Initial word learning in children:

A simple neural net model that does not require language-specific innate constraints

P. Thomas Schoenemann

Department of Anthropology
Indiana University

E-mail: toms@indiana.edu

Running headline: Neural Net Model of Word Learning

Acknowledgements:

This work was originally presented at the 1st Workshop on Language Acquisition and Change, City University of Hong Kong, May 2001. This paper has benefited from discussions with Craig Martell, William S.-Y. Wang, Vincent Sarich, and Reina Wong. Any errors that remain are of course my own.

Abstract:

Children rapidly learn the meanings of words even though the utterances they hear typically contain many words with different referents, and the intended referents of particular words is often unclear (Quine's problem of referential indeterminacy). It has been proposed that they possess innate, language-specific mechanisms that constrain their interpretations of utterances in particular ways. An alternative model instead argues that children perform loose mappings between both the words they hear and the possible interpretations for these words, thereby building up increasingly strong associations between salient dimensions of the child's conceptual awareness and utterances that they hear. Siskind {, 1996 #1604} demonstrated this could be modeled computationally. It is further shown that a simple artificial neural net can also model this process: words can come to be associated strongly with specific concepts even though the net is never explicitly trained to associate single words with pre-specified meanings.

Introduction:

The complexity of natural language, combined with the apparent ease with which children learn it, have lead many to argue for the existence of innate, language-specific cognitive mechanisms (Chomsky, 1972, 1980; Bickerton, 1990; Pinker, 1994, among others). One riddle involves the question of how children learn the meanings of words. Given that the specific pattern of sounds (or hand gestures) that is used to symbolize a particular concept will vary from language to language, some sort of learning must be involved. However, as Quine (Quine, 1960) pointed out, there are an indefinite number of possible referents to any given signal a child might hear. If someone points to a dog and says, “Look at the dog!”, how do children know the word “dog” refers to the dog, and not the grass it is standing on, or only the dog’s nose, or even the tip of the finger of the person pointing? How do children learn to associate a particular pattern of sounds or hand gestures with a specific concept (or set of concepts) in the face of this referential indeterminacy?

One possibility is that there are innate word learning biases or constraints that guide the child to make the correct association. A number of such learning constraints have been proposed based on empirical studies (see P. Bloom, 1996 for a review). For example, Markman (1987; 1990) has argued that children assume that words refer to the whole object, not some part of the object (e.g., “dog” refers to the whole animal, not just its nose, or some other part of its body). Markman (1990) has also shown that children tend to assume that words refer to items of the same kind (animals, tools, structures, etc.), not to items that are thematically related (e.g., items relevant to making a meal: foods, pots and pans, household appliances, etc.), in spite of the fact that in sorting tasks they tend to naturally associate things that are thematically related. Markman and Wachtel (1988) argue that children are biased to assume that the meanings of words are mutually exclusive. Clark (1987) argues instead for a weaker bias: children tend to assume words contrast in meaning in some way, but word meanings can overlap. The general idea of language-specific constraints on word learning is supported by a number of current researchers (e.g., Waxman & Booth, 2000; Woodward, 2000; 2001).

However, there are a number of difficulties with this approach. First, the constraints proposed so far explain only limited classes of words (usually count nouns), so that other theories must be posited to explain, e.g., the acquisition of verb meanings (P. Bloom, 1996, 2000; Tomasello & Akhtar, 2000). In fact, it has been argued that some of these strategies would actually hinder the acquisition of non-object noun meanings (Tomasello & Akhtar, 2000). In addition, it is not clear that these biases help explain how children initially begin learning word meanings, since they do not appear to operate from the beginning of word learning (P. Bloom, 2000; Smith, 2000). While it is still possible that they are innate biases that just happen to operate later, this pattern is also consistent with some form of learned strategy concerning the likely references for new words. In any case, Quine's problem is unresolved even with these strategies.

