Generalization, Representation, and Recovery in a Self-Organizing Feature-Map Model of Language Acquisition

Ping Li (Ping@Cogsci.Richmond.Edu)
Department of Psychology
University of Richmond
Richmond, VA 23173, USA

Abstract

This study explores the self-organizing neural network as a model of lexical and morphological acquisition. We examined issues of generalization, representation, and recovery in a multiple feature-map model. Our results indicate that self-organization and Hebbian learning are two important computational principles that can account for the psycholinguistic processes of semantic representation, morphological generalization, and recovery from generalizations in the acquisition of reversive prefixes such as *un-* and *dis-*. These results attest to the utility of self-organizing neural networks in the study of language acquisition.

Introduction

Language learning is characterized by the learner's ability to generalize beyond what is heard in the input. One current debate on connectionist models of language acquisition concerns the issue of generalization (Elman, 1998). Probably the best-known example in this debate has to do with the acquisition of the English past tense: children generalize *-ed* to irregular verbs, producing errors like *falled*, *breaked*, and *comed*. Connectionist researchers argue that their networks, like human children, display generalizations in a U-shaped pattern of learning (Rumelhart & McClelland, 1986; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1991). In contrast, symbolic theorists argue that generalization is rule-based (Pinker, 1991; Pinker & Prince, 1988).

Most of this debate has revolved around a specific cluster of connectionist models, the back-propagation network as a model of language acquisition. Several limitations are known to the back-propagation algorithm, especially in the context of language acquisition: in particular, back-propagation relies on a gradient-descent weight adjustment process to reduce the error between desired and actual outputs. According to the well-known "no negative evidence" argument (Baker, 1979; Bowerman, 1988), 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 the back-propagation network.

In this study, we explore self-organizing neural networks, in particular, the self-organizing feature maps as a potential class of models of language acquisition. In contrast to back propagation, 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. The self-organizing process extracts an efficient and compressed internal representation from a high-dimensional input space and expresses this new representation in a map structure (Kohonen, 1989). There are three important properties of self-organizing feature maps that make them particularly well suited to the study of language acquisition.

(1) *Self-organization*. Self-organization in these networks typically occurs in a two-dimensional 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, according to how similar by chance the input pattern is to the weight vectors of the units. Once a unit becomes active in response to a given input, the weight vectors of 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, every time an input is presented, an area of units will become activated on the map (the activity "bubbles"), and the maximally active units are taken to represent the input. Initially activation occurs in large areas in the map, but gradually learning becomes more focused so that only the maximally responding units are active. This process continues until all the inputs have found some maximally responding units.

(2) *Representation*. As a result of this self-organizing process, the statistical structures implicit in the high-dimensional space of the input are represented as topological structures on a two-dimensional space. Because the network develops activity bubbles to capture the input space, similar inputs will end up activating the same units or units in nearby regions, yielding a new similarity structure that becomes clearly visible on the map. This self-organized representation has clear implications for language acquisition: the formation of activity bubbles may capture critical processes of the emergence of lexical categories in children's acquisition of the lexicon. In particular, the network organizes information first in large areas of the map and gradually zeros in on small areas; this zero-in process is a process from diffuse to focused patterns of activity that leads to continuous adaptation of the network's representation. This process can naturally explain many generalization errors reported in the literature: for example, substitutions of *put* for *give* ("put me the bread") or *fall* for *drop* ("I falled it") reflects the child's recognition of diffuse lexical similarities but not the focused fine distinctions between words (Bowerman, 1982). Miikkulainen (1997) showed that in a lesioned self-organizing feature map, behaviors of dyslexia (e.g., producing *dog* in response to *sheep*) can result from partial damage

to the semantic representation (in effect a diffuse representation of meaning).

