distributional properties in large-scale speech corpus allows a statistical system to acquire syntactic categories. These studies are in many ways consistent with the empirical approach of distributional analysis of the linguistic input, as advocated by Maratsos and Chalkley (1980), among many others. They have revealed the power of distributional information in the input, and revived an interest in how the child could analyze the linguistic input to derive accurate representations of the syntax and semantics of words. It is along this line of interest that we seek to understand the acquisition of word meanings in this study.

As a starting point for our approach, we would like to consider two major hypotheses on how children start the process of lexical acquisition: the semantic bootstrapping hypothesis (Grimshaw, 1981; Pinker, 1984, 1987) and the syntactic bootstrapping hypothesis (Landau & Gleitman, 1985; Gleitman, 1990). The semantic bootstrapping hypothesis assumes that

* This project was supported by a Faculty Research Grant from the School of Arts and Sciences, University of Richmond. We would like to thank Ashley Diefendorf and Kimberly Getty for their assistance with extracting parental/caregivers' speech from the CHILDES database.

children can use ontological/perceptual categories of the world, such as agents and events, to guide their initial mappings to the grammatical function of the word, such as subject and verb; in other words, children can make initial lexical classifications on the basis of the contingencies between perceptual categories and linguistic categories. The syntactic bootstrapping hypothesis, in contrast, assumes that the young child can explore the grammatical context in which a word occurs to help them map the initial meanings of the word. The basic assumption is that certain classes of words typically occur in certain syntactic frames and grammatical structures, and the information of these frames and structures provides a useful initial guide to the child as to what the word could mean. Proponents of this hypothesis argue against an unduly emphasis on the perceptual-linguistic contingencies; instead, they argue for the important role of grammatical context.

The emphasis on the grammatical context in which words can occur is consistent with the approach that emphasizes the distributional properties of the input in the acquisition process. However, the role of grammatical context appears dubious to some researchers for the acquisition of word meaning. In particular, Pinker (1994) argued that syntactic bootstrapping allows the learner to identify only kinds of meaning ("frame meaning") and not semantic contents ("root meaning"). This is because syntactically similar words like *tear* and *break* occur in more or less the same syntactic frames but their semantic contents are different. Thus, if the child uses only syntactic bootstrapping, he or she would not be able to distinguish the meanings of words like *tear* and *break.*

This criticism of syntactic bootstrapping, we believe, is valid as long as one considers "grammatical context" in its narrow sense. Syntactic bootstrapping is severely limited in helping the learner to derive the semantic contents of words when it is only concerned with what we call "local co-occurrence" contexts, that is, grammatical or contextual constraints in the immediate environment of the target word: for example, direct objects, complement clauses, and prepositional phrases as post-contexts for verbs. Pinker's criticism would not apply, however, when we start to expand the notion of context from local co-occurrences to "global lexical co-occurrences". We use global lexical co-occurrence to refer to a word's *total experience* in the context of all other words with which it co-occurs. This total experience is measured by counting what words occur before and after a given target word, and how frequently they do so within a large, variable-size window. For example, *tear* tends to co-occur with words like *paper*, *apart*, and *shreds* in a sentence, while *break* tends to co-occur with words like *cup*, *glass*, *pieces*, *window* and *toy*. Thus, global co-occurrence is best seen as a measure of the contexts in which the word was used. This contextual history provides the word with a meaning that is tied to actual usage. Young children may very well explore this type of contextual history at early stages of lexical acquisition.

The notion of global lexical co-occurrence has recently been captured in a high-dimensional semantic space model of language and memory, the Hyperspace Analogue to Language (henceforth the HAL model, see Burgess and Lund, 1997; Lund and Burgess, 1996). In HAL a word is anchored by reference to not only other words immediately preceding or following it, but also words that are further away from it in a variable co-occurrence window, with each slot (occurrence of a word) in the window acting as a constraint dimension to define the function of the target word. The meaning of a word, in this perspective, emerges from the multiple constraints in a high-dimensional space of language use.

	barn	horse	past	raced	the
barn		2	4	3	6
horse					5
past		4		5	3
raced		5			4
the		3	5	4	2

Table 1 Global Co-occurrence Matrix for the Sentence *The horse raced past the barn*. The values in the matrix rows represent co-occurrence values for words which preceded the word (row label). Columns represent co-occurrence values for words following the word (column label). Cells containing zeroes were left empty in this table.