Furthermore, general evolutionary principles suggest that special, language-specific innate structures are inherently unlikely (Schoenemann & Wang, 1996; Schoenemann, 1999). It has long been recognized that adaptive evolutionary change most often occur through the modification of existing features, rather than the evolution of completely new mechanisms (Jacob, 1977; Schoenemann, 1999). In addition, since behavioral flexibility allows an organism to make adaptive changes during an individual's lifetime, without having to wait for genetic modifications (which generally take many generations to accomplish), it is also generally recognized that behavioral change drives genetic change, and not the other way around (Mayr, 1978). Languages must necessarily adapt in each and every generation to the cognitive abilities already present in the population (M.H. Christiansen, 1994). The idea that language rests on the evolutionary modification of prior cognitive functions is consistent with a broad range of research, including the fact that the left hemisphere appears to be biased towards processing of temporal information of all kinds, not just linguistic information (Tzeng & Wang, 1984), as well as the discovery of 'mirror' neurons that respond both to actions perceived in others as well as the same actions performed by self (Rizzolatti & Arbib, 1998; Arbib, 2001). This general view is held by a number of models of language evolution (e.g., Wang, 1991; Deacon, 1997; Schoenemann, in press).

With respect to word learning, therefore, an evolutionary perspective would lead us to expect <u>a priori</u> that pre-existing general-purpose mechanisms are the most likely explanation. It is of course <u>possible</u> that special language-specific mechanisms evolved, but Occam's razor requires us to thoroughly investigate explanations that do not posit special structures.

In fact, alternative models have been proposed for word learning in children that deny the necessity of language-specific innate constraints. A number of researchers have pointed out that the process of word learning is fundamentally a social process, involving the interaction of a language incompetent child with a series of language competent individuals (Nelson, 1988; 1991; Ahktar & Tomasello, 2000). Steels and Kaplan (2000) have shown that modeling a social language game between visually grounded robotic agents can lead to the development of a shared lexicon. Ahktar and Tomasello (2000) and Bloom (2000) specifically tie word learning to children's understanding of other's intentions, and review evidence that this plays a key role in helping children map the correct concepts to particular words. Smith (2000) reviews research suggesting that various word acquisition constraints are the <u>result</u> of a general associative developmental process, rather than <u>a priori</u> language-specific innate constraints. The idea that associative processes might explain how to get around Quine's problem of referential indeterminacy has been suggested by several authors (Pinker, 1989; Savage-Rumbaugh & Rumbaugh, 1993; Fisher, Hall, Rakowitz, & Gleitman, 1994; Siskind, 1996). The proposal is that children may be performing loose mappings between words they hear and possible interpretations for these words. If there is something consistent across all observed uses of a given word, then general associative mechanisms might be expected to gradually strengthen word-meaning pairings. What is consistent across uses might be sets of basic perceptual features (e.g., 'round shape', 'red color'), types of actions (e.g., 'run', 'hit'), emotional states of others (e.g., 'happy', 'angry'), and so forth. Basically, anything the child can conceive of at the time the word(s) are spoken would be potential referents for these words. There are likely to be innate mechanisms that bias children to package certain sets of perceptual features in certain ways, such that it is easier for them to form specific concepts. However, these mechanisms would not need to be specific to language (even though language itself allows for the creation of new concepts). From an evolutionary perspective, we must assume that a reasonably complex conceptual world existed prior to the development of language (Schoenemann, 1999). Thus, there must be some conceptual awareness and thought independent of language, and this is what language learning critically depends upon. Also note that this model is not inconsistent with the idea that social context and understanding of intentions are crucial to word learning. These can be conceived of as biases that help focus the child's attention on certain aspects of their perceptual input, and allow them to further limit the possible interpretations a child might make regarding a given set of words.

The benefit of the argument that word meanings emerge as a consequence of their repeated associations with particular concepts is that – in contrast to the constraint theories mentioned above - it explains the acquisition of any type of word (not just particular types of nouns) as long as the child already has the requisite concept(s). It is important to remember that even constraint theories implicitly accept that the child must already have concepts for words (P. Bloom, 2000).The pattern-matching hypothesis is both simpler and potentially more powerful, and requires a minimum of pre-existing language-specific cognitive structures. No particular assumptions need to be made about the extension of meanings to given words. If adults use words for which the meanings overlap in some odd but consistent way, children will learn these odd meanings simply by paying attention to the patterns of co-occurring associations. Thus constraints like

Clark's (1987)"principle of contrast" could be explained simply as a learned recognition that words tend to have meanings that differ in some way from others, without having to posit this constraint as existing innately from the outset of word learning. Furthermore, it explains how children could learn words at the very earliest stages of language acquisition, prior to any understanding of syntax that can provide cues to word meaning (Gleitman, 1990). It should be stressed that the argument is not that children do not use information provided by syntax or from various constraints or biases of word meaning, but rather that these are learned strategies, not innate constraints.