(3) *Hebbian learning*. Hebbian learning is essentially a co-occurrence learning mechanism, according to which the associative strength between two neurons is increased if the neurons are both active at the same time (Hebb, 1949). The amount of increase is proportionally to the level of activation of the two neurons. Different self-organizing maps can be connected via Hebbian learning, such as in Miikkulainen's (1997) multiple feature-map model: initially all units on one map are connected to all units on the other map; as self-organization takes place, the associations become more focused, so that in the end only the maximally active units on the two (or more) maps are associated. Hebbian learning has strong implications for language acquisition in that it can account for how the child abstracts relationships between phonological, semantic, and morphological properties of words on the basis of how often these properties co-occur and how strongly they are co-activated in the representation.

Because of these properties, self-organizing networks (a) allow us to track the development of the lexicon as an emergent process more clearly in the network's self-organization (from diffuse to focused patterns or from incomplete to complete associative links); (b) allow us to model one-to-many or many-to-many associations between forms and meanings in the development of the lexicon and morphology, and (c) provide us with a set of biologically more plausible and computationally more relevant principles to study language acquisition without relying on negative evidence to learn. They 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, et al., 1998). 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).

In this paper I focus on the problem of the English reversive prefixes that has been discussed by Whorf (1956) and Bowerman (1982) in the context of morphological generalization. In English, one can use the prefix *un-* to indicate the reversal of an action in verbs like *unbuckle*, *uncoil*, *undress*, *unfasten*, and *untie*, but not **unfill*, **unhang*, **unkick*, **unpush*, or **unsqueeze*. Why is *un-* allowed with some verbs but not with others? Whorf hypothesized that there is some underlying semantic category that licenses the use of *un-* (roughly "a covering, enclosing, and surface-attaching meaning"). Because this category functions only covertly (i.e., by the restrictions it places on *un-*), he called it a "cryptotype". To Whorf, the problem is that the precise meaning of the cryptotype is "subtle" or even "intangible", but the prefix that it licenses is productive. Bowerman argued that the notion of cryptotype, though elusive, might play an important role in children's acquisition of *un-*. Her data showed that children produce generalization errors like **unbury*, **unhang*, **unhate*, **unopen*, **unpress*, **unspill*, or **unsqueeze* starting from about age 3. She had two hypotheses on the role of cryptotype: (a) "generalization via cryptotype", i.e., recognition of the cryptotype leads to overly general uses (overgeneralizations); e.g., *bury* fits the cryptotype just as *cover* does, so say **unbury*. (b) "recovery via cryptotype", i.e., children use the cryptotype to recover from overgeneralization errors; e.g., *hate* does not fit the cryptotype meaning and only verbs in the cryptotype can take *un-*, so stop saying **unhate*.

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? In Li (1993) Li & MacWhinney (1996) we attempted to answer this question by simulating *un-* and its cryptotype in a back-propagation network. We hypothesized that cryptotypes are intangible only because traditional symbolic methods are less effective for analyzing the complex semantic structure: words in a cryptotype vary in the number of relevant semantic features, the strength of activation of each feature, and the degree of overlap of features. These complex structural properties lend themselves naturally to distributed representations and connectionist learning. We trained a network to map semantic features of verbs to three prefixation patterns: *un-*, its competitor *dis-*, and no-prefixation. Our results indicated that (a) the network formed internal representations of semantic categories that corresponded roughly to Whorf's cryptotype, on the basis of learning limited semantic features of verbs and morphological classes; (b) the network produced overgeneralization errors similar to those reported by Bowerman (1982), Clark et al (1995), and those observed in the CHILDES database.

In this study, we examine the representation of cryptotypes, the generalization of prefixes, and the recovery from generalizations in a self-organizing feature-map model. As discussed above, self-organizing feature maps learn on the basis of self-organization, produce representations in a map structure, and form associative connections via Hebbian learning. These properties have recently been implemented in DISLEX, a multiple feature-map model of the lexicon (Miikkulainen, 1997). In this study, we use DISLEX as a basis to simulate generalization, representation, and recovery. We think that the self-organizing and Hebbian learning processes as simulated in DISLEX can help us to understand the representational basis of morphological generalization and the learner's recovery from generalizations.

