CHAPTER EIGHT

IN SEARCH OF MEANING: THE ACQUISITION OF SEMANTIC STRUCTURES AND MORPHOLOGICAL SYSTEMS [*]

Li, Ping
University of Richmond

1. Introduction

The representation of language has been traditionally considered as a construction out of basic structural building blocks in the form of symbols and rules. This approach in general looks at linguistic representations statically. A contrasting approach, in the spirit of recent developments in connectionist networks and statistical learning, attempts to capture linguistic representations dynamically. It considers linguistic representations as emergent properties that evolve out of a continuously developing and adapting system. A shortcut to the understanding of this approach might come from the following example. Structured, rule-like representations in a connectionist network can emerge in much the same way as a hexagonal structure emerges from the honeycomb: every honeybee packs a given amount of honey to the honeycomb from multiple directions, but no honeybee has a grand planning for the hexagonal structure (Bates, 1984).[1] In this paper, I provide such an account of the emergence of semantic representations, in connection with morphological learning in language acquisition.

Lexical semantics and its acquisition by children has been a hotly debated issue in the last thirty years. Until recently, most researchers in this domain have thought that there is a fixed set of conceptual and semantic properties associated with each lexical item, and that the child's task is to acquire the necessary conceptual frameworks and the semantic properties. Recent computational models of language processing suggest that lexical semantics may be emergent properties, in particular, that lexical categories can be acquired by the computation of statistical regularities inherent in the input data. These models are in many ways consistent with the empirical approach of distributional analysis (dating back to structural linguistics; Saussure, 1916) that emphasizes the child's ability to analyze the linguistic input (e.g., Maratsos & Chalkley, 1980). They can be classified roughly into two categories. First, proposals from statistical analyses of large-scale text corpora indicate that lexical-semantic representations may emerge from multiple contextual and lexical co-occurrence constraints in a high-dimensional space. Second, connectionist (or neural network) models indicate that lexical-semantic structures can emerge from statistical learning of form-form and form-meaning mappings. In what follows, I will briefly consider both types of models, but the focus of this chapter will be on the second.[2]

High-dimensional semantic space and lexical representation

There have been a number of proposals that high-dimensional semantic space can provide accurate and faithful representations of lexical semantics through multiple contextual or lexical co-occurrence constraints in large text corpora. Two models have emerged most prominently in the last few years: the HAL model (Hyperspace Analogue to Language), advocated by Burgess and Lund (1997), and Lund and Burgess (1996); and the LSA model (Latent Semantic Analysis), developed by Landauer and Dumais (1997), and Landauer, Foltz, and Laham (1998). These two models are highly compatible with each other, although the specific methods used are different. In the following, I will focus on the HAL model as our research has linked this model specifically to children's acquisition of lexical semantics.

According to HAL, the meaning and function of a given word are determined by lexical co-occurrence constraints in a high-dimensional input space, that is, by what items may precede a word and what may follow it, and how often they do so. HAL focuses on global rather than local lexical co-occurrences: A word is anchored with reference not only to other words immediately preceding or following it, but also to 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 meaning and function of the target word.

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.

	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 that 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. See Burgess and Lund (1997). Reproduced with authors' permission.

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 or elements contribute equally to the meaning representation. 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 is referred to as global lexical co-occurrence that contributes to the richness of meaning. In short, global lexical co-occurrence is a measure of a word's *total experience* in the context of other words. The meanings of a word, in this perspective, emerge from multiple constraints in a high-dimensional space of language use.

Although models like HAL are not originally designed for language acquisition, they have significant implications for the acquisition of word meanings. Redington, Chater, and Finch (1998) used a similar method as HAL to capture lexical syntactic categories in child language. In another study, Li, Burgess, and Lund (2000) applied the HAL method to the analysis of parental speech in the CHILDES database (Child Language Database Exchange System; see MacWhinney 2000, for a description of the database). We analyzed 3.8 million words from the speeches of parents and caregivers addressed to children, and found that a reasonable size of speech corpus (e.g., 3.8 million words) with a reasonable amount of co-occurrence constraints (e.g., 50 co-occurrence elements) can yield accurate and faithful semantic representations of English words.[3] 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). In short, global lexical co-occurrences can provide useful and powerful cues to the young child in the acquisition of word meanings.

Emergent semantic structures in connectionist networks

A second set of models, consistent and complimentary with the computational approach discussed above, are the connectionist models of language processing and language learning. Recent years have seen rapidly developing interests in the application of connectionist models to the study of language acquisition (see Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Klahr & MacWhinney, 1998 for overview). This interest dates back to Rumelhart and McClelland's (1986) connectionist model of the learning of the English past tense and the debates thereafter (MacWhinney & Leinbach, 1991; Pinker, 1991, 1999; Pinker & Prince, 1988; Plunkett & Marchman, 1991, 1993; Seidenberg, 1997). Connectionist models rely on the use of a large number of connected micro-processing units (called 'nodes' or 'neurons') that activate in parallel and adjust weights of connections between one another through learning and processing. Two key assumptions of these networks have to do with (a) representation – knowledge is represented as patterns of activation distributed across the processing units, and (b) learning – new knowledge is formed through the adaptation of the strengths or weights of connections that hold among the processing units. These assumptions differ from traditional cognitive assumptions about knowledge representation that involves discrete symbolic representations of concepts, categories, and grammatical rules. With regard to language acquisition, advocates of connectionism argue that linguistic representations (of the lexicon, morphology, and grammar) are "emergent properties" due to the interaction of the processing units with the linguistic environment in the form-meaning mapping process. This view contrasts with the traditional psycholinguistic approaches that emphasize the mental representation of rules and the innateness of grammatical and semantic categories.

Connectionist principles of distributed representation, weight adjustment, and nonlinear learning provide a mechanistic account of how syntactic and semantic structures can emerge out of learning. For example, Elman (1990, 1995) showed that a simple recurrent network is able to derive internal representations of semantic as well as syntactic categories in a task of predicting the next word in the sentence. Lexical categories such as nouns and verbs, animate and inanimate, and human and animals emerge clearly in the network's hidden-unit representations after the network has been trained to map the current word in the input stream to the next word. What the network does is similar to the process of detecting lexical co-occurrence constraints in the input (as does the HAL model). Note that both Elman's network and the HAL method can be likened to the "distributional analysis" technique used by structural linguistics (Bensch, 1991), although structural linguistics did not have today's powerful statistical machinery and computational tools.

Li (1993) and Li and MacWhinney (1996) discussed more explicitly how a connectionist network can develop internal representations of semantic structures. Using the acquisition of the English reversive prefix *un-* as an example, they examined the role of cryptotypes in determining overgeneralization patterns, competition principles, and plasticity of learning. In three simulations, they showed that structured semantic representations can emerge from connectionist learning: the network formed internal representations of semantic categories that correspond to Whorf's cryptotypes, on the basis of learning limited semantic features of verbs and morphological classes. More important, the network produced overgeneralization errors similar to those reported by Bowerman (1982), Clark, Carpenter, and Deutsch (1995), and those observed in the CHILDES database, indicating that emergent semantic structures underlie patterns of productivity in child language.

In this paper, I take a more in-depth look at the issue of the acquisition of semantic structure along with the acquisition of morphological systems. I will focus on the second set of models discussed above, the connectionist approach to language acquisition, summarizing results from our studies. Our results indicate how semantic structures can emerge from the learning of probabilistic associations that hold between lexical items and morphological markers. Moreover, understanding gained from connectionist semantic acquisition directly helps us to identify psycholinguistic and computational mechanisms of generalization and overgeneralization in language acquisition.

