CHAPTER FOUR

Language acquisition in a self-organising neural network model

Ping Li
Department of Psychology, University of Richmond, Virginia, USA

INTRODUCTION

One crucial aspect of human language learning is the learner's ability to generalise existing patterns to novel instances. This ability often leads to various erroneous generalisations in learning. "Overgeneralisation" is one such type of error, characterised by the learner's use of a linguistic pattern that is broader in scope than the corresponding adult uses (Bowerman, 1982; Brown, 1973; Clark, 1987; Pinker, 1989). Perhaps the best-known example of overgeneralisation is the acquisition of the English past tense: children generalise *-ed* to irregular verbs, producing errors like *falled*, *breaked*, and *comed* (Brown, 1973; Kuczaj, 1977). But just what leads to children's overgeneralisations has been under intensive debate. Using the acquisition of the English past tense as an example, researchers have debated whether language acquisition should be characterised as a symbolic, rule-based learning process or as a connectionist, statistical learning process. Symbolic theorists assume that overgeneralisation errors result from the child's internalisation and application of linguistic rules (Ling & Marinov, 1993; Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992; Pinker, 1991, 1999; Pinker & Prince, 1988), whereas connectionists argue that overgeneralisations reflect the child's ability to extract statistical regularities from the input (MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1991, 1993; Rumelhart & McClelland, 1986; Seidenberg, 1997).

In contrast to the well-known overgeneralisation patterns, learners sometimes also exhibit "undergeneralisation"—generalisations that are narrower in scope than the corresponding adult usage. A typical example of undergeneralisation is one in which young children initially restrict tense-aspect morphology to specific semantic categories of verbs. For example, early on, English-speaking children use the progressive marker *-ing* only with atelic verbs that indicate durative processes (e.g. *walk*, *swim*, and *play*), whereas they use the past-perfective marker *-ed* only with telic verbs that indicate actions with clear endpoints or end result (e.g. *spill*, *break*, and *fall*). Capitalising on these patterns in early child language, some investigators hypothesise that children have innate semantic categories that bias them towards certain grammatical distinctions as expressed by contrasting morphological markers (Bickerton, 1981, 1984). Other researchers disagree with such hypotheses, arguing that the undergeneralisation patterns reflect learners' statistical analyses of the distributional properties of verbs and morphology in the input language (see Li & Shirai, 2000, for a summary).

In this chapter, I present a self-organising neural network that attempts to model both overgeneralisations and undergeneralisations in language acquisition, without making strong assumptions about the innateness of semantic categories or the symbolic nature of categorical representation. In particular, I will examine: (1) the acquisition of the English reversive prefixes that has been discussed by Whorf (1956) and Bowerman (1982) in the context of morphological overgeneralisation; and (2) the acquisition of grammatical suffixes that has been discussed by Brown (1973), Bloom, Lifter, and Hafitz (1980), and, Li and Shirai (2000), in the context of morphological undergeneralisation. I take the following observations as a starting point for the current study:

- Most previous connectionist models of language acquisition have been concerned with the phonological properties that govern the use of verb forms, for example, in the acquisition of the English past tense (see Klahr & MacWhinney, 2000, for an overview). Few studies have paid attention to the meaning structure of words, perhaps because of the level of difficulty in representing meaning faithfully in connectionist networks (but see Burgess & Lund, 1997, and Li, Burgess, & Lund, 2000). Reversive prefixes and aspect suffixes in English provide ideal cases where the use of grammatical morphology is governed primarily by semantic rather than phonological properties of lexical items. Our model addresses the relationship between the acquisition of lexical semantics and the learner's ability to generalise morphological devices.
- Most previous models have used artificially generated input representations that are in many cases isolated from realistic language uses. In addition, these input patterns are in most cases "handcrafted" *ad hoc* by

the modeller. Representations of linguistic information constructed in this way are often subject to the criticism that the network works precisely because of the use of certain features in the representation (Lachter & Bever, 1988). To overcome potential problems associated with such approaches to linguistic representations, we attempt to use phonological and semantic representations that more closely approximate the reality of language use. Moreover, we rely on corpus-based linguistic data to establish the sequence as well as the structure of the input data.

- Most previous models have used supervised learning, in particular, the back-propagation learning algorithm as their basis of network training. Although significant progress has been made with these types of networks, there are serious problems concerning the biological and psychological plausibility of such networks. In particular, "back-propagation networks" are known to suffer from catastrophic forgetting (inability to remember old information with new learning), scalability (inability to handle realistic, large-scale problems), and above all, an error-driven learning process that adjusts weights according to the error signals from the discrepancy between desired and actual outputs. In the context of language acquisition, these problems become more transparent. In particular, it would be a very strong argument that the feedback process used in back-propagation resembles realistic processes of child language learning. Such considerations lead us to self-organising neural networks, in particular, the self-organising feature maps, in which learning proceeds in an "unsupervised" fashion, without explicit teaching signals as in back-propagation nets.

This chapter is organised as follows. First, I briefly discuss the two linguistic problems—the use of reversive prefixes in connection with covert semantic categories and the use of grammatical suffixes in connection with aspectual semantic categories. I then describe the acquisition of the reversive prefixes and aspectual suffixes and the corresponding overgeneralisation and undergeneralisation patterns. Next, I present a self-organising neural network model that captures the processes underlying overgeneralisation and undergeneralisation. Finally, I conclude with general remarks on the significance of self-organising neural networks in unravelling the computational and psycholinguistic mechanisms of language acquisition.