Method

Network Architecture

DISLEX is a multiple feature-map model of the lexicon, in which different self-organizing maps dedicated to different types of linguistic information (orthography, phonology, or semantics) are connected through associative links via Hebbian learning. During learning, an input pattern activates a unit or a group of units on one of the input maps, and the resulting bubble of activity propagates through the associative links and causes an activity bubble to form in the other map. If the direction of the associative propagation is from phonology or orthography to semantics, comprehension is modeled; production is modeled if it goes from semantics to phonology or orthography. The activation of co-occurring lexical and semantic representations leads to continuous organization in these maps, and to adaptive formations of associative connections between the maps. Figure 1 presents a schematic diagram of the architecture of the model.

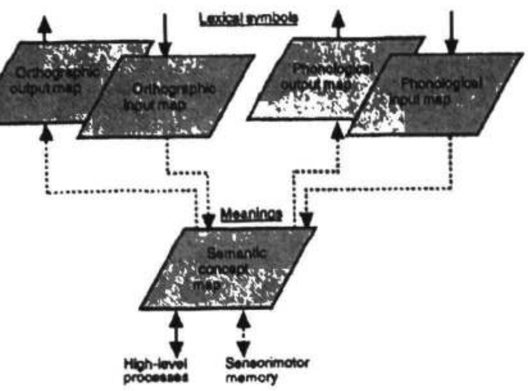

Figure 1: A multiple feature-map model of the lexicon (Miikkulainen, 1997)

In this study, we applied DISLEX to the examination of lexical and morphological acquisition. We constructed two self-organizing maps, each of the size of 25 x 25 units, one for the organization of phonological input (henceforth the phonological map), and the other for the organization of semantic input (the semantic map). We used no orthographic maps since we were modeling acquisition in young children who are preliterate.

Input Representations

The input data to our network were 228 verbs based on Li (1993) and Li and MacWhinney (1996). Forty-nine of them were verbs with the prefix *un-*, 19 of them were verbs with the competitor prefix *dis-*, and the remaining 160 were verbs with no prefixes (*un-* and *dis-* both indicate the reversal of the action of the verb, as in *untie* and *disassemble*). The relative higher proportion of the last type of verbs (i.e., zero verbs) as compared with *un-* and *dis-* verbs was intended to reflect the distribution of these forms in the input to children.

Previous connectionist models of language acquisition have often relied on the use of artificial input/output representations (e.g., randomly generated patterns of phonological or semantic representations) or representations that are constructed ad hoc by the modeler. For example, in our previous studies we represented each verb as a pattern of 20 semantic features, selected on the basis of our linguistic analyses (see Li, 1993; Li & MacWhinney, 1996). However, the use of this type of representation is subject to the criticism that the model works just because of the presence of these features in the representation. In this study, we wanted to use more linguistically grounded input data to simulate lexical and morphological acquisition. Thus, we represented our inputs as follows.

Phonological representations to our network were based on a syllabic template coding developed by MacWhinney and Leinbach (1991). Instead of a simple phonemic representation, this representation reflects current autosegmental approaches to phonology, according to which the phonology of a word is made up by combinations of syllables in a metrical grid, and the slots in each grid made up by bundles of features that correspond to phonemes, *C*'s (consonants) and *V*'s (vowel). The MacWhinney-Leinbach model used 12 *C*-slots and 6 *V*-slots that allowed for representation of words up to three syllables. For example, the 18-slot template *CCC VV CCC VV CCC VV CCC* represents a full trisyllabic structure in which each *CCCVV* is a syllable (the last *CCC* represents the consonant endings). Each *C* is represented by a set of 10 feature units, and each *V* by a set of 8 feature units.

Semantic representations to our network were based on the lexical co-occurrence analyses in the Hyperspace Analogue to Language (HAL) model of Burgess and Lund (1997). HAL represents word meanings through multiple lexical co-occurrence constraints in large text corpora. In this representation, the meaning of a word is determined by the word's global lexical co-occurrences in a high-dimensional space: 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 of the target word. Thus, a word is represented as a vector that encodes the multiple constraints (dimensions) in a high-dimensional space of language use. We used 100 dimensions for the unit length of the vectors.