2. Cryptotype as an Emergent Category and as a Trigger for Overgeneralization

Whorf's cryptotype

In one of the classic papers of early cognitive linguistics, Whorf (1956) presented the following puzzle. In English, the reversive prefix *un-* can be used productively with many verbs to indicate the reversal of an action, for example, as in *uncoil*, *uncover*, *undress*, *unfasten*, *unfold*, *unlock*, *untie*, or *untangle* (the meaning of reversal can also be expressed by other prefixes such as *dis-* or *de-*). However, many seemingly parallel forms are not allowed, such as **unbury*, **unfill*, **ungrip*, **unhang*, **unpress*, **unspill*, **unsqueeze*, or **untighten*. Why is *un-* prefixation allowed with some verbs but not others? None of the standard categories of Latin grammar can be used as a basis for a rule to tell us when we can use *un-* and when we cannot.

Whorf's puzzle was deeper than this simple discrepancy. He reminded us that *un-* is a productive device in English morphology, and that despite the difficulties that linguists have in characterizing its use, native speakers do have an intuitive feel for which verbs can be prefixed with *un-* and which cannot. He presented the following thought experiment: if a new verb, *flimmick* is coined to mean "to tie a tin can to something", then native speakers are willing to accept the sentence, "He unflimmicked the dog" as expressing the reversal of the "flimmicking" action; if *flimmick* means "to take apart", then they will not accept "He unflimmicked the puzzle" as describing the act of putting a puzzle back together. The constrained productivity of *un-* prompted Whorf that there was some underlying or covert semantic category, a cryptotype, that governs the productive use of *un-*. According to Whorf, cryptotypes only make their presence known by the restrictions they place on the possible combinations of overt forms. When the overt prefix *un-* is combined with the overt verb *tie*, there is a covert cryptotype that licenses the combination *untie*. This same cryptotype also blocks a combination such as **unmove*. To Whorf, the deep puzzle was that while the use of the prefix *un-* is productive, the cryptotype that governs its productivity is unclear: "we have no single word in the language which can give us a proper clue to its meaning or into which we can compress this meaning; hence the meaning is subtle, intangible, as is typical of cryptotypic meanings."

Although cryptotype seemed puzzling, Whorf did propose that there was "a covering, enclosing, and surface-attaching meaning" (Whorf, 1956:71) that could be the basis of the cryptotype for *un-*. Whorf was correct in noting that verbs that take *un-* usually have one or more of the covering, enclosing, or surface-attaching meaning. But it is not clear whether we should view this cryptotype as a single unit, three separate meanings, or a cluster of related meanings. Nor is it clear whether these notions of attachment and covering fully exhaust the subcomponents of the cryptotype. Subsequent analyses have suggested certain additional components not initially considered by Whorf. For example, Marchand (1969) and Clark *et al.* (1995) argue that verbs that license *un-* all involve a change of state, usually expressing a transitive action. This transitive action typically reaches a terminal point in time (encoded by a *telic* verb; Comrie, 1976), or some end state or result (an *accomplishment* verb; Vendler, 1967). When the meaning of a verb does not involve a change of state or does not indicate telicity or accomplishment, the verb cannot take *un-,* thus the ill-formedness of verbs like **unswim*, **unplay*, and **unsnore*.

Cryptotype and morphological productivity in child language

Whorf's discussion shows clearly how cryptotype is important to the use of *un-* in the adult language. Bowerman was the first to point out that the notion of cryptotype might also play an important role in children's acquisition of *un-*.

According to Bowerman (1982, 1983, 1988), children's acquisition of *un-* tends to follow a U-shaped pattern, a pattern that children display in other areas of morphological acquisition as well, such as the acquisition of the English past tense (Brown 1973; Kuczaj 1977). Children initially produce *un-* verbs in appropriate contexts, treating *un-* and its base verb as an unanalyzed whole. This initial stage of rote control is analogous to the child's saying *went* without realizing that it is the past-tense form of *go*. Productivity of *un-* comes at the next stage, when children realize that *un-* is independent of the verb to indicate the reversal of an action.

The next stage in the acquisition of *un-* begins at around age 3. At this stage, children start to produce overgeneralizations in spontaneous speech such as **unarrange*, **unbreak*, **unblow*, **unbury*, **unget*, **unhang*, **unhate*, **unopen*, **unpress*, **unspill*, **unsqueeze*, or **untake* (Bowerman, 1982). These overgeneralizations have also been observed in Clark *et al.* (1995) in both experimental and naturalistic data with children from ages 3 to 5, for example, **unbend*, **unbury*, **uncrush*, **ungrow*, **unstick*, and **unsqueeze*. Similar examples can also be found in the CHILDES database, such as **unblow*, **unbuild*, **uncatch*, **uncuff*, **unhand*, **unlight*, **unpull*, **unstick*, and **unzipper* (see Li & MacWhinney, 1996, for a more complete list of examples of overgeneralization errors). During this period, children also make certain 'overmarking' errors. For example, the child might say **unopen* and really only means to say *open*, or *unloosen* to mean *loosen*. In such cases, the base forms *open* and *loosen* have a reversive meaning that triggers the attachment of the prefix, even when the action of the base meaning is not actually being reversed. These errors are analogous to redundant past-tense marking as in **camed* and redundant plural marking as in **feets* (Brown, 1973). As children grow older, overgeneralization or overmarking errors gradually disappear.

A traditional explanation of the U-shaped pattern in children's morphological acquisition goes like this: initially they rely on rote learning, then they develop a general rule and apply it productively (and overgeneralize it), and finally they recover from productive errors (this is much like what has been argued for the acquisition of the English past tense). For productivity to take place at the second stage, Bowerman correctly pointed out that cryptotype plays an important role. But how could the child extract the cryptotype and use it as a basis for morphological generalization or recovery, when the cryptotype is intangible even to linguists like Whorf? (see Whorf's comments on the subtle and intangible nature of the cryptotype as discussed earlier).

A connectionist account of cryptotype and its acquisition

A connectionist perspective provides us with a natural way of capturing Whorf's insights of cryptotype as well as its acquisition in a formal mechanism. In our view, there can be several 'mini-cryptotypes' that work together as interactive 'gangs' (McClelland and Rumelhart, 1981). For example, "enclosing" verbs, such as *coil*, *curl*, *fold*, *reel*, *roll*, *screw*, *twist*, and *wind,* all seem to share a meaning of circular movement. Similarly, "attaching" verbs, such as *clasp*, *fasten*, *hook*, *link*, *plug*, and *tie*, all involve hand movement. Other verbs such as *bind*, *buckle*, *fasten*, *latch*, *leash*, *lock*, *strap*, *tie*, and *zip* form a mini-cryptotype that share a "binding" or "locking" meaning. Still another cluster of verbs such as *cover*, *dress*, *mask*, *pack*, *veil*, and *wrap* forms the "covering" mini-cryptotype. These mini-cryptotypes or mini-gangs interact collaboratively to support the formation of the larger cryptotype that licenses the use of *un-*, in terms of summed activation, as illustrated as in Figure 1.

The mini-gangs collaborate rather than compete because their members are closely related by the overlap of semantic features. For example, the verb *screw* in *unscrew* may be viewed as having both a meaning of circular movement and a meaning of binding or locking; *zip* in *unzip* may be viewed as sharing both the "binding/locking" meaning and the "covering" meaning, and both *screw* and *zip* involve hand movements. Moreover, a feature may also vary in the strength with which it is represented in different verbs. For example, circular movement is an essential part of the meaning of the verb *screw*, but less so for *wrap* (one can wrap a small ball with a soft tissue paper without turning around either the object or the wrapping paper). These properties of feature overlap and degraded featural composition lend themselves naturally to properties of connectionist models. Distributed patterns, weighted connections, nonlinear learning as embodied in connectionist networks seem to be ideal for handling the elusiveness and gradience of these semantic structures.