The example in Table 1 illustrates the notion of global lexical co-occurrence more clearly. It shows a matrix using a 5-word moving window for just one sentence (*the horse raced past the barn*). Within this five-word window, co-occurrence values are inversely proportional to the number of words separating a specific pair of words. A word pair separated by a four-word gap, for instance, would gain a co-occurrence strength of 1, while the same pair appearing adjacently would receive an increment of 5. The product of this procedure is an N-by-N matrix, where N is the number of words in the vocabulary being considered. This matrix illustrates how the matrix acquires information about meaning. Consider, for example, the word *barn*. The word *barn* is the last word of the sentence and is preceded by the word *the* twice. The row for *barn* encodes preceding information that co-occurs with *barn*. The occurrence of the word *the* just prior to the word *barn* gets a co-occurrence weight of 5 since there are no intervening items. The first occurrence of *the* in the sentence gets a co-occurrence weight of 1 since there are four intervening words. Adding the 5 and the 1 results in a value of 6 recorded in that cell. A word meaning vector is formed by concatenating the row and column values for the lexical item. Of course, not all vector values/elements (what we call “dimension” later in the paper) contribute equally to the meaning representation, so in our experiments vector length was systematically varied (10, 50 or 100 elements). The most appropriate elements are those that contribute most to the contextual meaning and this is determined by identifying which vector elements have the greatest contextual diversity (see Lund & Burgess, 1996, for details). It is this more complex pattern of co-occurrence, which we refer to as global lexical co-occurrence, that contributes to the richness of meaning.

In this study, we show that global lexical co-occurrences allow the learner to identify not only kinds of meaning -- what Pinker calls “frame meaning”, but also semantic contents – what he calls “root meaning”. Thus, our global co-occurrence approach avoids the drawbacks associated with local co-occurrences approaches such as syntactic bootstrapping. We also provide an account of the factors that could affect the analysis of global co-occurrence information in adult input speech during the child’s acquisition of word meanings.

2 Method

2.1 Materials

Because we are concerned with the information of global lexical co-occurrences in the input speech, we take as our target corpus the parental or caregivers’ speech from the CHILDES database (MacWhinney and Snow, 1985; MacWhinney, 1995). We extracted from the database all the utterances of the parents, caregivers, and experimenters. Although not all of these utterances are child-directed, they form a representative sample of the speech that children are exposed to (e.g., dinner table talks, activities of free plays, and storytelling). The total number of lexical items in this corpus is about 3.8 million words (tokens). This number falls well within the range of the number of items that an average three-year-old is exposed to; according to one estimate, a three-year-old has been exposed to 10-30 million words on the average (Hart & Risley, 1995). This number is also small relative to the regular HAL data based on over 300 million words from the Usenet. Thus, the corpus presents a challenge to our assumption about the role of global lexical co-occurrences in language acquisition. We wanted to see how far the statistical information alone in the linguistic input could allow the child to derive meaningful representations of words, if the child computes global lexical co-occurrences of the words.

2.2 Procedure

First, we ran a standard HAL analysis on the parental/caregivers’ speech corpus in the CHILDES database to see whether the derived vectors yield accurate lexical representations of word meaning. Second, to determine precisely the role of global lexical co-occurrences, we systematically varied three parameters in further analyses, using the analysis of the whole corpus as a basis.

(1) The size of the moving window was varied from 2, 5, to 10 words. We assume that the larger the window size is for tracking co-occurrence, the more global the lexical co-occurrence is under consideration.
(2) The size of the corpus was varied from 10% (0.38 million words), 50% (1.9 million words), to 100% (3.8 million words) of the corpus. The larger the corpus size is, the more accurately the corpus reflects properties of global lexical co-occurrences in the language.
(3) The size of the vector dimensions was varied from 10, 50, to 100. As we discussed earlier, each vector dimension corresponds to a lexical constraint in the definition of a word. The larger the number of vector dimensions is, the more lexical co-occurrences are introduced to mutually constrain the meaning of the word.

A complete factorial design with these three variables (3 x 3 x 3) yielded 27 sets of data for analysis (including the set of the "full-corpus representations" from the first step of analysis, i.e., the representations based on a 10-word moving window, the full corpus, and 100 vector dimensions). The results are discussed below.

3 Results and Discussion

3.1 Accuracy of Word Meaning Representations from Global Lexical Co-occurrence Analyses

Our analysis of the complete parental/caregivers' speech corpus yielded a total of 4,417 word-type representations. There was very little cleaning up or pre-processing of the original text data. Among the 4,417 representations there were about 300 strings that were unintelligible (e.g., *bch*, *tha*), probably due to transcription errors or errors of the speakers (e.g., false starts, made-up words, etc.).