THE INTERACTION BETWEEN VERB SEMANTICS AND MORPHOLOGY

Prefixes, suffixes, and verbs

Language is an interactive system. In contrast to early conceptions about systems of language (Chomsky, 1957), linguists and cognitive scientists now accept that linguistic components interact across levels: between syntax and semantics, between syntax and phonology, and between semantics and morphology, and so on. In this chapter, I focus on the interaction between semantics and morphology, one that can be best illustrated with examples from the use of the English reversive prefixes such as *un-* and *dis-* and the use of aspectual suffixes like *-ed* and *-ing*. The centrepiece of grammatical morphology in a sentence is the verb, and thus the study of verbs along with prefixes and suffixes is the main focus of our present research.

In one of the classic papers of early cognitive linguistics, Whorf (1956) presented the following puzzle on prefixation. In English, the reversive prefix *un-* can be used productively with many verbs to indicate the reversal of an action, for example, as in *undress*, *unfasten*, *unlock*, or *untie*. Similar reversal meanings can also be expressed by other prefixes such as *dis-* or *de-*. However, English prevents the use of *un-*, *dis-*, or *de-* in many seemingly parallel forms, such as the ill-formed **undry*, **unkick*, or **unmove*. Why? Whorf proposed that there is an underlying semantic category that governs the use of *un-*: a "cryptotype" or covert semantic category. According to Whorf, cryptotypes only make their presence known by the restrictions that 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 prohibits combinations such as **unkick*. To Whorf, the deep puzzle is that while the use of the prefix *un-* is a productive morphological device, the cryptotype that governs its productivity is elusive: "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."

Whorf did propose the "covering, enclosing, and surface-attaching meaning" as a core meaning for the cryptotype of *un-*. However, 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; for example, Marchand (1969) and Clark, Carpenter, & Deutsch (1995) argue that verbs that license *un-* all involve a change of state, usually expressing a transitive action that leads to some end state or result, as encoded by telic verbs. When the meaning of a verb does not involve a change of state or telicity, the verb cannot take *un-*, thus the ill-formedness of verbs like **unswim*, **unplay*, and **unsnore*.

An alternative prefix, *dis-*, shows many similar properties with *un-*, although Whorf did not discuss this prefix in the context of cryptotype. For example, the base verbs in *disassemble, disconnect, disengage, disentangle*, and *dismantle* all fit Whorf's cryptotypic meanings of binding, covering, and attaching. As a result, many *dis-* and *un-* verbs are competitors, for example, *disconnect* versus *unlink*, or *disengage* versus *uncouple*. These two suffixes, however, do not overlap completely: *dis-* is used for many abstract mental verbs to which *un-* does not apply (e.g. *disassociate, disengage*, and *disentangle*) and, overall, *un-* is much more productive than is *dis-* in modern English.

An equally interesting domain as the above where semantics meets morphology is the use of inflectional suffixes that mark aspectual contrasts, for example, between perfective and imperfective. According to Comrie (1976), imperfective aspect presents a situation with an internal point of view, often as ongoing (progressive) or enduring (continuous), whereas perfective aspect presents a situation with an external perspective, often as completed. In English, the imperfective–perfective contrast is realised in the difference between the progressive *-ing* and the past-perfective *-ed*.[1] Thus, *-ing* marks the progressive aspect—an ongoing event (e.g. "John is walk*ing*"), *-ed* marks the perfective aspect—a completed event (e.g. "John has walk*ed* for an hour"), and *-s* marks the habitual aspect—a routinely performed action or an enduring state (e.g. "John walk*s* for an hour everyday"). In contrast to the grammatical aspect expressed by suffixes, linguists also recognise the importance of "lexical aspect" or "inherent aspect": the temporal properties of a verb's meaning, for example, whether the verb encodes an inherent endpoint or end result of a situation. There are various linguistic descriptions of lexical aspect, but we adopt here a three-way classification (Mourelatos, 1981; Parsons, 1990): (1) *processes*—verbs that encode situations with no inherent endpoint (e.g. *walk*); (2) *events*—verbs that encode situations with inherent endpoint or end result (e.g. *break*); and (3) *states*—verbs that encode situations as homogeneous involving no dynamic or successive phases (e.g. *know*).

The complex relationship between grammatical aspect and lexical aspect is another clear case where morphology interacts with semantics. Like the derivational prefixes *un-* and *dis-*, uses of the inflectional suffixes *-ed, -ing* and *-s* are also in many cases constrained. For example, in English, *-ing* rarely occurs with state verbs; thus, while "John knows the story" is good, "John is knowing the story" sounds odd (Smith, 1983). There are also combinatorial constraints between *-ing* and event verbs; for example, "John is noticing a friend" is distinctly odd. These kinds of constraints are sometimes referred to as

[1] Note that *-ed* marks both past tense and perfective aspect in English, just as *-s* marks both present tense and habitual aspect. In other languages, separate affixes are often used for expressing tense and aspect.

"naturalness of combination" (Comrie, 1976) between verbs and morphology, which may ultimately reflect the intricate relationships between language use and event characteristics. For example, many events with an end result last for such a brief period of time that any comment on them is likely to occur only after the event has ended, for example, situations denoted by verbs like *drop*, *fall*, and *crash* (cf. Brown, 1973). Thus, it is rare for speakers to describe the "ongoing-ness" of such events with *-ing* but more natural to describe them using past-perfective forms. In some languages the less natural combinations may be prohibited altogether from the grammar (see Li & Shirai, 2000).

The prediction that we can derive from these kinds of constraints is that natural speech will exhibit strong associations between given types of verbs and given types of morphology (for example, the perfective-to-event associations). A further, perhaps more important, prediction is that children are able to explore the statistical relationships that exist between verb semantics and morphology in language acquisition. I will return to these predictions when considering empirical and modelling studies of acquisition.