Task and Procedure

Upon training of the network, a phonological input representation of the verb was inputted to the network, and simultaneously, the semantic representation of the same input 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 prefix-able with *un-* or *dis-*, the phonological representation of *un-* or *dis-* was also co-activated with the phonological and the semantic representations of the verb stems. 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.

To observe effects of learning on the network's representation, generalization, and recovery, we designed four stages to train the network. (1) A verb's phonological representation was co-activated with its semantic representation on a one-to-one basis, which means that the network saw only the verb's phonological representation and its semantic representation simultaneously. This was done to model the whole-word learning stage, at which children have not analyzed morphological devices as entities separate from the verb stems (Bowerman, 1982). (2) One-to-one mapping was relaxed, so that the phonological and semantic representations of verb stems (e.g., *tie*, *connect*), prefixed verbs (*untie*, *disconnect*), and the prefixes themselves (*un-*, *dis-*) were all co-activated in the network. (3) Twenty-five novel verbs were introduced to the network in order to test whether generalizations would occur in our network as in children's speech. These were verbs on which previous studies have reported children's generalizations (e.g, **ungrip*, **unpress*, and **untighten*; see Bowerman, 1982; Clark et al., 1995). Generalization was tested by inputting the verbs to the network without having the network self-organize the verbs or learn the phonological-semantic associations. (4) Self-organization and Hebbian learning resumed for the novel verbs introduced at Stage 3 in order to test if the network could recover from generalizations.

All simulations were run on a SUN Ultra 1 workstation, using the DISLEX codes configured by Miikkulainen (1999).

Results and Discussion

To analyze the simulation results, we focus here on three levels of analysis: the network's representation of verb semantics, its patterns of morphological generalization, and its ability to recover from generalization errors.

The Representation of Cryptotype

In this study, we wanted to analyze whether our network developed structured representation as a function of the self-organization of verb semantics. In particular, we wanted to see how the patterns of activity formed in the maps can capture Whorf's notion of cryptotype.

As discussed earlier, a distinct property of self-organizing feature maps is that the structures in the network's new representation are clearly visible as activity bubbles or patterns of activity on the two-dimensional map; this property obviates the need of extra steps of mathematical analysis (e.g., cluster analysis or principal component analysis) as required in other connectionist networks. In our network, the self-organization process extracted the semantic structures from the high-dimensional space of the HAL semantic vectors and expressed them on the two-dimensional map as concentrated patterns of activity. Figure 2 presents a snapshot of the network's self-organization of 120 verbs after the network was trained for 600 epochs at Stage 1.

An examination of the semantic map shows that the network has clearly developed forms of representation that correspond to the category of cryptotype that Whorf believed governs the use of *un-*. In Li and MacWhinney (1996) we suggested that a connectionist model provides a formal mechanism to capture Whorf's cryptotype, in that there can be several "mini-cryptotypes" that work collaboratively as interactive "gangs" (McClelland & Rumelhart, 1981) to support the formation of the larger cryptotype. The idea of "mini-cryptotype" is realized 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