Because it was difficult to do a word-by-word analysis of the over 4,000 representations, to test on the accuracy of these representations, we selected a manageable size of representative vocabulary of 509 words from the 680-word list of the Communicative Development Inventories (CDI) for toddlers (Dale & Fenson, 1996). These 509 words are all content words, mostly nouns, verbs, and adjectives. Other function words in the 680-word list, which are connectives, articles, auxiliaries, etc., were not included in our analysis.

Each word representation consists of a context vector. The length of the vector depends on the number of dimensions used, with each dimension representing the strength of a lexical co-occurrence constraint on a continuous scale from 0 to 1. For example, the word *break* is represented as follows, with 100 dimensions:

Break 0.135 0.000 0.000 0.000 0.006 0.098 0.004 0.131 0.010 0.002 0.026 0.024 0.060 0.004 0.230 0.035 0.007 0.406 0.007 0.005 0.009 0.0114.122 0.094 0.006 0.032 0.050 0.023 0.054 0.098 0.016 0.011 0.083 0.012 0.041 0.294 0.017 0.018 0.003 0.016 0.000 0.016 0.018 0.011 0.012 0.000 0.000 0.028 0.131 0.026 0.009 0.016 0.062 0.012 0.009 0.036 0.015 0.008 0.000 0.006 0.004 0.011 0.005 0.003 0.038 0.005 0.041 0.000 0.047 0 0.000 0.110 0.023 0.003 0.050 0.011 0.015 0.000 0.083 0.019 0.046 0.368 0.010 0.055 0.007 0.040 0.039 0.041 0.000 0.000 0.005 0.008 0 0.025 0.024 0.049 0.012 0.026 0.614 0.088 0.056

We computed for each of the 509 items their ten nearest neighboring words according to the metric similarity of the vectors. We visually inspected these words and their neighbors to determine whether semantically similar words are represented in vectors with similar values. The following is a sample of five representative words (representing action, food, animal, quality, and time) and their ten nearest neighbors (including the target word itself) from the global co-occurrence analysis of the full-corpus representations. These neighbors reflect the contextualized meaning of a word. For example, *break* has as its neighbors *spill*, *hit*, and *fall* -- all contextually related and providing strong meaning constraints given the sample of language experience provided by the model.

Break*: break, spill, hit, fall, throw, hurt, pull, knock, bang, tear*
Breakfast*: breakfast, supper, lunch, dinner, dessert, juice, snacks, snack, yogurt, Xmas*
Duck*: duck, cow, pig, bear, dog, doggie, horse, frog, mouse, rabbit*
Good*: good, nice, great, pretty, beautiful, very, terrible, funny, wonderful, big*
Night*: night, last, yesterday, today, once, before, nursery, school, time, first*

In addition to the analysis of the nearest neighbors, we constructed a connectionist network of a self-organizing feature map (SOFM) to represent each vector of the full-corpus representations on a two-dimensional space so that we can visually inspect the similarities of all the 509 words at once. The SOFM compresses multiple features in a high-dimensional space into a 2-D space, and by way of this compression it extracts an efficient representation of the similarity structure implicit in the data (Kohonen, 1982, 1989; Miikkulainen, 1993; see Li, 1999 for an application of SOFM in language acquisition). In this new representation, similar vectors will activate units in nearby regions of the map. Thus, SOFM allows us to clearly visualize the complex relationship inherent in the vectors encoding high-dimensional, global lexical co-occurrences. Our SOFM network, after 500 epochs of iterative updating of its weight space, was able to represent the 509 words with 89% accuracy (i.e., it was able to clearly distinguish and represent about 453 out of 509 words). Figure 1 presents a snapshot of the 2500-unit map (50 x 50) that captures the 509 words (only a portion, about one-third, of the large map could be shown here).