The acquisition of lexicon and morphology

The above discussion demonstrates the close interactions between verb semantics and grammatical morphology in adult language. How do children acquire such interactions?

Bowerman (1982) was among the first to point out the important role of lexical semantics in children's morphological acquisition. In particular, she argued that the onset of lexical or morphological errors signals a change or reorganisation in the child's mental lexicon: Words that are not initially recognised as related are later on grouped together. Thus, we should pay attention not only to the acquisition of morphology *per se*, but also to the developing semantic structure in the child's lexicon. Bowerman illustrated the point with the acquisition of *un-*. Her data suggest that children follow a U-shaped learning curve in learning *un-*, a pattern also found in other areas of morphological acquisition (e.g. the acquisition of the English past tense). At the initial stage, children produce *un-* verbs in appropriate contexts, treating *un-* and its base verb as an unanalysed whole. This initial stage of rote control is analogous to the child's saying *went* without realising that it is the past tense of *go*. At the second stage (from about age 3), children produce overgeneralisation errors like **unarrange*, **unbreak*, **unblow*, **unbury*, **unget*, **unhang*, **unhate*, **unopen*, **unpress*, **unspill*, **unsqueeze*, and **untake* (Bowerman, 1982, 1983; see also Clark et al., 1995 for similar errors in naturalistic and experimental settings). At the final stage of this U-shaped learning, children recover from these errors and overgeneralisations cease.

Bowerman (1982, 1983) proposed that Whorf's notion of cryptotype might play an important role in children's acquisition of *un-*. Cryptotype might influence acquisition at either the second stage or the final stage of the U-shape: (1) "generalisation via cryptotype"—recognition of the cryptotype leads to overly general uses (overgeneralisations); e.g. *tighten* fits the cryptotype just as *tie* does, so the child says **untighten*; or (2) "recovery via cryptotype"—children use the cryptotype to recover from overgeneralisation errors; e.g. *hate* does not fit the cryptotype meaning, and given that only verbs in the cryptotype can take *un-* the child stops saying **unhate*. Both of these possibilities have some empirical evidence in Bowerman's data. However, there is an important question unanswered: How could the child extract the cryptotype and use it as a basis for morphological generalisation or recovery, if the cryptotype is intangible even to linguists like Whorf? (See Whorf's comment on the elusiveness of cryptotype, p. 118.)

In an earlier connectionist model (Li, 1993; Li & MacWhinney, 1996), we hypothesised that cryptotypes seemed intangible because of the limitations of traditional symbolic methods for analysing complex semantic structures. The meanings of a cryptotype constitute a complex semantic network, in which words in a cryptotype can vary in: (1) how many semantic features are relevant to each word; (2) how strongly each feature is activated in the representation of the word; and (3) how features overlap with each other across members in the cryptotype. For example, the verb *screw* in *unscrew* may be viewed as having both the "circular movement" and the "locking" meaning; circular movement is an essential part of the meaning for *screw*, but less so for *wrap*. These complex structural relationships in lexical semantics make a rule-based analysis less effective, if not impossible, but lend themselves naturally to distributed representations and nonlinear processes in neural networks. In this chapter, I further argue that a self-organising neural network can derive cryptotype representations by identifying the complex nonlinear structure from high-dimensional space of language use.

Turning to the acquisition of suffixes, the major empirical findings are that young children show strong patterns of association between verb semantics and morphology in the acquisition of aspect, that is, undergeneralisations in which children restrict morphology to specific categories of verbs. In particular, English-speaking children tend to use the progressive marker *-ing* with process verbs only, whereas they associate the past-perfective marker *-ed* with event verbs. These associations are very strong initially, but weaken over time. Cross-linguistic data suggest similar patterns in children's acquisition of other languages (Li & Shirai, 2000). These patterns prompted some researchers to argue for the existence of innate or prelinguistic categories. In particular, Bickerton (1984) argued strongly that the patterns reflect the functioning of a language bioprogram, in which certain semantic distinctions, for example, distinctions between state and process and between punctual and nonpunctual categories, are hardwired, and that the learner simply needs to find out how they are instantiated in the target language. For example, Brown (1973) observed that English-speaking children do not use the progressive *-ing* with state verbs. To Bickerton, this is strong evidence for the state–process distinction: Children's early use of morphology is to mark bioprogrammed semantic distinctions, not grammatical distinctions.[2]

Thus, the key developmental issue here is whether the empirical patterns reflect innate biases originating from predetermined semantic categories. In this chapter, I present an alternative proposal that rejects the strong version of the nativist argument on innate semantic categories. Earlier discussions (p. 120) have predicted that, in parental input, there are strong associations between verb semantics (lexical aspect) and morphological categories (grammatical suffixes). A further prediction is that children are able to explore the statistical relationships between verbs and morphology in language acquisition. Li and Bowerman (1998) propose that the initial verb–suffix associations could arise as a result of the learner's analyses of the semantics–morphology co-occurrence probabilities in the linguistic input. In the following, I present a connectionist model that implements this proposal. The goal is to demonstrate that a neural network model that draws on realistic linguistic corpus can capture complex semantic structures that are often difficult for symbolic analyses. Through modelling, we can identify more clearly how semantic representations emerge as a function of learning rather than innate hardwiring.

SELF-ORGANISING NEURAL NETWORK AND LANGUAGE ACQUISITION

Modelling semantics in connectionist networks

As mentioned earlier, most previous connectionist models have explored only the formal characteristics, particularly the phonological properties of words. It is relatively straightforward to represent such formal properties, for example, by using acoustic or articulatory features of phonemes (Li & MacWhinney, 2002; MacWhinney & Leinbach, 1991; Miikkulainen, 1997).