In addition, this model would explain how apes are able to learn to associate arbitrary symbols with concepts (Savage-Rumbaugh & Rumbaugh, 1993). Thus we potentially have a way of explaining the apparent evolutionary continuity of cognitive ability with a common mechanism: exactly what an evolutionary perspective predicts.

Not only does this model solve the referential indeterminancy problem without language-specific innate constraints, but Siskind (1996) has shown that the pattern-matching hypothesis can be modeled computationally. Siskind's model handles certain kinds of noise as well as homonymy (assuming the variant word meanings are completely distinct) by allowing multiple interpretations to coexist (e.g., the model allows for a "$band_1$" and "$band_2$" to have different meanings). It also impressively demonstrates increasingly fast word learning as more and more words are learned (eventually learning words after only one or two exposures) just like human children. It cannot, however, deal with polysemy, nor the presence of idioms or metaphors, nor with the problem of fragmentary utterances. It also assumes children can segment speech into words. However, his model is fundamentally designed to show proof-of-concept of the idea of cross-situational inference for solving word-meaning mapping problems. It postulates a number of formal inference rules about what to do with certain kinds of word-meaning associations, as well as a set of tables mapping word symbols to necessary and possible conceptual symbols. Siskind himself specifically states that he does not claim children actually employ the particular algorithm presented in his paper (p. 40).

The goal of the present paper is to show that some key aspects of this idea - that word meanings emerge as a consequence of their repeated associations with particular concepts -can be modeled using simple artificial neural nets. Specifically, neural nets can learn to associate a specific word with a specific concept even though they are never explicitly trained to associate any single word with a single concept. Because neural nets are specifically modeled on general principles of how the brain actually works at the neuronal level, showing that neural nets can learn this way provides additional support for the argument that word meanings can be learned without innate language-specific mechanisms. Furthermore, the model presented here deals with certain problems that are not handled by Siskind's (1996) model: polysemy, idioms and metaphors, and fragmentary utterances.

Neural nets:

Neural nets are composed of simple processing units (analogous to neurons), which are massively interconnected with other units in a simple hierarchical organization. Thus, neural nets are simulations of networks of neurons. Just as in real brains, neural nets have specialized input and output structures. Incoming information from the input nodes is manipulated in various ways as it is passed through the network, ultimately causing certain patterns of activation in the output structures (output nodes). Figure 1 shows a simple neural net architecture. The actual net used for the simulation discussed in this paper has more nodes – see below – but for visual simplicity a smaller network is depicted here. A neural network is trained by adjusting the internal connections between nodes such that given input patterns will produce specific output patterns. This training is accomplished, typically, by presenting the net with a series of correct input and output pairs. For each pair of associations, the net's internal connections are adjusted slightly to favor a specific output given a specific input. Neural nets have been shown to be able to learn some quite impressive tasks (see Churchland & Sejnowski, 1994 for several examples). In particular, they have been increasingly used to model aspects of language learning for which it had long been assumed language-specific innate constraints were necessary, such as in learning the rules regarding verb use from naturalistic input (Seidenberg, 1997), and learning to segment speech into words on the basis of subtle, noisy probabilistic cues inherent in the input (Morten H. Christiansen, Allen, & Seidenberg, 1998).

[figure 1 about here]

However, a simple neural net model is clearly not appropriate to the problem of word learning because the net typically knows unambiguously that a given input (i.e., a word) should produce a given specific output (a concept). This would beg the question we are trying to answer: how does the net (or child) learn the meanings of words without ever being able to know exactly the intended meaning of any utterance? The modification in the present simulation is to train the net using inherently ambiguous learning sets. The net will be trained to associate a group of words at a time with larger sets of possible concepts in each learning trial. The set of possible concepts will always include the intended concepts (correct given the input words), but will also include a number of distractor words which are meant to represent other conceivably correct concepts (conceivably correct given the context in which the words may have been heard). At the end of the training process, the net is tested to see how strongly single words are associated with specific meanings, even after never having been trained on specific associations. This model is broadly similar to that proposed by McClelland and Rumelhart (1986) with the biggest difference being that ambiguous learning sets are used. McClelland and Rumelhart (1986) showed that a simple net (without a hidden layer) could map words to meanings such that specific words activated specific meanings, and vice versa. In other words, setting the meaning to a particular value would activate the word for this meaning. It could also activate the correct meaning from only a partial activation of the word as input. However, it was trained to associate exactly one meaning with exactly one word, which does not solve the referential indeterminacy problem. The question is whether neural nets are robust in the face of ambiguous learning sets.