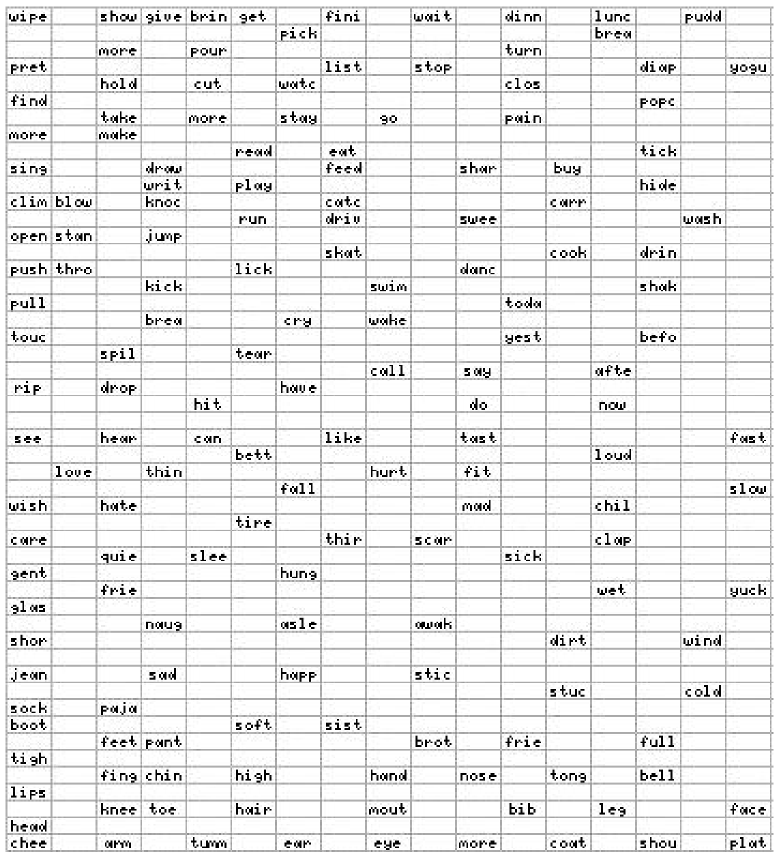

Figure 1 A SOFM representation of 509 words derived from global lexical co-occurrence analyses. Words longer than four letters are truncated.

An examination of this map shows clearly that words with similar meanings are grouped together, while words with different meanings are represented on different regions of the map. For example, on the upper right-hand corner of the map, we can find that words related to eating like *dinner, lunch, breakfast, pudding, yogurt,* and *popcorn* are mapped toward the same region. Another group of body-part words, *lips, head, cheek, feet, finger,* *knee, arm, chin, toe, tummy, hair,* and *ear,* appear toward the lower left-hand corner. Still other groups of similar words occur in various other areas of the map (e.g., *fast* and *slow, sad* and *happy, today* and *yesterday, asleep* and *awake*, and so on). The formation of groups of words on the same region of the map, and the relative distance between these groups, indicate clearly that the vectors provide representations of kinds of meanings. Moreover, the vectors also represent clearly the semantic contents of words. This is shown by two types of distances on the map: the relative distances between groups on the map reflect gross differences between dissimilar words (such as those between food words and body-part words), and the relative distances within groups reflect subtle differences between similar words (such as those between *dinner, lunch* and *breakfast* on the one hand and *pudding, yogurt,* and *popcorn* on the other).

This analysis applies not only to nouns and adjectives, but also to verbs. Meanings of verbs are often considered harder to grasp than meanings of nouns, because verbs are not referentially as clear as nouns, and verbs often occur in the same syntactic contexts or have the same syntactic frames (e.g., *break* vs. *tear*). In fact, much of the "bootstrapping" discussion has focused on how children learn verbs. An examination of Figure 1 reveals that the two kinds of relative distances hold for verbs as well. For example, one can find the following clusters of verbs in nearby regions of the map: actions using hands: *push, pull, throw,* and *touch*; actions using feet, *knock, jump, run,* and *kick;* breaking actions, *drop, spill, break, tear,* and *hit*, daily activities, *write, draw, read, play, eat,* and *feed*; and stative verbs, *wish, love, hate,* and *think*, Within each group of these verbs, individual verbs are represented by different units on the map, and more similar verbs are generally closer to each other than less similar verbs.