[2] A somewhat different, but related view is advocated by Slobin (1985). He suggested the examination of the morphology–semantics mapping by identifying what are "basic" to the learner—constructs that are "prelinguistic" or "privileged" in the initial stages of language acquisition. Slobin's basic child grammar contains a prestructured "semantic space" with universal semantic notions or categories, such as *process* and *result* for the acquisition of tense and aspect. However, because the issue of innateness is less fundamental to the basic child grammar than to the language bioprogram hypothesis, we do not consider it as a nativist theory in this debate. See Li and Shirai (2000) for an analysis of Slobin's perspectives in the context of the nativist–functionalist debate in language acquisition.

It is much more difficult to represent the meaning of words, and thus the modelling of lexical semantics represents a challenge to connectionist language research.

In previous connectionist models involving semantics, researchers have generally constructed semantic representations for a specific set of words on the basis of their own linguistic analyses (Li, 1993; Li & MacWhinney, 1996; MacWhinney, 1998; Ritter & Kohonen, 1989). Alternatively, they use a localist coding to approximate semantics (Cottrell & Plunkett, 1994; Joanisse & Seidenberg, 1999). For example, in our model of the acquisition of prefix, we constructed 20 semantic features for the *un-* verbs, including the general characteristics of actions, relationships between objects, and joint properties of objects that were designed to capture the semantic range of the verbs that can be prefixed with or without *un-* (Li, 1993; Li & MacWhinney, 1996). We presented these features to native speakers of English, and asked them to rate the extent to which a given feature applies to a given verb. A feature-by-verb matrix was derived for each rater, and the mean ratings for each verb became our semantic vectors.

In our modelling, these feature vectors were submitted as input to a feed-forward network with back-propagation learning, and the network's task was to predict which verb could take *un-*, its competitor *dis-*, or no prefix. Two major results were found in our simulations. First, our network formed internal representations of semantic categories that captured Whorf's semantic cryptotypes, on the basis of learning the 20 semantic features. The cryptotype emerged as a function of the network's identification of the relationship that holds between *un-* and the multiple weighted features shared by the *un-* verbs. Our results suggest that in learning of the use of *un-*, the child, like our network, may be computing the combinatorial constraints on the co-occurrences between the prefix, the verb forms, and the semantic features of verbs. Such a process allowed the system to extract a meaningful representation of the *un-* verbs. Second, our network produced overgeneralisation errors similar to those reported in empirical research, for example, **unpress*, **unfill*, and **unsqueeze*. More interestingly, these overgeneralisations were all based on the cryptotype representations that the network developed, indicating clearly that semantic representations served to trigger morphological generalisation. They provided support for Bowerman's (1982) "generalisation-via-cryptotype" hypothesis, but showed no evidence for the "recovery-via-cryptotype" hypothesis.

Although results from these initial simulations are encouraging, the way that semantic features were derived in our model, as in many connectionist models, is subject to the criticism that the network worked precisely because of the use of the "right" features (cf. Lachter & Bever, 1988). It can be argued, for example, that in coding the features the modeller preprocesses the meaning, and what the network receives is very different from what the learner is exposed to in a realistic learning situation. Consideration of this problem led us to look for semantic representations whose actual features are blind to the modeller. High-dimensional space models figure prominently in our search, especially the hyperspace analogue to language (HAL) model of Burgess and Lund (1997, 1999). Li (1999) used HAL semantic representations based on lexical co-occurrence analyses. In HAL, the meaning and function of a given word are determined by lexical co-occurrence constraints in large-scale speech corpora. 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 in the window (occurrence of a word) acting as a constraint to define the meaning of the target word. Global lexical co-occurrence is a measure of a word's total contextual history—what words occur before and after a given word, and how frequently. In this perspective, the semantics of a word can be represented as a vector that encodes the global lexical constraints in a high-dimensional space of language use. Figure 4.1 presents a schematic representation of such vectors: each of the 25-dimension vectors represents the semantics of a word, with each unit representing the degree of a given lexical co-occurrence constraint.

To verify if HAL can be used successfully to capture the acquisition of word meaning, Li, Burgess, and Lund (2000) analysed 3.8 million word tokens from parental speech in the CHILDES English database (MacWhinney, 2000). We found that the HAL method can derive accurate lexical semantic representations, given a reasonable size of speech such as our CHILDES parental speech (rather than a huge amount of speech such as the Usenet data for the original HAL model). The implication of our study is that young children can acquire word meanings if they exploit the considerable amount of contextual information in the linguistic input by computing multiple lexical co-occurrence constraints. The limitation of the study is that no learning was involved in the representation of word meanings, as it was based purely on extraction of statistical information. I will return to this

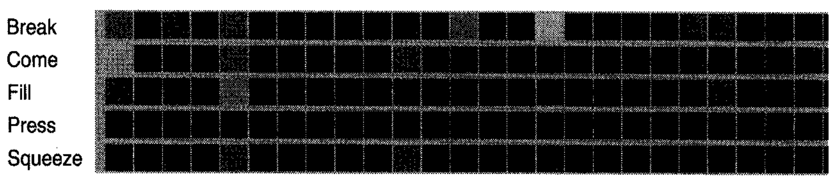

Figure 4.1. A grey-scale representation of HAL vectors for five words. Each dimension of the 25-unit vector represents the degree of a given lexical co-occurrence constraint. A high degree of constraint is represented as white or grey, and a low degree of constraint as dark or black (on a continuous scale from 0 = all black, to 1 = all white).

point in the Conclusions section, where I suggest a developmental learning model of the HAL type.