The model used here is deliberately very simple, designed simply to show proof-of-concept. Possible elaborations to make it more realistic will be discussed below. The specific network used consisted of three layers, as in Figure 1. The input layer has 20 nodes, each of which stand for a different specific word that the net "hears". The output layer also consists of 20 nodes, each of which stand for different specific concepts. The net has a hidden layer of 40 nodes, which are connected to each of the input and output nodes. For example, node 4 in the input layer might stand for the word "man". A key assumption in this simulation is that words can in some way be segmented from the speech stream (this is clearly a simplification, but we will argue below that this model might be elaborated to show how segmentation of words and learning of word meanings might both be accomplished by the same associative mechanism). Exactly how words might be represented in real brains is not dealt with here, except to say that if real brains can distinguish words, they must represent them differently in some way. Presumably different words are represented as different networks of activation among neurons in real brains, but our short-cut will be to assume that these different network activation states can be signified by single nodes. Note that it would be possible to have complex network activations (simulating different words) be the inputs for our neural net model, but that would not materially change the basic point demonstrated here. When a particular word is presented to the net, the input node corresponding to that word is set to 1. Otherwise, the node is set to 0.

Similarly, the output nodes represent single concepts. The assumption here is that the net already "knows" certain concepts, and the problem is simply one of mapping possible words to possible concepts. Furthermore, we are assuming for the moment that the net has at least one concept for each word. Node 17 of the net might represent the concept

[man], for example. Just as in the case of inputs, we will not worry about exactly how concepts are represented in real brains (though presumably they are also distinct networks of activation). Whether a concept is signified by a single node or a network of activation is not crucial for the present argument. The values for both the hidden and output nodes can be set to any value between 0 and 1. A value of 1 for an output node means that that particular concept is maximally activated by the inputs. There is nothing constraining more than one output node from being maximally activated at the same time.

The network is trained to associate sets of inputs (representing words) with sets of outputs (representing concepts). For the present simulation, the net is trained to weakly associate three words, chosen at random, with six possible concepts in each round. The three correct concepts are always included among the concepts the net is trained on, but the other three concepts are randomly chosen (with respect to the correct concepts). This is meant to model the idea that children have some conceptual awareness of an array of possible meanings for the words they hear, but that they do not know exactly what any given word means. The other randomly chosen associations are meant to model other possible features of the child's conceptual awareness that are present at the time the words are spoken. They need not be thought of as concrete objects specifically present in front of the child, but rather anything that the child could potentially conceive of at that moment (e.g., 'fast', 'sleeping', 'anger'). An example training set might have, for example:

"man" "woman" "kiss"

as inputs, and:

[ball] [woman] [run] [kiss] [man] [push]

as outputs. The weights of the net are then adjusted slightly to favor each of these outputs equally, when presented with these inputs.

The net is trained using the standard back-propagation algorithm (Rumelhart, Hinton, & Williams, 1986). Tlearn neural net software package was used to run these simulations (available free on the internet: http://crl.ucsd.edu/innate/tlearn.html; Plunkett & Elman, 1997). The rate of learning was set at 2. For simplicity, weights (describing how activation in one node affects activation in a connected node) are randomly set at the start. This means the net starts with a random association between inputs (words) and outputs (meanings).

The net is tested after every 50 rounds by being presented, one at a time, each word as input. This means that the input node corresponding to a particular word is set to 1, activation is fed forward through the net, and the activation of each output node is logged. The goal will be to see if, after having been trained to associate <u>sets</u> of words with <u>larger sets</u> of possible concepts (only half of which will actually be correct in any given trial), the net activates the correct <u>single</u> concept for a given <u>single</u> word as input (for example, does the output node corresponding to [kiss] show the greatest activation when only the input node corresponding to the word“kiss” is activated?). Thus the goal is to see if the net can learn individual associations in the face of ambiguous input, without ever having been trained explicitly on any given association.