3.2 Variables that Affect the Power of Global Co-occurrence Analyses

The above results show that word representations defined by global lexical co-occurrences can successfully capture the range of "root meaning" as well as "frame meaning" of words. Thus, global lexical co-occurrence provides very useful and important information that could be explored in the child's induction of word meanings. Several points should be noted on the analysis in 3.1. First, our analysis was based on the complete sample of 3.8 million words in the corpus. Although this amount of speech, as discussed earlier, does not exceed the amount of what a three-year-old child is exposed to, we wanted to push the limit to see if a further reduction in the input would still allow the derivation of meaningful representations of words. Second, our analysis assumed a moving window of ten words, which means that for any given word, we computed the lexical co-occurrence information of nine other words either before it or after it. To track lexical co-occurrences of ten words at a time may be placing too much demand on the child's memory capacity. Thus, we wanted to experiment with a reduced size of moving window, to see if a smaller window size would still allow accurate representation of word meanings. It might be the case that many of the global lexical co-occurrences can be captured in a smaller window size, and a larger moving window contains redundant information. Third, each of the vector representations discussed in 3.1 had 100 dimensions, which means that we used the lexical co-occurrence information of 100 other words to define any given word-- in other words, we used 100 pieces of mutually constraining information to support the meaning of the word. Although one could well imagine that any word is associated with at least 100 other words in the language and that children are able to learn these associations, we wanted to see if a reduction in the number of associated lexical constraints would still lead to meaningful lexical representations. Section 2.2 outlined the factorial design of our experiment. Table 2 presents the results from the analyses of the new representations.

In Table 2, we used percent agreement as a gross measure of the degree to which each of the 26 new sets of representations matched up with the full-corpus representations discussed in 3.1 (i.e., the representations with a 10-word moving window, the full corpus, and 100 dimensions). We first computed the ten nearest neighboring words for each of the 509 lexical representation in each of the 27 sets. We then compared how many neighboring words are in agreement for each of the 509 targets, between each of the 26 new sets and the “full” set.

	Vector Dimensions								
	10			50			100		
Window Size	10%	50%	100%	10%	50%	100%	10%	50%	100%
2	10	15	18	25	39	46	28	48	57
5	13	19	21	30	51	64	33	62	82
10	17	25	27	30	54	70	33	63	100

Table 2 Percent Agreement of Nearest Neighboring Words between the Full-corpus Representation and Representations of Specific Window Size, Corpus Size, and Vector Dimension

Table 2 indicates clearly that each of the three variables, window size, corpus size, and vector dimensions, is important in the computation of global lexical co-occurrences. If any one of the variables is severely degraded, the analysis does not yield good representations. For example, when the vector dimension was 10, the match between any of the nine sets of representations and the full set was very poor (only 27% for the set with 10-word moving window and 100% corpus). Similarly, when the corpus size was 10%, the match was also poor, even with a larger window size and more vector dimensions. However, the reduction in window size seems to be less dramatic than the reduction in vector dimensions or corpus size. For example, with a window size of 2 words, there was a 57% match if the analysis was done on the full corpus with a 100 vector dimensions. This suggests that for global lexical co-occurrences to work, it is more critical to have a reasonably large number of lexical constraints and a reasonably large amount of speech; by contrast, a relatively small window size can well do the job of tracking lexical co-occurrences, probably because there is redundant information in a larger-than-necessary moving window (cf. an increase from 5 to 10 words in window size did not increase the match for 10% and 50% corpus under 50 and 100 vector dimensions). This result suggests that the learner does not need a very large memory capacity to be able to detect global lexical co-occurrences.

Table 2 also shows that it is possible to get a good match when none of the three variables is severely degenerate. In other words, meaningful representations could occur not only with a smaller moving window, but also with a smaller corpus, or fewer dimensions, if the other variables compensate for the information loss. For example, with a window size of 5 words, corpus size of 100%, and vector dimensionality of 50, the match reached 64%; an increase in window size would lead to a match of 70%, and an increase in vector dimensions would lead to a match of 82%. Progressive increases of one or more of these variables in general lead to better matches, that is, more accurate representations.

4 Conclusion

The role of co-occurrence in language is an old philosophical issue. To what extent temporal or spatial contiguity determines the function of particular linguistic elements? According to Saussure, the function of a given linguistic "entity" (e.g., a word) is defined entirely by reference to the complex relationships that hold between this entity and other entities, much like that the role of a chess piece is determined by its relationship with other pieces on the chessboard (Saussure, 1916). It is further true, argued Saussure, that it does not matter what materials make the entities or the pieces, as long as their structural relationships can be clearly defined. Saussure's arguments on structural properties of linguistic units bear striking similarity to modern-day theories of distributional properties of language, statistical regularities in the input, and global lexical co-occurrences in natural speech.