Self-organising feature maps and language representation

Like most previous connectionist models of language acquisition, the model of Li (1993) and Li and MacWhinney (1996) was based on the standard back-propagation learning algorithm. Although significant progress has been made with models based on back-propagation, there are some known limitations associated with these models (see the Introduction section). Some of these problems become most transparent when considered in the context of language acquisition. For example, a strong assumption has been made that the language learner can be considered as a "hypothesis generator": each time the learner hears some linguistic information, he or she will compare it with existing knowledge and make a guess as to what should be correct in the target language (Plunkett & Juola, 1999). However, there is so far no psychological evidence that the language learner is a hypothesis generator of this nature (i.e. as a back-propagation machine). Children do not receive constant feedback about what is incorrect in their speech, or the kind of error corrections on a word-by-word basis as provided to the network (consider the "no negative evidence problem" in language acquisition; see Baker, 1979; Bowerman, 1988). The gradient descent mechanism used in back-propagation also leads to other problems, for example, local minima, the problem that the network is entrapped into a local landscape and unable to move to the global error minimum (Hertz, Krogh, & Palmer, 1991).

Consideration of these problems led us to look for models that bear more biological and psychological plausibility. We turned to a class of self-organising neural networks, the self-organising feature maps (SOFMs). SOFMs belong to the class of "unsupervised" neural networks, because learning in these networks does not require the presence of a supervisor or an explicit teacher; learning is achieved by the system's self-organisation in response to the input. During learning, the self-organising process extracts an efficient and compressed internal representation from the high-dimensional input space and projects this new representation onto a two-dimensional space (Kohonen, 1982, 1989, 1995). Several important properties of SOFMs and related features make them particularly well suited to the study of language acquisition. We briefly discuss three of them here and their implications for language acquisition.

(1) *Self-organisation.* Self-organisation in these networks typically occurs in a two-dimensional topological map, where each unit (or "neuron") 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 that unit and its neighbouring 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 so-called activity "bubbles"), and the maximally active units are taken to represent the input. Initially activation occurs in large areas of the map, but gradually learning becomes focused so that only the maximally responding unit or units are active. This process continues until all the inputs have found some maximally responding units.

(2) *Representation.* As a result of this self-organising process, the statistical structures implicit in the high-dimensional input space are represented as topological structures on the two-dimensional space. In this new representation, similar inputs will end up activating the same units in nearby regions, yielding meaningful activity bubbles that can be visualised on the map. The self-organising process and its representation have clear implications for language acquisition: The formation of activity bubbles may capture critical processes for the emergence of semantic categories in children's acquisition of the lexicon. In particular, the network organises information first in large areas of the map and gradually zeros in onto smaller areas; this zeroing-in is a process from diffuse to focused patterns, as a function of the network's continuous adaptation to the input structure. This process allows us to model the emergence of semantic categories as a gradual process of lexical development. It naturally explains many generalisation errors reported in the child language literature: for example, substitutions errors (e.g. *put* for *give*, *fall* for *drop*; Bowerman, 1978) often reflect the child's initial recognition of diffuse similarities but not fine-grained distinctions between the words. It also explains language disorders that result from the breakdown of focused activation or the inability to form focused representations (Miikkulainen, 1997; Spitzer, 1999).

(3) *Hebbian learning.* Hebbian learning is not an intrinsic feature of a SOFM, but several SOFMs can be connected via Hebbian learning, such as in the multiple feature-map model of Miikkulainen (1993, 1997). Hebbian learning is a well-established biologically plausible learning principle, 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 may be proportional to the level of activation of the two neurons. In a multiple SOFM model, all units on one map are initially connected to all units on the other map. As self-organisation takes place, the associations

become more focused, such that in the end only the maximally active units on the corresponding maps are associated. Hebbian learning combined with SOFMs has strong implications for language acquisition: It can account for the process of how the learner establishes relationships between word forms, lexical semantics, and grammatical morphology, on the basis of how often they co-occur and how strongly they are co-activated in the representation.

Thus, models based on the above properties: (1) allow us to track the development of the lexicon clearly as an emergent property in the network's self-organisation (from diffuse to focused patterns or from incomplete to complete associative links); (2) 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 (3) provide us with a set of biologically plausible and computationally relevant principles to study language acquisition without relying on negative evidence to learn. They are biologically plausible because the human cerebral cortex can be considered as essentially a self-organising map (or multiple maps) that compresses information on a two-dimensional space (Kohonen, 1989; Spitzer, 1999), and computationally relevant because language acquisition in the natural setting (especially organisation and reorganisation of the lexicon) is largely a self-organising process that proceeds without explicit teaching (MacWhinney, 1998, 2001).

A number of studies have employed SOFMs for language research. An earlier attempt was made by Ritter and Kohonen (1989), who constructed a network that takes semantic features of animals (e.g. *small-size*, *has hair*, *can fly*) and organises them on a feature map. In the input, each animal was represented as a combination of these features in a feature vector and, after 2000 epochs of self-organisation, the network developed meaningful representations of types of animals. Wild predators (e.g. *tiger*, *lion*, *wolf*) were grouped together on one area of the map, whereas birds (e.g. *hawk*, *owl*, *goose*) were grouped nearby on another area. Within each group, similar animals were closer to each other than were dissimilar ones. Although Ritter and Kohonen's model used only a dozen or so animal words with a highly idealised feature representation, their results showed that interesting semantic structures could develop from the network's self-organisation of relevant features, and that the new representations in a SOFM can correspond closely to the hierarchical structure of human conceptual relationships.