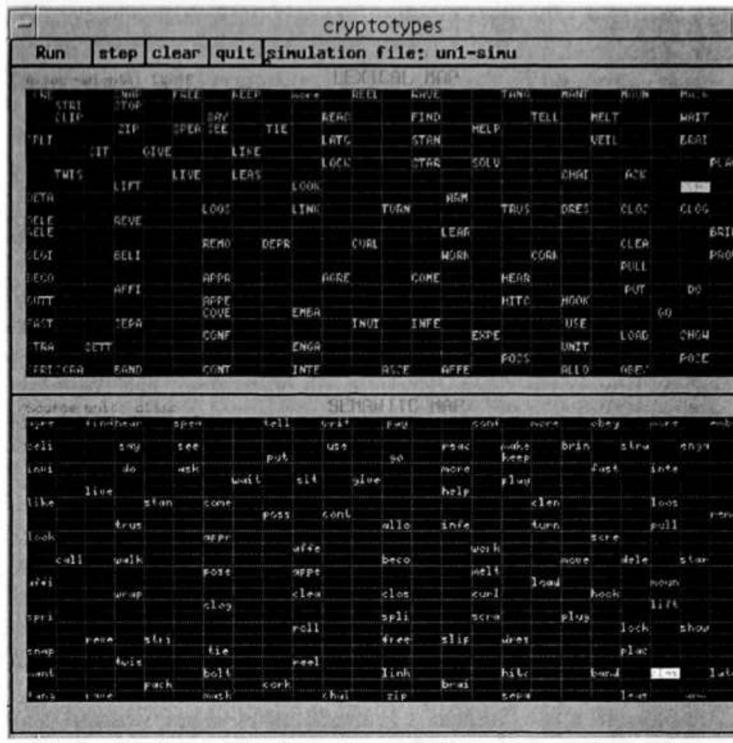

Figure 2: Phonological and semantic representations in DISLEX after the network was trained on 120 verbs for 600 epochs. The upper panel is the phonological map (in capital letters), and the lower panel the semantic map. Words longer than four letters are truncated.

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 and 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 the figure, 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-*.

Representation and Generalization

Neural networks are considered to be able to generalize to novel patterns (Elman, 1998). But do they show the same types of generalization as children do? And on what basis do they generalize? Our simulation results indicate that our network was not only able to capture the elusive cryptotype category by way of self-organization, but also able to generalize on the basis of its representation of the cryptotype. For example, the network produced overgeneralization errors that match up with empirical data when tested for generalization at Stage 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.

First, most of these overgeneralizations involve verbs that fall within the *un-* cryptotype. These verbs (e.g., *connect, freeze, grip, peel, plant, press, spill, stick,* and *tighten*) were mapped to the bottom layer on the semantic map within which we identified the network's representation of the cryptotype. Earlier, we pointed out two hypotheses regarding the role of cryptotype in children's acquisition of un- according to Bowerman: "generalization via cryptotype" and "recovery via cryptotype". Our results here are 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). Consistent with our previous simulations, we found no flagrant violations of the cryptotype in the network's generalizations such as **unhate* or **untake* (as in Bowerman's data); hence there was no basis for the recovery via cryptotype hypothesis, that is, that the learner can use the representation of cryptotype to recover from overgeneralizations.

Second, all the above generalizations were simulated production errors, in which case patterns of activity in the semantic map were propagated through associative links to the phonological map. The ability to simulate both comprehension and production through associative connections is a distinct property of DISLEX (see Network Architecture). 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 can associate the semantics of *tighten* with *un-* because of *clench*, even though the network learned only the association for *un-clench* and not *un-tighten* (i.e., at an earlier stage *tighten* was not included in the training). 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 structures in the lexical representation (which in turn is a result of self-organization) and from the associative learning of semantics and phonology.

Finally, generalization errors in our data were not limited only to morphological generalizations. We also found lexical generalizations similar to those reported by Bowerman (1982) and Miikkulainen (1997) (see Introduction). Most important, these generalizations demonstrate further the intimate relationship between representation and generalization. For example, our network produced *see* in response to *say*, *detach* in response to *delete*, *begin* in response to *become*, due to its representation of these pairs of words in the same region on the phonological map. These generalizations well resemble lexical errors in surface dyslexia (Miikkulainen, 1997). Similarly, the network comprehended *see* as *speak*, *arm* as *clasp*, and *unscrew* as *hook*, due to its representation of these pairs of words in the same region on the semantic map, and these generalizations resemble lexical errors in deep dyslexia. Again, self-organization of lexical information and Hebbian learning of associative connections account for the origin of this type of lexical generalizations.

Mechanisms of Recovery from Generalizations

Our last analysis of the simulation results involves the network's ability to recover from generalization errors. The network in Li (1993) and Li and MacWhinney (1996) suffered, by and large, from the failure to recover from overgeneralization errors. This failure, we hypothesized, was due to the gradient-descent error-adjustment process used in back-propagation. Can our self-organizing network recover from generalizations? If so, what computational mechanisms permit its recovery?