In this study, we attempted to understand how structural properties inherent in continuous speech, in particular, global lexical co-occurrence information, can help the child to learn word meanings. We assume that local co-occurrences between adjacent linguistic units (as incorporated in the syntactic bootstrapping theory) are relatively less informative about meanings of words (Lund, Burgess, & Audet, 1996). By contrast, an aggregate of multiple co-occurrences in a high-dimensional space of language use provides useful and important information that could be explored by the child in the induction of word meanings. Our analyses of a realistic corpus of parental/caregivers' speech in the CHILDES database lend support to the role of global lexical co-occurrences in language acquisition. In particular, we propose that lexical semantic representations could be derived through the computation of lexical co-occurrences in the linguistic input. Not only kinds of meanings (i.e., "frame meaning", Pinker, 1994) but also semantic contents ("root meaning") can be accurately represented by this method. Our results suggest that young children can learn word meanings by exploiting the considerable amount of contextual information in the input to compute multiple higher-order lexical constraints. This approach relies on a few simple assumptions about what the learner does. One important assumption is that the learner has the ability to track continuous speech with some limitation on working memory, which can be modeled with a weighted moving window of a variable size; another assumption is that the learner is sensitive to lexical co-occurrences during language processing. Such statistical abilities seem to be readily available to the child at a very early age, as recent studies of statistical learning in infants have revealed (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1997).

In conclusion, global lexical co-occurrences provide powerful cues to the child in the acquisition of word meanings. Of course, a myriad of other types of information grounded in the actual learning situation is available to the child as well. What we have demonstrated here is how amazingly the learner can derive accurate semantic representations even without those other visual or perceptual cues in parent-child interactions. When coupled with those other cues, analyses of global lexical co-occurrences are expected to yield even more accurate and faithful representations of word meanings.

References

Burgess, C., & Lund, K. 1997. Modelling parsing constraints with high-dimensional semantic space. *Language and Cognitive Processes,* 12:1-34.

Elman, J. 1990. Finding structure in time. *Cognitive Science,* 14: 179-211.

Elman, J. 1998. Generalization, simple recurrent networks, and the emergence of structure. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society,* eds. M.A. Gernsbacher & S. Derry. Mahwah, NJ: Lawrence Erlbaum.

Dale, P., & Fenson, L. 1996. Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers.* 28: 125-127

Gleitman, L. 1990. The structural sources of verb meaning. *Language Acquisition,* 1: 3-55.

Grimshaw, J. 1981. Form, function, and the language acquisition device. *The logical problem of language acquisition*, eds. C. Baker & J. McCarthy. Cambridge. MA: MIT Press.

Hart, B., & Risley, T. 1995. *Meaningful differences in the everyday experiences of young American children*. Baltimore: Paul Brookes Publishing.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43: 59-69.

Kohonen, T. 1989. *Self-organization and associative memory*. Heidelberg: Springer-Verlag.

Landau, B., & Gleitman, L. 1985. *Language and experience: Evidence from blind children*. Cambridge, MA: Harvard University Press.

Li, P. 1999. Generalization, representation, and recovery in a self-organizing feature-map model of language acquisition. *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society,* eds. M. Hahn & S.C. Stoness. Mahwah, NJ: Lawrence Erlbaum.

Lund, K., & Burgess, C. 1996. Producing high-dimensional semantic space from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*, 203-208.

Lund, K., Burgess, C., & Audet, C. 1996. Dissociating semantic and associative word relationships using high dimensional semantic space. *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society,* eds. M. Hahn & S.C. Stoness. Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B., & Snow, C. 1985. The Child Language Data Exchange System. *Journal of Child Language*, 12**:** 271-296.

MacWhinney, B. 1995. *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.

Maratsos, M., & Chalkley, M. 1980. The internal language of children's syntax: The ontogenesis and representation of syntactic categories, *Children's language* (Vol.2), ed. K. Nelson. New York: Gardner Press.

Miikkulainen, R. 1993. *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. Cambridge, MA: MIT Press.

Pinker, S. 1984. *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pinker, S. 1987. The bootstrapping problem in language acquisition. *Mechanisms of language acquisition*, ed. B. MacWhinney. Hillsdale, NJ: Lawrence Erlbaum.

Pinker, S. 1994. How could a child use verb syntax to learn verb semantics? *Lingua*, 92: 377-410.

Redington, M., & Chater, N., & Finch, S. 1998. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science,* 22: 425-470.

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

Saffran, J., Newport, E., Aslin, R., Tunick, R., & Barrueco, S. 1997. Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science,* 8: 101-105.

Saussure, F. de. 1916. *Cours de linguistique générale*. Paris: Payot. (English translation*: A course in general linguistics*. New York: Philosophical Library; Chinese translation: *Putong Yuyanxue Daolun*. Beijing: Peking University Press).