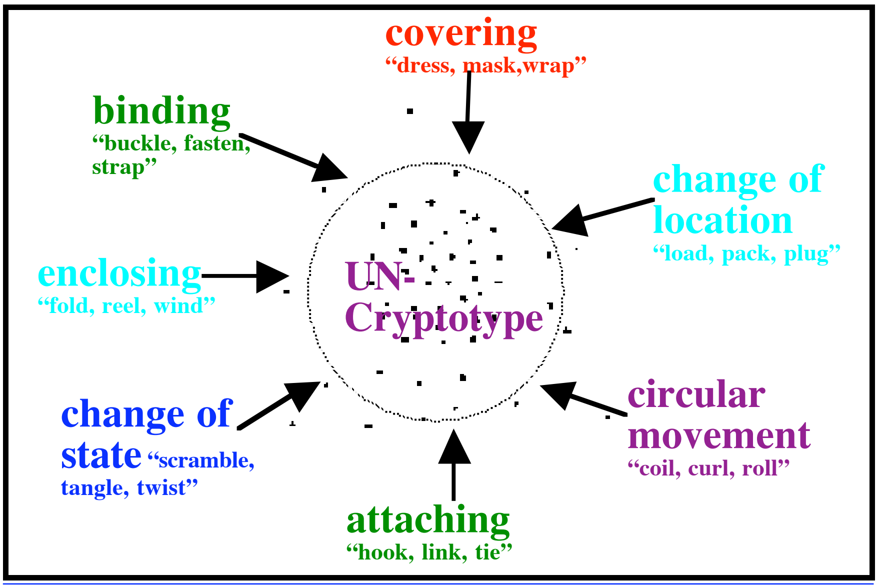


Figure 1. *Multiple features support the formation of the* un- *cryptotype. Arrows represent the feature-to-category connections; the weights or strengths of connections are omitted. Dots in the center of the circle represent words that fit the core of the category, while dots near the border of the circle represent borderline cases.*

In the last few years, our laboratory has carried out connectionist simulations to study the issue of semantic structure and overgeneralization, using the acquisition of *un-* as an example. In the following sections, I will discuss two major models in this endeavor. The first model uses a standard feed-forward network to simulate the acquisition of cryptotypes and prefixes. The second model uses a self-organizing neural network, which has also been recently applied to the acquisition of semantic and grammatical structures in children and in bilingualism. Readers who are interested in the technical details of these models should consult Li and MacWhinney (1996), Li and Farkas (2002), Li (2003), and Li, Farkas, and MacWhinney (2004).

3. A feed-forward network that learns to map semantic features of verbs to prefixes

Method

Connectionist networks that use the back-propagation algorithm (henceforth 'backpropagation networks') are perhaps the most popular class of networks and are most widely applied in studies dealing with language. A standard back-propagation network consists of three layers of processing units (Rumelhart, Hinton, & Williams, 1986). In this type of network, information is first encoded at the input layer, then it funnels through the hidden layer, where internal representation is formed, and finally results are produced at the output layer (hence the nickname of 'feed-forward networks'). Each layer consists of different units, representing different states/processes of information processing (from input to output). Learning in this case is a function of adjusting the weights of the connections between units across the layers. The adjustment is done through the back-propagation algorithm, according to which the network discovers a discrepancy between its actual output and the desired output, and then an error signal is propagated back through the system, so that weights are adjusted in a way such that the next time the same input will lead to an output that matches more closely to the desired output (for technical details of the algorithm, see Hinton, Rumelhart, & McClelland, 1986).

In our simulation, we used 160 verbs as input to our network. They consisted of 49 verbs that can take the prefix *un*-, 19 verbs that can take the competing prefix *dis*- (see Li & MacWhinney, 1996, for the rationale of including *dis*- verbs, and the competition between *un*- and *dis*- in both child and adult languages), and 92 randomly selected verbs that can take neither prefix (henceforth 'zero verbs'). Each verb was represented by a semantic pattern (a vector) that consists of 20 semantic features. These features were selected in an attempt to capture basic linguistic and functional properties inherent in the semantic range of these verbs. In order to objectively determine the values of each semantic feature, we presented 15 native English speakers with the 160 verbs along with the 20 semantic features, and asked them to judge the semantic relevance of each feature to each verb. A feature-by-verb relevance matrix was derived for each subject, and the final input vectors were derived by averaging the matrices from all subjects. A hierarchical clustering analysis on these vectors attests to the validity of our method, as distance metrics in this analysis reflected the similarities and differences between words.

The task of the network was to take the semantic vectors of English verbs as input, and map them onto different prefixation patterns in the output: *un*-, *dis*-, and zero. Figure 2 shows the network architecture and examples.

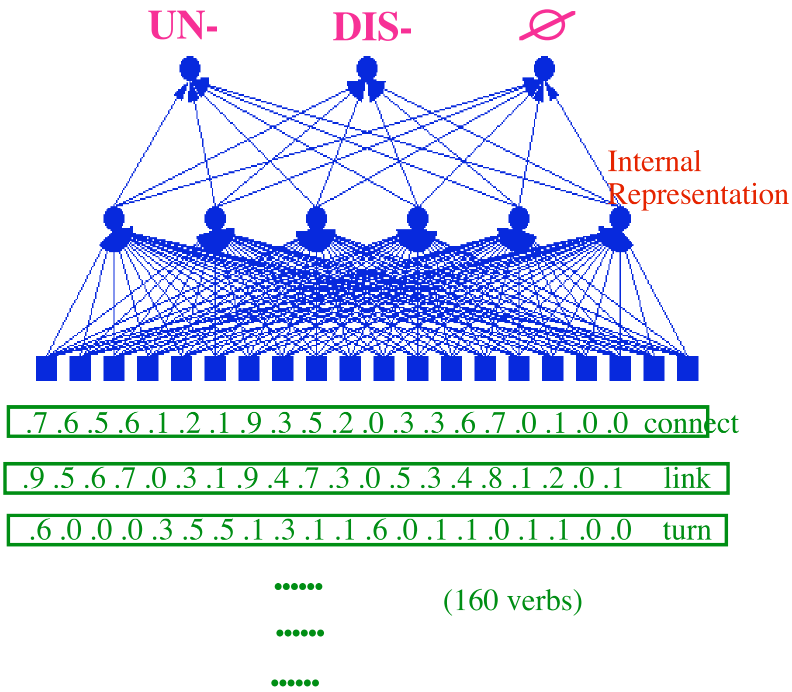


Figure 2. *The feed-forward network that learns to map semantic features of verbs to prefixation patterns (*un-, dis-, *Ø).*

Results and Discussion

Connectionist networks are dynamic systems that explore the regularities in the input-output mapping processes through the adjustment of connection weights (to and from the hidden units) and the activation of the hidden units. To analyze how our network developed internal representations, we used the hierarchical cluster analysis to probe into the activation of the hidden units at various points in time during the network's learning (see Elman, 1990, for an application of this method). Figure 3 (in Appendix) presents such an analysis at three time points, the early (3a), intermediate (3b), and later stages of learning (3c), respectively. Focusing here on the verbs that share the enclosing-rotating meaning (most of which can be prefixed with *un-*), we can see how the network developed structured semantic representations. These cluster trees indicate that early on with little learning, there was not much meaningful structure in the data, and thus, the enclosing-rotating verbs were scattered all over the cluster tree. Gradually as learning progressed, these verbs started to form smaller groups at several levels. Finally when learning reached a stable situation, they were all grouped under one cluster.