Our network displayed a significant ability of recovery from generalization errors. When tested for generalizations at Stage 3, no learning took place in the network for self-organization or associative connection. When tested for recovery at Stage 4, self-organization and Hebbian learning resumed. Within 200 epochs of new learning during this stage, our network recovered from the majority of the overgeneralizations tested at Stage 3. 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.

As discussed earlier, adjustment of associative connections via Hebbian learning in DISLEX is 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. For example, *un-* and *tighten* were co-activated because of *clench* at Stage 3; at Stage 4 *un-* and *clench* continue to be co-activated, but *un-* and *tighten* do not get co-activated. Hebbian learning determines that the associative connection between *un-* and *clench* remains to be strong, but that between *un-* and *tighten* gets eliminated, thereby simulating what happens at the final phase of the U-shaped learning when errors disappear. This result models the process that children's overgeneralizations are gradually eliminated when there is no auditory support in the input about specific co-occurrences that they expect (MacWhinney, 1997). 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* (e.g., Clark, 1987, MacWhinney, 1987).

Hebbian learning coupled with self-organization provides a simple but powerful computational principle to account for the recovery process. Restructuring of associative connections often goes hand-in-hand with the reorganization of the corresponding maps. For example, at Stage 4, the network developed finer representations for verbs such as *clench* and *tighten*: as the associative strengths of these verbs to *un-* varied, their representations also became more distinct. This process in our simulation is consistent with the criteria approach of Pinker (1989) which argues that children recover from generalizations by recognizing fine and subtle semantic and phonological properties of verbs (although we do not assume as Pinker does that fine distinctions among verbs rely on the child's innate capacity). Interestingly, in the few cases in which our network did not recover from generalizations, the network was unable to make the fine distinctions between verbs on the basis of meanings; for example, because it was unable to separate on the semantic map *stick* from *screw*, *press* from *zip*, and *freeze* from *bolt*, it continued to produce the erroneous **unstick*, **unpress*, and **unfreeze*. This inability might be due to resource limitations (i.e., size of the map); we are currently investigating this problem using much larger feature maps (e.g., map of 50 x 50 units).

Conclusion

In this paper I showed that self-organizing neural networks can be used successfully to model and provide insights into language acquisition, particularly with respect to issues of generalization, representation, and recovery. Our simulated DISLEX model, without receiving hand-crafted features, was able to capture elusive semantic categories such as Whorf's cryptotype, display overgeneralization errors as children do, and recover significantly from overgeneralizations. Although the simulation results presented here are preliminary, we think that they serve to deepen the link between previous empirical and modeling results and new models in neural networks. Future research in this direction will involve the development of more realistic training schedules (e.g., incremental learning), the use of input representations that are grounded in children's language (e.g., semantic vectors based on lexical co-occurrence analysis of the CHILDES database), and the development of network architecture that is better suited to the task of morphological acquisition (e.g., use of separate morphological maps that allow the interaction among morphemes and verb stems on both phonological and semantic levels).

Acknowledgments

This research was supported by a Faculty Research Grant from the University of Richmond. I am very grateful to Brian MacWhinney and Risto Miikkulainen for their help, comments, and discussions at various stages of the project, and to Curt Burgess and Kevin Lund for making available their HAL semantic vectors to our modeling.

References

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

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

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

Clark, E.V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

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

Elman, J. (1998). Generalization, simple recurrent networks, and the emergence of structure. In M. Gernsbacher & S. Derry (eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum.

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

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

Li, P. (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, Stanford University, 162-178.

Li, P., & MacWhinney, B. (1996). Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science, 8,* 3-30.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

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

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

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

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

Pinker, S. (1991). Rules of language. *Science, 253,* 530-535.

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

Rumelhart, D. & McClelland, J. (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. II). Cambridge, MA: MIT Press.

Whorf, B. (1956). Thinking in primitive communities. In J. B. Carroll (Ed.), *Language, thought, and reality.* Cambridge, MA: MIT Press.