Results:

Figure 2 shows a summary of the activation levels of the output nodes as a function of the number of training sets presented to the net. The output nodes are divided into target meaning vs. non-target meaning nodes. A target meaning node is the node that represents the appropriate concept for a given input word. The non-target meaning nodes are all other nodes for that word. At the beginning, there is no distinction between target and non-target meaning node activation levels. This means that when a particular word is input into the net, no particular concept node is activated more than others. After training on 50 sets of word-concept pairings, the target meaning output nodes show on average slightly greater activation than the other non-target meanings. However, there is quite a large degree of variability across words, with some showing very little activation at all. However, by the time the net has been trained on only 150 word-meaning sets, there is a clear separation between target and non-target activation levels. The net reaches essentially perfect activation after training on ~1000 word-meaning sets. That is, upon activating a particular word in input, the appropriate concept node is highly activated, and all other concept nodes are hovering around 0.1.

[figure 2 about here]

Discussion:

The net learned to associate single words with single concepts, yet it rapidly did so without ever having been explicitly trained to associate single words with single concepts. This is perhaps not surprising. Connections between particular word-concept pairs are continually strengthened, while associations between a word and other concepts are not. Nevertheless, this demonstrates that the problem of referential indeterminacy can easily be overcome if there is something consistent about the conceptual state of the child during repeated exposures to a given word, even if there are other words and concepts are present at the same time.

The particular model presented above does make a number of assumptions. As pointed out, it assumes that children already have concepts for which words can map onto. Again, however, this is also implicitly assumed by constraint models, and is what we should expect from an evolutionary perspective in any case (Schoenemann, 1999).

In addition, the model assumes that inappropriate meanings that happen to be conceptualized by the child at the time the word is presented are random, over successive trials, with respect to the intended meaning. This is clearly not strictly the case. A given child might conceivably happen to be exposed to a word (e.g., “car”) when looking out of a window, but not in other instances. This child, under the present model, would initially tend to assume the word “car” means something more like ‘wheeled moving object visible outside my bedroom window’. However, eventually this association will necessarily be refined, because it is only a matter of time before the child also hears the word in other circumstances. Thus, this model explains why children often do use words in peculiar ways, both more restrictive ways (e.g., the car example above; L. Bloom, 1973) and more general ways. For example Bowerman (1978) reported that her daughter used the word “moon” to refer to many other objects of similar shape (e.g., a dishwasher dial). This apparently anomalous, individualistic understanding of word meanings early in word acquisition has been widely reported (reviewed in Nelson, 1988). Thus, though in this model potentially distracting associations are specifically randomized, thus leading to relatively quick mappings of words to meanings, quirky non-random associations would merely lead to exactly the kind of idiosyncratic understanding and use of words that appears to be seen in children (with real neural nets).

Another assumption this model makes is that the child can segment the speech chain into words. This assumption is of course shared by constraint theories. It is possible that the present model could be extended to explain how word segmentation might be accomplished at the same time that word-meaning mapping is being done. This might be done by training the net on associations between sequences of phonemes and concepts. In a classic paper on nets that learn sequences of inputs (as opposed to fixed sets of inputs as in the current model), Elman (1990) showed that a particular kind of recurrent network architecture could allow a net to learn word boundaries simply by being trained to predict the next letter in a stream of letters constituting a sentence. If such a recurrent net were trained to predict word meanings based on sequences of phonemes, it seems likely that it would at first have no clear association of particular sets of phonemes with particular meanings. However, over time the net would probably pass through a period in which supersets of phonemes would be associated loosely with sets of meanings. For example, the sound sequences indicated by the words 'soft' and 'fur' might not associate individually with their respective concepts, but if both are activated consecutively , they might activate both concepts equally. This remains to be tested, but holds promise for explaining how words are segmented from streams of phonemes.