These snapshots provide a picture of the developmental trajectories in the network's integration of semantic structures during the meaning-form mapping process. They illustrate how a mini-cryptotype, such as the enclosing-rotating category, which supports the use of *un-*, can emerge from learning the mapping of verb semantics to prefixation.

In the studies reported by Li & MacWhinney (1996), we used an incremental learning procedure, in which the network took in the input gradually, verb by verb. Learning with this procedure also lent us insights into the formation of cryptotype in the network. Figure 4 shows a cluster tree of the network's hidden-unit representation when the network learned 50 verbs. In this graph, we can observe two general clusters: one for the *un-* verbs, and the other for the zero verbs – verbs that cannot be prefixed with *un-* or *dis-*. Our interpretation of these clusters is that the network acquired a distinct representation for the *un-* verbs by identifying the mini-cryptotypes inherent in these verbs. For example, most of the verbs in the *un-* cluster share the cryptotypic meaning of binding or locking: *bind*, *chain*, *fasten*, *hitch*, *hook*, *latch*, etc. However, not all mini-cryptotypes were identified at this time, and they emerged at different stages as discussed above. Figure 4 also shows, for example, that the network had not yet developed a clear representation for the enclosing verbs: the verbs *ravel* and *coil* were correctly categorized into the *un-* cluster, but the verb *roll* was incorrectly treated as a zero verb.

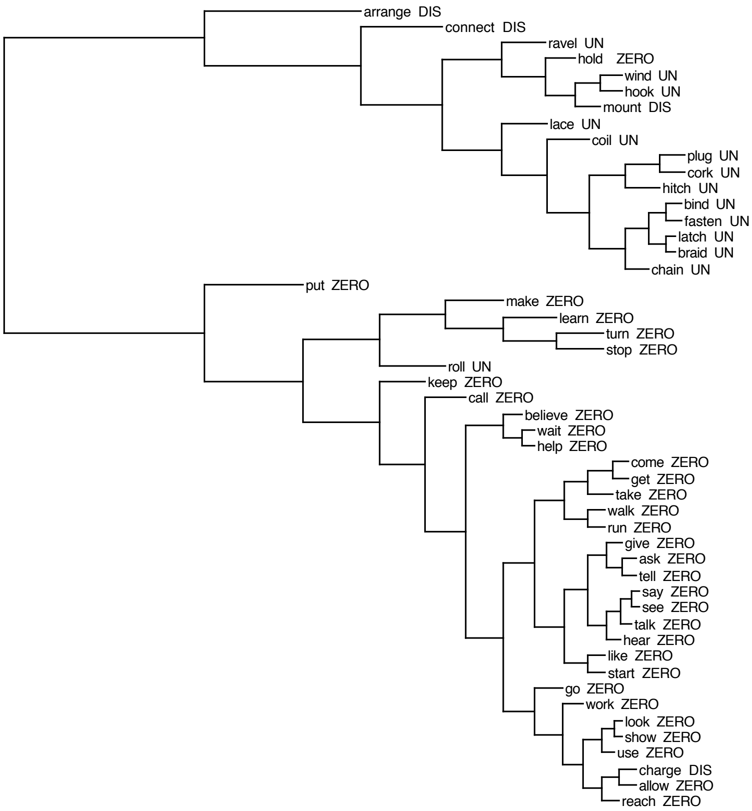


Figure 4. *A hierarchical cluster analysis of the network's hidden-unit representations after the network has learned 50 verbs.*

Note that our network received no discrete label of the semantic category associated with *un*-, nor was there a single categorical feature that tells which verb should take which prefix (hence Whorf's problem). All that the network received was semantic featural information distributed over different input patterns. Over time, however, the network was able to identify the regularities that hold between distributed semantic patterns and patterns of prefixation, and developed a structured representation in the mapping process. The structured representations in the network thus emerged as a function of its learning of the association between form and meaning, not as a property that was given *ad hoc* to the network by the modeler.

The emerging representations also clearly capture Whorf's notion of cryptotype. The meaning of a cryptotype constitutes a complex semantic network, in which verbs differ from one another with respect to (a) how many features each verb contains, (b) how strongly each feature is represented in the verb, and (c) how strongly features overlap with one another within a verb (all true with the input to our network). It is these complex relationships that give rise to the notion of cryptotypes.

The emergence of cryptotype representations in our network can be viewed as a replacement for the traditional analytic frameworks of categories and rules (Lakoff, 1987; MacWhinney, 1989). In this perspective, children's learning of *un-* is not simply the learning of a symbolic rule for the use of the prefix with a class of verbs (given that it is not even clear what the rule is), but the accumulation of the connection strength that holds between a particular prefix and a set of semantic features distributed across verbs. The learner groups together those verbs that share the largest number of features and take the same prefix. Over time, the verbs gradually form clustered patterns, with respect to both meaning and prefixation pattern. This learning process can best be described as a statistical procedure in which the child implicitly tallies and registers the frequencies of co-occurrence of semantic features, lexical items, and morphological devices.

Bowerman (1982, 1983) suggested that there are two possible roles for cryptotypes to influence the learning of *un-*. (a) "Recovery via cryptotype": cryptotypes help the child to overcome overgeneralizations made at an earlier stage, if these overgeneralizations involve verbs that fall outside the cryptotype, such as **uncome*, **unhate*, and **untake* (Bowerman, 1982); (b) "Generalization via cryptotype": cryptotypes trigger productivity and leads to overgeneralizations. This occurs because, once children have identified the cryptotype, they will overgeneralize *un-* to all verbs that fit the cryptotype, irrespective of whether the adult language actually allows *un-* with these verbs. Our simulation results provide support for the second role of cryptotype in inducing overgeneralizations that fall within the realm of the cryptotype. Figure 4 showed how the network included *hold* and *mount* in the *un-* category. These verbs were included apparently because of their semantic similarity with members of the cryptotype, most of which can take *un-* (e.g., *bind*, *chain*, *fasten*, *hitch*, *hook*, *latch*). Examining the output patterns of *hold* and *mount* in the network, we found that *un-* was overgeneralized on these verbs. Similar overgeneralization errors produced by the network included **unbury*, **uncapture*, **unfill*, **unfreeze*, **ungrip*, **unhold*, **unloosen*, **unmelt*, **unpeel*, **unplant*, **unpress*, **unsplit*, **unsqueeze*, **unstrip*, **untack*, and **untighten*, most of which fit the cryptotype meaning. Our network produced few simulated errors that were flagrant violations of the cryptotype meaning, such as forms like **uncome* reported by Bowerman (1982), thus our results provide no direct evidence for the first role of cryptotype as hypothesized by Bowerman. In our simulations, overgeneralizations occurred typically after the network had developed structured cryptotype representation, indicating that cryptotype served as a trigger for morphological overgeneralization.

These results match up well with available empirical data. For example, one child in Bowerman's study produced errors such as **uncapture*, **unpeel*, **unpress*, **unsplit*, **unsqueeze*, and **untighten*, similar to those in our network. The overgeneralizations that the child produced all fell within the cryptotype, and her acquisition of *un-* as a reversive prefix went hand in hand with her discovery of the cryptotype meanings of the verbs. In Clark *et al.*'s (1995) naturalistic data, the child's innovative uses of *un-* also respected the cryptotype from the beginning. Clark *et al.* noted that the child's use of *un-* matched the semantic characteristics of the cryptotype even when the conventional meanings of the verb in the adult language did not: **unbuild* was used to describe the action of detaching lego-blocks, **undisappear* was used to describe the releasing of the child's thumbs from inside his fists.[4] Thus, once the learner (child and network alike) formed a structured representation that corresponds to the cryptotype for *un-*, the representation guides the learner's behavior in productive morphological use.