Miikkulainen's (1993) research represents another important step in using SOFMs for language research. He proposed an integrated model of memory and natural language processing, in which multiple SOFMs dedicated to different levels of information are connected. A subcomponent of this model is DISLEX (Miikkulainen, 1997), a SOFM model of the lexicon. In DISLEX, different maps correspond to different linguistic information (orthography, phonology, or semantics) and 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 maps, and the resulting bubble of activity propagates through the associative links and causes an activity bubble to form in the other map. The activation of co-occurring form–meaning representations leads to adaptive formations of the associative connections between the maps. DISLEX successfully models the mental lexicon in normal and disordered language processing. Miikkulainen showed that in a lesioned SOFM, behaviours of dyslexia (e.g. producing *dog* in reading *sheep*) can result from partial damage to the semantic representation. The network also displayed behaviour of surface dyslexia (e.g. producing *ball* in reading *doll*), which results from partial damage to the form representations.

MacWhinney (2001) further considered the use of SOFMs in the domain of lexical acquisition. A normal English-speaking child, starting from the age of 2, learns an average of nine words per day, ending up with an active vocabulary of about 14,000 words by age 6 (Carey, 1978). Apparently, this size of lexicon exceeds the capacity of most current connectionist models (the "scalability" problem in connectionism). To answer this challenge, MacWhinney trained two feature maps to associate with each other, one representing lexical semantics, and the other phonological features. In a simplified scheme, the phonology or semantics of an input was represented by four units with random values. The two maps were associated through Hebbian learning, as in the DISLEX model. It was found that a network with 10,000 nodes was able to learn the form–meaning associations of up to 6000 words, with an average error of less than 1 per cent. MacWhinney suggested that it would be possible to increase the size of the feature map to learn more words and that given the enormous number of cells in the human brain, the size of the feature map is not an important limiting constraint on lexical acquisition by children.

The above studies all attest to the utility and importance of self-organising neural networks in language research. However, they suffer from the same problems we discussed earlier, either because the semantic representations were too simplified, or because the target lexicon of the model was too small and unrealistic, or both. In considering these problems, Li (1999, 2000) explored SOFMs as a feasible model of language acquisition in the context of lexicon and morphology. In what follows, I will present a sketch of the model and its implications for language acquisition; for details, see Li (1999, 2000), and Li and Shirai (2000, Chapter 7).

A SOFM MODEL OF LEXICAL AND MORPHOLOGICAL ACQUISITION

On the basis of the discussions above, I present two modelling studies: (1) the acquisition of semantic cryptotypes of verbs in the context of derivational prefixes; and (2) the acquisition of lexical aspect of verbs in the context of inflectional suffixes. In each case, the model simulates the development of the lexicon and morphology in young children. The goal is to show: (1) how a SOFM model can capture processes of semantic organisation that lead to distinct semantic categories, categories that have been claimed to be either intangible (e.g. cryptotypes) or innate (e.g. state and result); and (2) how such a model can derive semantic–morphological associations as observed in child language, on the basis of analysing distributional information in realistic linguistic data. Evidence from such a model should provide insights into psycholinguistic mechanisms underlying lexical and morphological acquisition.

The model consisted of two SOFMs, each of the size of 25 × 25 units, one for the organisation of lexical form, including the phonology of verb stems and affixes (the lexical map), and the other for the organisation of semantic information (the semantic map). Because the simulation of suffixes involved twice as many verbs as the simulation of prefixes, the size of the maps for the suffix model was correspondingly expanded (50 × 50 units). Figure 4.2 illustrates the model diagrammatically.

Method

Input and representation. As we model the acquisition of prefixes and suffixes, the input data to our network consist mainly of lexical representations of verbs with which the affixes co-occur. In the case of prefixes, we selected 228 verbs according to the *Webster's New Collegiate Dictionary* and the Francis and Kucera (1982) corpus. The 228 verbs for prefixes include 49 *un-* verbs, 19 *dis-* verbs, and 160 verb stems with no prefixes. In the case of suffixes, we selected 562 verbs from the CHILDES parental corpus (see Li & Shirai, 2000) with the following criterion: A verb was included in our training

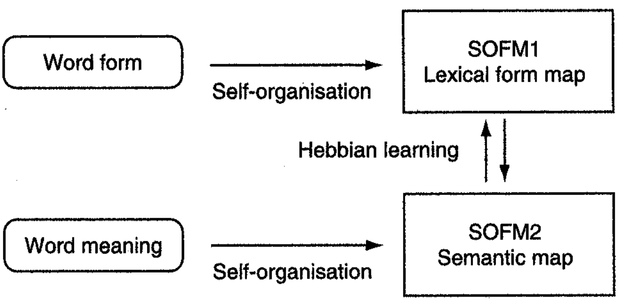

Figure 4.2. A SOFM model of lexical and morphological acquisition. The model consisted of two SOFMs: one self-organises on lexical form (the lexical map), and the other self-organises on word meaning (the semantic map). The associations between the two maps are trained via Hebbian learning.

data if it occurred in the parental corpus for five or more times at a given age period (see the Stages of training section below).

To represent the phonology of the verbs, we used a syllable-based template coding developed by MacWhinney and Leinbach (1991). This coding scheme has the advantage over traditional phonemic representations in that it can more accurately capture phonological similarities of multisyllabic words. A word's representation is made up by combinations of syllables in a metrical grid, and the slots in each grid are made up by bundles of features that correspond to phonemes, *C*s (consonants) and *V*s (vowel). For example, the 18-slot template *CCCVV CCCVV CCCVV 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 10 feature units, and each *V* by 8 feature units, making a total of 168 units for each phonological vector (see Li & MacWhinney, 2002, for a more recent version of this representation).

The semantic representations of verbs to our network were based on lexical co-occurrence analyses in the HAL model (Burgess & Lund, 1997). As discussed earlier, HAL measures the semantics of a word by its total contextual history, encoded as a vector that represents multiple lexical co-occurrence constraints from large-scale corpora. Of course, not all lexical constraints contribute equally to the representation, so we extracted 100 components that have the greatest contextual diversity as the appropriate vector dimensions (see Lund & Burgess, 1996, for details). Thus, each semantic representation is formed by a 100-unit vector.