One benefit of this extension of the model is that it has the potential for explaining how idioms, metaphors, and fragmentary utterances can be learned (which is a limitation of Siskind's 1996 model). Children initially tend to treat formulaic expressions like idioms as single lexical items (MacWhinney, 1978; Elman, 1990). A temporal extension of the model discussed here would potentially have the ability to associate strings of variable length (Elman, 1990). Thus idioms, metaphors and fragmentary utterances would simply be treated as long words. Assuming the child can have a conceptual understanding of what the idiom is meant to refer to, they should be learned in exactly the same manner as described here for words.

The contention here is not that children lack innate predispositions, but rather that we do not need to postulate <u>language-specific</u> innate mechanisms to explain word learning. Nor is it claimed that children never use syntactical information, or that they don't ever use constraints as rules of thumb for guessing the meanings of words. The argument is that that these are simply helpful "cranes" (Smith, 2000) that develop later as a result of earlier word learning. The model presented here shows how word learning can get started prior to any of these strategies. It is also not disputed that a child's understanding of other's intentions plays a crucial role in word learning. It is simply noted that some non-trivial level of referential indeterminacy still exists even if the child can eliminate a number of possible referents by being able to make educated guesses about what speakers are thinking. The associative model implemented here solves Quine's problem, and at the same time would work in conjunction with a child's understanding of intentionality.

The work of Steels and colleagues (Steels & Kaplan, 2000) represents an alternative explanation. Their model involves robot agents who learn words through an interactive game in which meanings and words are both created at the same time. Agents gradually converge on meanings for words as they adjust their understanding of each other's utterances on the basis of feedback. While it is clear that human children do get some feedback (even if not explicit "negative evidence") through various forms of implicit cueing when miscommunication occurs, it is not clear how many words are learned via an interactive game akin to that played by the robots. In any case, the model suggested here would work independent of – and in addition to – other methods of word learning.

It is possible to model word learning as a process of emerging word-meaning associations accomplished by accumulating loose mappings between perceived words and possible conceptual interpretations. The fact that a simple artificial neural net model can do this suggests that real neural nets have at least this capability (though the subtlety and complexity of the associations are likely to be many orders of magnitude greater). Such an explanation makes the riddle of language evolution that much more comprehensible, because it relies on pre-existing associative and conceptual abilities rather than de novo, innate, language-specific cognitive mechanisms.

Literature cited:

Ahktar, N., & Tomasello, M. (2000). The social nature of words and word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (New York: Oxford University Press.

Arbib, M. A. (2001). Co-evolution of human consciousness and language. *Ann N Y Acad Sci* **929**, 195-220.

Bickerton, D. (1990). *Language & Species*. Chicago: University of Chicago Press.

Bloom, L. (1973). *One Word at a Time: The Use of Single Word Utterances before Syntax*. The Hague: Mouton.

Bloom, P. (1996). Overview: Controversies in language acquisition. In P. Bloom (Ed.), *Language Acquisition: Core Readings* (Cambridge, Massachusetts: MIT press.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, Massachusetts: The Mit Press.

Bowerman, M. (1978). The acquisition of word meaning. In N. Waterson & C. Snow (Eds.), *Development of Communication: Social and Pragmatic Factors in Language Acquisition* (New York: Wiley.

Chomsky, N. (1972). *Language and Mind* (Enlarged ed.). New York: Harcourt Brace Jovanovich, Inc.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences* **3**, 1-61.

Christiansen, M. H. (1994). *Infinite Languages, Finite Minds: Connectionism, Learning and Linguistic Structure*. Unpublished Unpublished PhD dissertation, University of Edinburgh, Edinburgh, Scotland.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language & Cognitive Processes* **13**, 221-268.

Churchland, P. S., & Sejnowski, T. J. (1994). *The Computational Brain*. Cambridge, Massachusetts: MIT Press.

Clark, E. V. (1987). The principle of constrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (Hillsdale, NJ: Erlbaum.

Deacon, T. W. (1997). *The symbolic species : the co-evolution of language and the brain*. New York: W.W. Norton.

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

Fisher, C., Hall, G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua* **92**, 333-375.

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

Jacob, F. (1977). Evolution and tinkering. *Science* **196**, 1161-1166.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Reseach in Child Development* **43**.

Markman, E. M. (1987). How children constrain the possible meanings of words. In U. Neisser (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization* Cambridge: Cambridge University Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science* **14**, 57-77.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology* **20**, 121-157.