In subsequent simulations, our network also displayed a limited amount of recovery from overgeneralization errors. Typically, recovery was best when the network had developed only partial or unstable semantic structures at relatively early stages of learning, and it became increasingly difficult when a fixed structure had emerged at later stages of learning (Li & MacWhinney, 1996). This is because the back-propagation learning algorithm proceeds in such a way that early on, the network's weight configurations are not fully committed and more flexible to change, but later on as the network learns more and more words, it settles on a more stable weight space that makes adjustment difficult if not impossible (see Elman, 1993: 91-93 for a detailed discussion of how the learning algorithm determines weight adjustment over time). This situation does not seem to match with what we know about child language: most children eventually recover from all overgeneralization errors, no matter how late. Even tough plasticity might be particularly characteristic of early learning (Spitzer, 1999), older children and adults are still able to change, adapt, and recover from errors, unlike the network studied here (Bownds, 1999). This mismatch, along with other considerations discussed below, prompted us to study another type of connectionist model, the self-organizing neural network, to account for lexical acquisition.

4. A self-organizing network that learns to map semantic features to prefixes

Although most previous connectionist model of language acquisition have relied on the use of feed-forward networks with back-propagation, researchers have started to see their limitations. In addition to its limited ability to recover from overgeneralizations, there were two other major limitations to the network that we used. First, our network, like most previous models, received semantic input features selected on the basis of linguistic analyses on the part of the modeler. Input representation in this way is subject to the criticism that the network worked (e.g., displayed cryptotype representation) precisely because of the use of certain semantic features (cf. Lachter & Bever, 1988). To overcome potential limitations associated with this problem, in the new simulations we used semantic representations that are based on analyses of global lexical co-occurrences from a large text corpus (see previous discussion of HAL, and Method below). Second, back-propagation relies on a gradient-descent weight adjustment process to reduce the error between desired and actual outputs, but this type of adjustment seems unrealistic for child language learning. According to the well-known "no negative evidence" argument (Baker, 1979; Bowerman, 1988; Pinker, 1989), children do not receive constant feedback about what is incorrect in their speech, or receive the kind of error corrections on a word-by-word basis as provided to a back-propagation network. Thus, back-propagation networks would seem to be poor candidates as models of language acquisition on grounds of their psychological or biological plausibility. Considerations of these problems lead us to self-organizing neural networks. Self-organizing networks are biologically more plausible because one could conceive of the human cerebral cortex as essentially a self-organizing map (or multiple maps) that compresses information on a two-dimensional space (Spitzer, 1999). They are computationally more relevant because one could argue that child language acquisition in the natural setting (especially organization and reorganization of the lexicon) is largely a self-organizing process that proceeds without explicit teaching (MacWhinney, 1998, 2001).

Method

In contrast to standard feed-forward networks, self-organizing networks use unsupervised learning that requires no presence of a supervisor or an explicit teacher; learning is achieved entirely by the system's self-organization in response to the input (Kohonen, 1982, 1989, 1995). Self-organization in these networks typically occurs in a two-dimensional map (self-organizing map), where each unit is a location on the map that can uniquely represent one or several input patterns. At the beginning of learning, an input pattern randomly activates one of the many units on the map. Once a unit becomes active in response to a given input, the weights to the unit and its neighboring units are adjusted so that they become more similar to the input and will therefore respond to the same or similar inputs more strongly the next time. In this way, the network gradually develops concentrated areas of units on the map (like the activity "bubbles") that respond to particular inputs. This process continues until all the inputs can elicit specific response patterns in the network. As a result of this self-organizing process, the statistical structures implicit in the multi-dimensional space of the input are represented in the two-dimensional space of the map.

Here we used the hierarchical feature map model of Miikkulainen (1993, 1997) in our simulations, because it combines multiple self-organizing maps in a single network. In this model, there is a semantic map that processes semantic information of the words, and there is a phonological map that processes phonological information of words (for more details of the application of the model, see Li, 1999, 2003). The two maps are connected via associative links trained by Hebbian learning, a well-established biologically plausible learning principle, according to which the associative strength between two units (semantic and phonological) is increased if the units are both active at the same time (Hebb, 1949).

The same set of verbs described in §3 was used as the input, but they were represented differently from the way they were represented in the previous simulations. The semantics of these words were encoded as patterns of global lexical co-occurrence constraints (Burgess & Lund, 1997; see §1), rather than patterns of semantic features selected on the basis of our own linguistic analyses. Each verb was represented as a pattern of 100 units, and the values of these units reflected the degree of a lexical co-occurrence constraint (on a continuous scale from 0 to 1). We also derived a phonological representation for each verb and the prefixes *un-* and *dis-*, according to MacWhinney and Leinbach (1991). In this representation scheme, each verb was encoded by 168 units in a syllabic template to represent the combinatorial constraints of phonology (see also Li & MacWhinney, 2002, for details).

Upon training of the network, a phonological representation of the verb was presented to the network, and simultaneously, the semantic representation of the same verb was also presented to the network. By way of self-organization, the network formed an activity on the phonological map in response to the phonological input, and an activity on the semantic map in response to the semantic input. Depending on whether the verb is prefixable with *un-* or *dis-*, the phonological representation of *un-* or *dis-* may also be co-activated with the phonological and the semantic representations of the verb stem. At the same time, through Hebbian learning the network formed associations between the two maps for all the active units that responded to the input. The network's task was to create new representations in the corresponding maps for all the input words and to be able to map the semantic properties of a verb to its phonological shape and its morphological pattern.

Results and Discussion

In our network, the self-organizing process extracted and compressed the high-dimensional information from the HAL semantic vectors and expressed the semantic similarities on the two-dimensional space as concentrated patterns of activity. Figure 5 presents a snapshot of the network's self-organization of 120 verbs after the network was trained for 600 epochs.