Task and procedure. Upon training of the network, a phonological representation of a verb was input to the network and, simultaneously, the semantic representation of the same verb was also presented to the network. By way of self-organisation, 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 compatible with a given affix in the language (prefix) or in the input speech (suffix), the phonological representation of the affix was also coactivated with the phonological and the semantic representations of the verb stem. As the network received input and continued to self-organise, it simultaneously learned associations between maps through Hebbian learning: Initially, all the units on one map were fully connected to all the units on the other map; as learning continued, only the units that were coactivated in response to the input were associated. If the direction of the associative propagation goes from phonology to semantics, *comprehension* is modelled; if it goes from semantics to phonology, *production* is modelled. As the goal of learning, the network should create new representations in the corresponding maps for all the inputs and link the semantic properties of a verb to its phonological shape and morphological pattern. All simulations were conducted with the DISLEX simulator (Miikkulainen, 1999).

Stages of training. To observe effects of the interaction between lexicon and morphology in learning, we designed four stages to train the network. In the case of prefixes, a given verb is paired with *un-*, *dis-*, or zero-marking according to whether the prefixation is allowed in the adult language. In the case of suffixes, a given verb is paired with *-ing*, *-ed*, *-s*, or zero-marking according to whether the verb co-occurs with the suffix in the parental speech.

For prefixes, the four stages were: (1) the phonological representation of a verb stem was coactivated with its semantic representation on a one-to-one basis in the input. This was done to model the whole-word learning stage—a stage at which children have not analysed morphological devices as separate entities from the verb stems (Bowerman, 1982); (2) phonological and semantic representations of verb stems (e.g. *tie*, *connect*), prefixed verbs (*untie*, *disconnect*), and the prefixes themselves (*un-*, *dis-*) were all coactivated in the input; (3) 25 novel verbs were introduced to the network to test whether generalisations would occur in our network as in children's speech. These were verbs on which previous studies have reported children's generalisations (Bowerman, 1982; Clark et al., 1995). Generalisation was tested by inputting the verbs to the network without having the network self-organise or learn the phonological–semantic associations; (4) self-organisation and Hebbian learning resumed for the novel verbs introduced at stage 3 to test if the network could recover from generalisations.

For suffixes, the four stages were based on the age groups of the input data (i.e. the age of the child for which adult input was available in our corpus—the input age). The four stages were: (1) input age 1;6 (13–18 months). Relatively few uses of suffixes occur in the CHILDES parental data before the child is 13 months old. For the period of 13–18 months, a total of 186 verbs fit our selection criteria (i.e. occurred five or more times); (2) input age 2;0 (19–24 months) included 324 verbs; (3) input age 2;6 (25–30 months) included 419 verbs; and (4) input age 3 (31–36 months) included 562 verbs. These stages reflect an incremental growth of vocabulary, and the verbs at a later stage always included verbs at the previous stage. It also reflected a coarse frequency coding: a verb or a suffix was presented to the network for the number of times it occurred across the four stages.

In the following sections, I report two sets of simulation results, one for prefixes, and the other for suffixes. However, the acquisition patterns are comparable for both types of morphology, to which we will return in the Conclusions section.

Results and discussion: Prefix simulations

In this section, I focus on three levels of analysis on the prefix simulations: the network's representation of the cryptotype, its patterns of overgeneralisation, and its ability to recover from the generalisation errors.

Representation of cryptotype. One of the major motivations for this study was whether neural networks can develop structured representation as a function of its self-organisation on verb semantics. In particular, I wanted to see how the patterns of activity formed in the semantic map can capture Whorf's covert, "intangible", category of cryptotype.

In Li and MacWhinney (1996) we suggested that there are several "mini-cryptotypes" that work collaboratively as interactive "gangs" (McClelland & Rumelhart, 1981) to support the formation of a larger cryptotype. For example, "enclosing" verbs, such as *coil*, *curl*, *fold*, *ravel*, *roll*, *screw*, *twist*, and *wind*, all seem to share a meaning of circular movement; another set of verbs such as *cover*, *dress*, *mask*, *pack*, *veil*, and *wrap* form the "covering" mini-cryptotype, and so on. Members in these mini-cryptotypes are closely related by overlapping semantic features. Previously, we have used hierarchical cluster analyses to identify the existence of mini-cryptotypes in our network, by analysing the hidden-unit activation patterns. In the current study, these mini-cryptotypes can be seen more clearly in the emerging structure of the SOFM's two-dimensional layout as activity bubbles. In our network, the self-organisation process extracted semantic structures from the input and projected the new representations on the semantic map. Figure 4.3 presents a snapshot of the network's representation after it was trained on 120 verbs for 600 epochs at stage 1.