Mayr, E. (1978). Evolution. *Sci. Am.* **239**, 47-55.

McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 2: Psychological and Biological Models). Cambridge, MA: MIT Press.

Nelson, K. (1988). Constraints on word meaning. *Cognitive Development* **3**, 221-246.

Nelson, K. (1991). Concepts and meaning in language development. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch & M. Studdert-Kennedy (Eds.), *Biological and Behavioral Determinants of Language Development* (Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: Harper Collins Publishers, Inc.

Plunkett, K., & Elman, J. L. (1997). *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*. Cambridge, MA: MIT Press.

Quine, W. V. O. (1960). *Word and Object*. Cambridge, Massachusetts: MIT Press.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends Neurosci* **21**, 188-194.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1: Foundations). Cambridge, MA: MIT Press.

Savage-Rumbaugh, E. S., & Rumbaugh, D. M. (1993). The emergence of language. In K. R. Gibson & T. Ingold (Eds.), *Tools, Language and Cognition in Human Evolution* (Cambridge: Cambridge University Press.

Schoenemann, P. T. (1999). Syntax as an emergent characteristic of the evolution of semantic complexity. *Minds and Machines* **9**, 309-346.

Schoenemann, P. T. (in press). Conceptual complexity and the brain: Understanding language origins. In J. W. Minett & W. S.-Y. Wang (Eds.), *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics* (Hong Kong: University of Hong Kong Press.

Schoenemann, P. T., & Wang, W. S.-Y. (1996). Evolutionary principles and the emergence of syntax. *Behavioral and Brain Sciences* **19**, 646-647.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science* **275**, 1599-1603.

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

Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (New York: Oxford University Press.

Steels, L., & Kaplan, F. (2000). Bootstrapping grounded word semantics. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: formal and computational models* (Cambridge: Cambridge University Press.

Tomasello, M., & Akhtar, N. (2000). Five questions for any theory of word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (New York: Oxford University Press.

Tzeng, O. J. L., & Wang, W. S.-Y. (1984). Search for a common neurocognitive mechanism for language and movements. *American Journal of Physiology (Regulatory Integrative Comparative Physiology)* **246 (15)**, R904-R911.

Wang, W. S.-Y. (1991). *Explorations in Language*. Taipei, Taiwan: Pyramid Press.

Waxman, S. R., & Booth, A. E. (2000). Principles that are invoked in the acquisition of words, but not facts. *Cognition* **77**, B33-B43.

Waxman, S. R., & Booth, A. E. (2001). On the insufficiency of evidence for a domain-general account of word learning. *Cognition* **78**, 277-279.

Woodward, A. L. (2000). Constraining the problem space in early word learning. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a word learner: A debate on lexical acquisition* (New York: Oxford University Press.