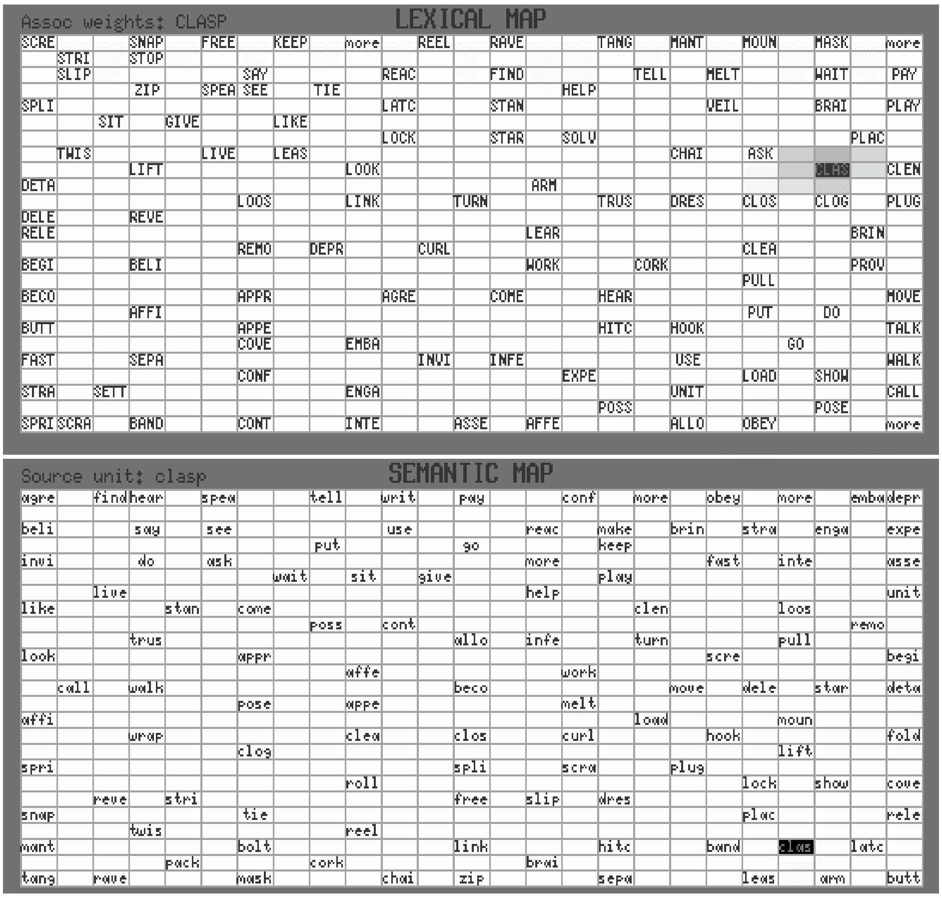


Figure 5. *A self-organizing map model that shows the organization of 120 verbs after the network was trained on these verbs for 600 epochs. The upper panel is the lexical phonological map (indicated by capital letters), and the lower panel the semantic map (indicated by lower-case letters). Words longer than four letters are truncated.*

An examination of the semantic map shows that the network has clearly developed forms of representation that correspond to cryptotype categories. Earlier we suggested that a connectionist model provides a formal mechanism to capture Whorf's notion of cryptotype, in that there can be several mini-cryptotypes that work collaboratively as interactive gangs to support the formation of the larger cryptotype. The idea of 'mini-cryptotype' is reflected most clearly in the emerging structure of the self-organizing map. Our network, without the use of *ad hoc* semantic features, formed clear mini-cryptotypes by mapping similar words onto nearby regions of the map. For example, towards the lower right-hand corner, verbs like *lock*, *clasp*, *latch*, *lease*, and *button* are mapped to the same region of the map, and these verbs all share the "binding/locking" meaning. A similar mini-cryptotype also occurs towards the lower left-hand corner, including verbs like *snap*, *mantle*, *tangle*, *ravel*, *twist*, *tie*, and *bolt*. Still a third mini-cryptotype can be found in the upper left-hand corner, including *hear*, *say*, *speak*, *see*, and *tell*, verbs of perceptions and audition. Finally, one can observe that *embark*, *engage*, *integrate*, *assemble*, and *unite* are being mapped toward the upper right-hand corner of the map, which all seem to share the "connecting" or "putting-together" meaning (interestingly, these are the verbs that can take the prefix *dis*-). Of course, the network's representation at this point is still incomplete, as self-organization is moving from diffuse to more focused patterns of activity; for example, the verb *show*, which shares similarity with none of the above mini-cryptotypes, is grouped with the binding/locking verbs. What is crucial, however, is that these mini-cryptotypes form the semantic basis for the larger cryptotype of *un*- verbs. As shown in Figure 5, the network has mapped most verbs in the cryptotype to the bottom layer of the semantic map, and these are the verbs that can take the prefix *un*-.

Moreover, our network was not only able to capture the elusive cryptotype by way of self-organization, but also able to generalize on the basis of its representation of the cryptotype. During testing of the network's productive ability, overgeneralization occurred with 50% of the testing words. For example, the network produced overgeneralization errors that match up with empirical data and our previous simulation results (see §3), including **unbreak*, **uncapture*, **unconnect*, **unfreeze*, **ungrip*, **unpeel*, **unplant*, **unpress*, **unspill*, **unstick*, **untighten*, etc. These overgeneralizations were based both on the network's representation of the meaning of verbs and on the associative connections that the network formed through Hebbian learning in the semantics-phonology mapping process. Again, like in our previous simulations, most of these overgeneralizations involve verbs that fall within the *un*- cryptotype. Thus, the results here are again consistent with the "generalization via cryptotype" hypothesis, that is, the representation of cryptotype leads to overly general uses of *un*- (see also discussion of the *clench* example below) rather than the narrowing down of its uses (as predicted by the "recovery via cryptotype" hypothesis).

One of the advantages of the self-organizing model is its ability to simulate comprehension and production through associative connections. The associative connections formed via Hebbian learning provide the basis for the production of overgeneralization errors. For example, the semantic properties of *tighten* and *clench* are similar and they were mapped onto nearby regions of the semantic map. During learning, the semantics of *clench* and *unclench* were co-activated, and the phonology of *clench*, *unclench*, and *un-* were also co-activated. When the semantics and the phonology of these items were associated through Hebbian learning, the network linked the semantics of *tighten* with the prefix *un-* because of *clench*, even though the network learned only the association for *un-clench* and not *un-tighten* (when *tighten* was withheld from training at an earlier stage). This associative process of correlating semantic features, lexical forms, and morphological devices simulates the process of learning and generalization in children's productive speech, and shows that overgeneralizations can naturally result from the semantic structure in the lexical representations (which in turn is a result of self-organization), and from the associative learning of semantics and phonology.

In §3 we discussed the failure of a feed-forward network in recovering from overgeneralization errors. We attributed that failure to the gradient-descent error-adjustment process used in the back-propagation algorithm. In self-organizing networks, recovery is a function of the adjustment of associative connections via Hebbian learning, proportional to how strongly the units in the associated maps (phonological and semantic maps in this case) are co-activated. When a given phonological unit and a given semantic unit have fewer chances to become co-activated, the strengths of their associative links are correspondingly decreased. We could compare this to a situation in which the learner receives no auditory support about the specific meaning-form co-occurrences that he or she expects in the production (MacWhinney, 1997). Given that the learning system is input-sensitive, over time, the meaning-to-form connections will weaken and therefore less likely to occur in production.

Indeed, our network displayed significant ability to recover from generalization errors. When tested for recovery with additional new learning (500 epochs), the network recovered from the majority of the overgeneralizations (75% recovery). Recovery in this case is a process of restructuring of the mapping between phonological, semantic, and morphological patterns, and the restructuring is based on the network's ability to reconfigure the associative links through Hebbian learning, in particular, the ability to form new associations between prefixes and verbs and the ability to eliminate old associations that were the basis of erroneous generalizations. For example, *un-* was overgeneralized to *tighten* because of *clench* earlier on; when tested for recovery, only *un-* and *clench* continue to be co-activated. Hebbian learning determines that the associative connection between *un-* and *clench* remains strong, but that between *un-* and *tighten* weakens and gradually decreases to zero. This simulates the situation in which the child receives no support in the input about the relationship between *un-* and *clench*. Of course, in the real learning situation, the strength of the connection between *un-* and *tighten* may also be reduced by a competing form such as *loosen* that functions to express the meaning of **untighten*, whereby principles of contrast or competition help to eliminate the erroneous combination (e.g., Clark, 1987; MacWhinney, 1987).

Note that the restructuring of associative connections often goes hand-in-hand with the reorganization of the corresponding maps. For example, as the associative strengths of *clench* and *tighten* to *un-* varied, the verbs' representations also became more distinct. This simulated result is consistent with Pinker's (1989) criteria proposal that children recover from generalizations by recognizing fine and subtle semantic and phonological properties of verbs. In the few cases in which our network did not recover from overgeneralizations, the network was unable to make the fine semantic distinctions between verbs (see Li, 1999, for details).

5. General Discussion and Conclusions

In this chapter I attempt to provide a computational perspective on a developmental issue. I started with two types of approaches to the problem of the acquisition of word meanings. I then gave a connectionist account of the acquisition of semantic structures and morphological systems, presenting modeling results from both a feed-forward network and a self-organizing network. I have chosen to examine a classical puzzle that Whorf presented some 40 years ago, the issue of cryptotype in connection with the use and acquisition of the English reversive prefix *un-*. This problem differs from many of the currently debated topics, for example, the acquisition of the English past-tense where the patterns of use largely depend on phonological constraints and where the focus of debate has been on the competition between regular rules and exceptions. The *un-* problem examined here is essentially semantic, and there seems to be no regular rule that governs the use of this prefix (hence "intangible", as Whorf named it). Our connectionist models provide some insights into the understanding of Whorf's puzzle, in particular, the understanding of the emergence of complex semantic structures in language acquisition and the role of a structured semantic representation in morphological productivity (e.g., overgeneralization). The simulation results suggest a dynamic learning picture in which the network extracts shared semantic information, develops representations of the cryptotype, and overgeneralizes morphological devices. Such results allow us to understand the processes underlying important phenomena such as the U-shaped behavior in language acquisition.

Current debates in cognitive science and psycholinguistics revolve around the issue of the nature of linguistic representation. Symbolic theories construe linguistic representations in terms of rules in physical symbol systems. A child is said to have a general rule in her mental representation, “adding *-ed* to make the past tense”, at some stage of language acquisition. This kind of description seems intuitively clear, and the rule offers a powerful mechanism for productivity. Connectionist models provide alternative explanations to this perspective, explanations that place emphasis on the statistical learning processes that lead to rule-like behaviors. In this chapter I have demonstrated that the acquisition of linguistic patterns, such as the prefixation of *un-*, can be construed as emerging out of basic processing capacities, that is, the processing of the intricate relationships among phonological and semantic features, lexical items, and morphological devices in a natural language. This perspective seems to be especially suited for the problem that we have at hand, the cryptotype problem that was once thought “subtle” and “intangible” in a symbolic framework. In our view, the reason for the intangibility of the cryptotype is probably that the semantic features that unite different members of a cryptotype are represented in a complex distributed fashion (e.g., feature overlaps across categories), such that they are not easily subject to traditional symbolic analysis, but are accessible to native intuition (according to Whorf). Native intuitions are clearly implicit representations of the complex semantic relationships among verbs and morphological markers, and connectionist networks provide mechanisms to capture these intuitions through weighted connections, distributed representations, and nonlinear dynamics.

Virtually the same story could be told about many other linguistic domains in which the problem is primarily semantically motivated. For example, the use of classifiers is one of the hardest problems for second language learners of Chinese, as well as a major challenge to linguistic theories (cf. Chao, 1968; Lakoff, 1987, Li & Thompson, 1981). Each noun in Chinese has to be preceded by a classifier that categorizes the object of the noun in terms of its shape, orientation, dimension, texture, countability, and animacy. The appropriate uses of most classifiers by native speakers are mostly automatic, yet it is difficult for linguists to come up with clear descriptions of symbolic rules that govern their uses. We can probably assume that native speakers have acquired a representation by a connectionist cryptotype-like mechanism in which multiple weighted semantic features connected in a network jointly support the use of classifiers. We have recently successfully applied this type of mechanisms and explanations to the study of the acquisition of inherent verb aspect and tense-aspect morphology in Chinese, English, and Japanese (see Li & Bowerman, 1998; Li, 2000, 2003; Li & Shirai, 2000). Following this line of research further, we have developed the DevLex model, a self-organizing neural network model for the development of the lexicon. We have applied DevLex to the modeling of monolingual and bilingual lexicon acquisition, simulating the formation of categorical representations, the confusion of competing lexical items in early speech, and the spurt of vocabulary in early word production (see details in Farkas and Li, 2002; Hernandez, Li, and MacWhinney, 2005; Li and Farkas, 2002; Li, Farkas, and MacWhinney, 2004).

In sum, we can start to understand some of the most difficult problems in language acquisition, for example, the acquisition of semantic structures, when we take a computational approach of the type discussed here. Structured semantic representations can emerge from statistical computations of the various constraints among lexical items, semantic features, and morphological markers in a high-dimensional space of language use, as they continuously evolve and develop. The evolution and development of semantic representations as acquired by children may be due to simple probabilistic procedures of the sort embodied in connectionist networks or statistical learning mechanisms for form-to-form and form-to-meaning mappings.

Acknowledgments

Preparation of this article was supported by grants from the National Science Foundation (#BCS-9975249; #BCS-0131829), and a Faculty Research Grant from the University of Richmond. I would like to thank Elizabeth Bates, Melissa Bowerman, and Jeffrey Elman for their discussions on the feed-forward network, Brian MacWhinney and Risto Miikkulainen for their comments and discussions on the self-organizing network, and Curt Burgess and Kevin Lund for making available the HAL semantic vectors for our modeling.

EndNotes:

• Correspondence concerning this article should be addressed to Ping Li, Department of Psychology, University of Richmond, Virginia, VA 23173, U.S.A. E-mail: pli@richmond.edu .

[1] Readers who are interested in details of connectionist theory and methods should read Rumelhart, McClelland, and the PDP Research Group (1986). For non-technical introduction of connectionism, read Bechtel and Abrahamsen (1991) and Spitzer (1999). For technical discussions, read (progressively more technical) Dayhoff (1990), Fausett (1994), Anderson (1995), and Hertz, Krogh, and Palmer (1991). For its relevance to developmental theories, read Elman et al. (1996) and Klair and MacWhinney (1998). For a comprehensive review of all major fields in neural networks, consult Arbib (1995).

[2] For some readers, these two sets of models may simply be viewed as the same kind of models, given that they both rely on statistical patterns and are in many ways closely related.

[3] Note that the 3.8 million words represent only a small portion of what the child is exposed to in the learning environment. According to one estimate, an average three-year-old has been exposed to 10-30 million words (Hart & Risley, 1995).

[4] Diary notes of my daughter's speech also include similar uses: *unbuild* the snowman" was used to refer to the detachment of decorative pieces from the snowman, and *untape* to refer to the removal of tape from a piece of paper that has been taped (child was 6 years and 9 months).

References

Baker, C
1979 Syntactic theory and the projection problem. *Linguistic Inquiry, 10,* 533-581.

Bates, Elizabeth
1984 Bioprograms and the innateness hypothesis: Commentary on Bickerton. *Behavioral and Brian Sciences, 7,* 188-190.

Bensch, P.A
1991 Neo-structuralism: A commentary on the correlations between the work of Zelig Harris and Jeffrey Elman. *Center for Research in Language Newsletter,* 5 (2).

Bowerman, Melissa
1982 Reorganizational processes in lexical and syntactic development. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge: Cambridge University Press.

1983 Hidden meanings: the role of covert conceptual structures in children's development of language. In D. Rogers & J. Sloboda (Eds.), *The acquisition of symbolic skills.* New York: Plenum.

1988 The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining language universals.* New York: Basil Blackwell.

Bownds, M.D
1999 *The biology of mind: Origins and structures of mind, brain, and consciousness.* Bethesda, MD: Fitzgerald Science Press.

Brown, Roger
1973 *A first language.* Cambridge, MA: Harvard University Press.

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

Chao, Y.-R
1968 *A grammar of spoken Chinese.* Berkeley: University of California Press.

Clark, Eve, K. Carpenter & W. Deutsch
1995 Reference states and reversals: Undoing actions with verbs. *Journal of Child Language, 22,* 633-662.

Comrie, Bernard

1976 *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge, England: Cambridge University Press.

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

1993 Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71-99.