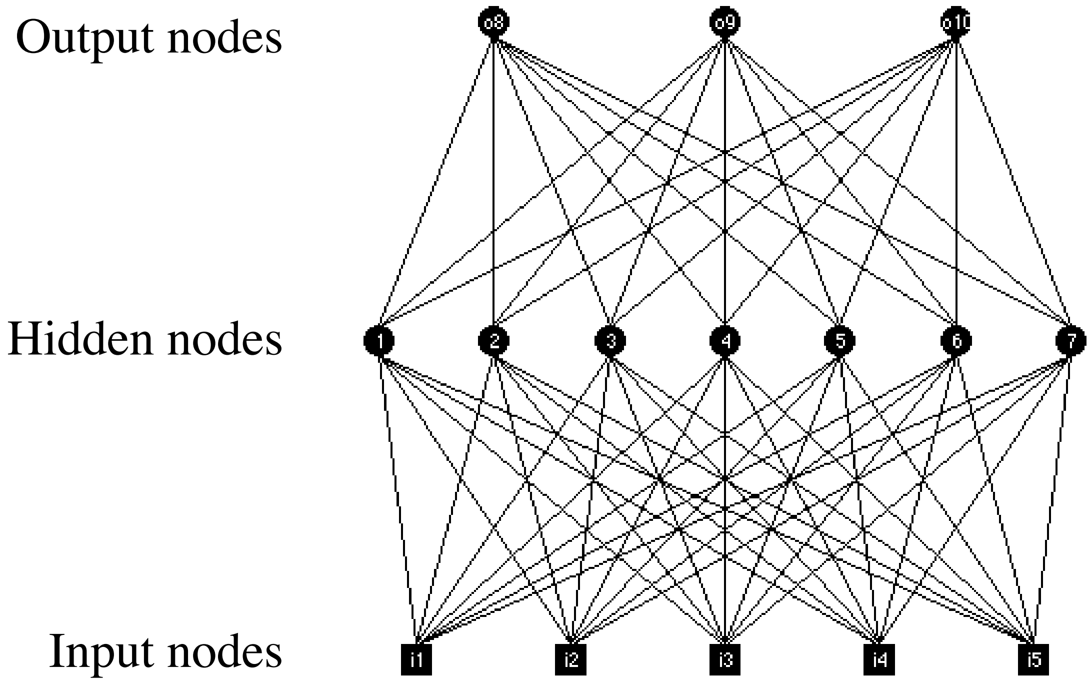

Figure 1: Example architecture of a simple feed-forward net with 5 input nodes, 7 hidden nodes, and 3 output nodes. Each node is connected with (and therefore helps set the value of) every node in the subsequent layer.

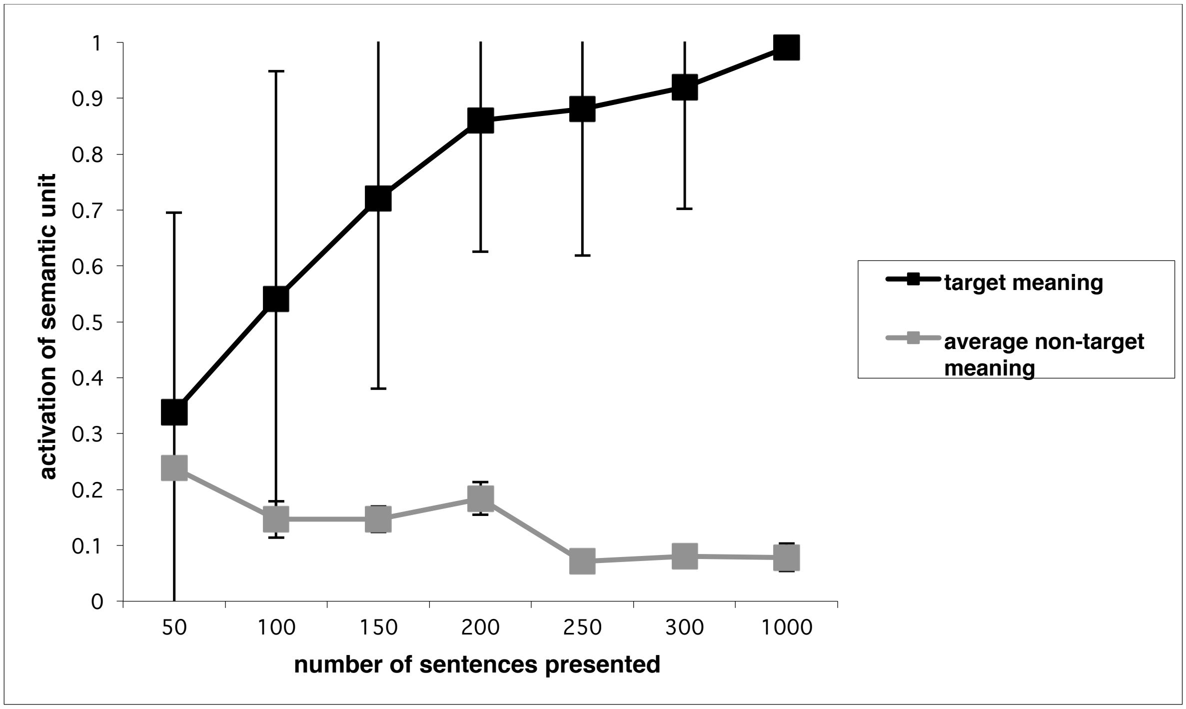

Figure 2: Average output activation levels for target vs. non-target meanings after activation of a single word on input as a function of the number of sentences presented. Error bars represent the standard deviation for the 20 input words.