1995 Language as a dynamic system. In R. Port & T. van Gelder (eds.), *Mind as motion.* Cambridge, MA.: MIT Press.

Elman, Jeffrey L., E. Bates, M. Johnson, A. Karmiloff-Smith, D. Parisi & K. Plunkett
1996 *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

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

Hebb, D
1949 *The organization of behavior: A neuropsychological theory.* New York, NY: Wiley.

Klahr, D., & B. MacWhinney
1998 Information processing. In Damon, W., Kuhn, D, and Siegler, W. (eds.), *The Manual of Child Psychology* (Vol. 2).

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

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

1995 *Self-organizing maps.* Heidelberg: Springer-Verlag.

Kuczaj, S
1977 The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16,* 589-600.

Lachter, J., & T. Bever
1988 The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. *Cognition, 28,* 195-247.

Lakoff, George
1987 *Women, fire, and dangerous things.* Chicago: The University of Chicago Press.

Li, Charles & Sandra Thompson
1981 *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Li, Ping
1993 Cryptotypes, form-meaning mappings, and overgeneralizations. In E. V. Clark (Ed.), *The Proceedings of the 24th Child Language Research Forum,* Center for the Study of Language and Information Publications, Stanford University.

1999 Generalization, representation, and recovery in a self-organizing feature-map model of language acquisition. In Hahn & S.C. Stoness (eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

2000 The acquisition of lexical and grammatical aspect in a self-organizing feature-map model. In L. Gleitman & Aravind K. Joshi (eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

2003 Language acquisition in a self-organising neural network model. In P. Quinlan (Ed.) *Connectionist models of development: Developmental processes in real and artificial neural networks.* Philadelphia & Brighton: Psychology Press.

Li, Ping & Melissa Bowerman
1998 The acquisition of lexical and grammatical aspect in Chinese. *First Language, 18,* 311-350.

Li, Ping, C. Burgess & K. Lund
2000 The acquisition of word meaning through global lexical co-occurrences. In E. Clark (ed.) *Proceedings of the 30th Child Language Research Forum.* Cambridge, MA: Cambridge University Press.

Li, Ping & Brian MacWhinney
1996 Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science, 8*, 1-28.

2002 PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, and Computers, 34,* 408-415.

Li, Ping & Y. Shirai
2000 *The acquisition of lexical and grammatical aspect.* Berlin and New York: Mouton de Gruyter.

MacWhinney, Brian
1989 Competition and lexical categorization. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization.* New York: Benjamins.

1998 Models of the emergence of language. *Annual Review of Psychology*, *49*, 199-227.

2000 *The CHILDES project: Tools for analyzing talk.* Hillsdale, NJ: Lawrence Erlbaum.

2001 Lexicalist connectionism. In P. Broeder & J.M. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches.* Oxford, UK: Oxford University Press.

MacWhinney, Brian & J. Leinbach
1991 Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40,* 121-157.

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

Marchand, H
1969 *The categories and types of present-day English word-formation: a synchronic-diachronic approach.* Munich: C.H. Beck'sche Verlagsbuchhandlung.

McClelland, Jay & David Rumelhart
1981 An interactive activation model of context effects in letter perception: Part 1. An account of the basic findings. *Psychological Review, 88,* 375-402.

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

1997 Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language, 59,* 334-366.

Pinker, Steven
1989 *Learnability and cognition: The acquisition of argument structure.* Cambridge, MA: MIT Press.

1991 Rules of language. *Science, 253,* 530-535.

1999 Out of the minds of babes. *Science, 283,* 40-41.

Pinker, Steven & Alan Prince
1988 On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73-193.

Plunkett, K. & V. Marchman.
1991 U-shaped learning and frequency effects in a multi-layer perceptron: Implications for child language acquisition. *Cognition*, 38: 43-102.

1993 From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48: 21-69.

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

Rumelhart, David & Jay McClelland
1986 On learning the past tenses of English verbs. In J. McClelland, D. Rumelhart, and the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. I). The MIT Press.

Rumelhart, David, G. Hinton & R. Williams
1986 Learning internal representations by error propagation. In J. McClelland, D. Rumelhart, and the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. I). The MIT Press.

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

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 Gangyao*. Beijing: Commercial Press).

Seidenberg, Mark
1997 Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275,* 1599-1603.

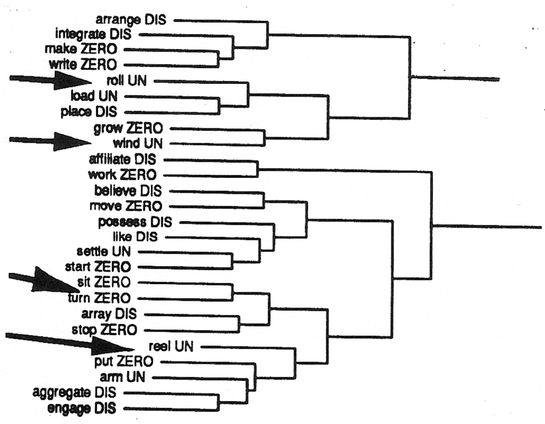
Spitzer, M
1999 *The mind within the net: Models of learning, thinking, and acting.* The MIT Press.

Vendler, Z
1967 *Linguistics in philosophy*. Ithaca: Cornell University Press.

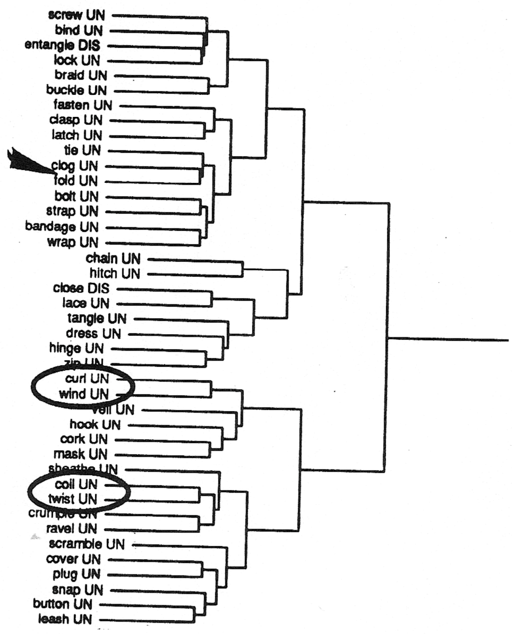
Whorf, Benjamin L
1956 Thinking in primitive communities. In J. B. Carroll (Ed.), *Language, thought, and reality*. The MIT Press.

APPENDIX

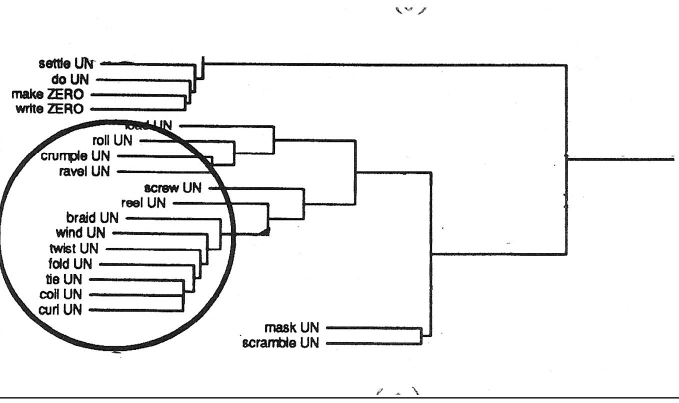
Figure 3 - Part A:



Part B:



Part C:


settle UN
do UN
make ZERO
write ZERO
roll UN
crumple UN
ravel UN
screw UN
reel UN
braid UN
wind UN
twist UN
fold UN
tie UN
coil UN
curl UN
mask UN
scramble UN