A close examination of the semantic map shows that the network developed clear representations that correspond to the cryptotype which Whorf believed governs the use of *un-*. Our network, without using *ad hoc* semantic features, mapped members in mini-cryptotypes onto nearby regions of the SOFM. For example, towards the lower right-hand corner, verbs like *lock*, *clasp*, *latch*, *lease*, and *button* were mapped to the same region, and these verbs all share the "binding/locking" meaning. Similarly, "attachment" verbs like *snap*, *mantle*, *tangle*, *ravel*, *tie*, and *bolt* occurred towards the lower left-hand corner, and verbs of perceptions and audition like *hear*, *say*, *speak*, *see*, and *tell* can be found in the upper left-hand corner. One can also observe that *embark*, *engage*, *integrate*, *assemble*, and *unite* are being mapped towards 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-organisation is moving from diffuse to more focused patterns of activity; for example, the verb *show*, which shares similarity with

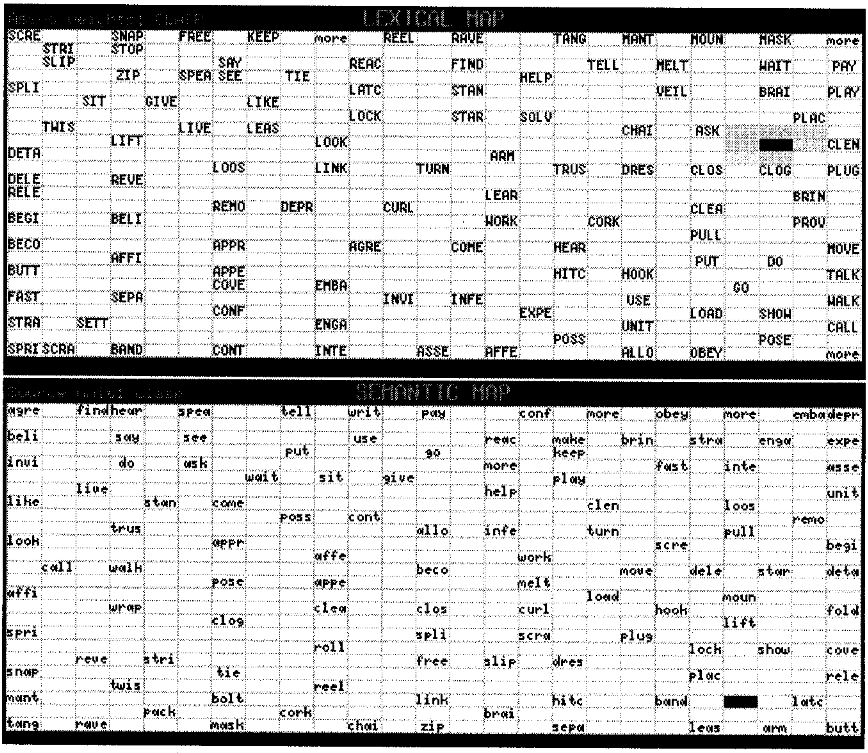

Figure 4.3. The representation of semantic cryptotypes in a SOFM. The upper panel is the lexical map, and the lower panel the semantic map. The phonological representations of words are in capitals, and the semantic representations in lower case. Words longer than four letters are truncated.

none of the above, is grouped with the binding/locking verbs. What is crucial, however, is that these clusters form the semantic basis for the overall cryptotype of the *un-* verbs. As shown in Figure 4.3, the network has mapped most verbs in the cryptotype to the bottom layer of the semantic map, and importantly, these are the verbs that can take the prefix *un-*.

These results from our model offer a tangible solution to the "intangible" aspects of Whorf's cryptotype. Connectionist learning provides us with a natural way of capturing Whorf's insights of cryptotype as well as its acquisition in a formal mechanism. It gives a precise account of how the *un-* cryptotype emerges from learning in a distributed representation: The formation of a cryptotype is supported by mini-cryptotypes that interact collaboratively, which are in turn supported by multiple weighted features shared by all the *un-* verbs through summed activation.

Representation and overgeneralisation. Connectionist networks can generalise learned patterns to novel instances, but do they show the same types of generalisation as children do? And on what basis do they generalise?

In our network, as discussed earlier, the two SOFMs can be connected via Hebbian learning: the phonological and semantic representations of a verb are coactivated in different maps, along with the corresponding prefixes that the verb can take in the language. Hebbian learning determines how strong the connections between the phonology, the meaning, and the affix should be during each stage of the learning. At the same time, the two maps also self-organise. In this way, Hebbian learning and self-organisation provide the network with focused pathways from form to meaning and from meaning to form. Thus, when the network receives a new input, it can readily "comprehend" the input (from form to meaning) or "produce" the input (from meaning to form) using its existing, learned pathways between the feature maps. This procedure also allows us to test the network's generalisation ability when meaningful representations have emerged from the maps.

The simulation results indicate that our network was not only able to capture the elusive cryptotype by way of self-organisation, but also able to generalise on the basis of this representation. For example, when tested for generalisation at stage 3, the network produced overgeneralisation errors (e.g. **unbreak*, **uncapture*, **unconnect*, **ungrip*, **unpeel*, **unplant*, **unpress*, **unspill*, **untighten*) that match up with empirical data. These overgeneralisations were based both on the network's established structure of semantic representations and on the associative connections that it formed in learning the meaning–form mappings. Several observations can be made on the network's overgeneralisations.

First, most of these overgeneralisations involve verbs that fall within the Whorfian cryptotype (e.g. *connect*, *grip*, *peel*, *plant*, *press*, *spill*, and *tighten*). Earlier, we pointed out two hypotheses regarding the role of the cryptotype in children's acquisition of *un-* according to Bowerman: "generalisation via cryptotype" and "recovery via cryptotype". Our results here are consistent with the first hypothesis, that is, the representation of cryptotype leads to overly general uses of *un-*. Consistent with our previous simulations, we found no violations of the cryptotype in the network's overgeneralisations such as **unhate* or **untake* (as found in Bowerman's data); hence there was no evidence for the hypothesis that the learner can use the cryptotype representation to recover from overgeneralisations.

Second, the associative pathways between the two maps formed via Hebbian learning provide the basis for the production of overgeneralisations. 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 coactivated, and the phonology of *clench*, *unclench*, and *un-* were also coactivated. When the semantics and the phonology of these items were associated through Hebbian learning, the network can associate the semantics of *tighten* with the phonology of *un-* because of *clench*, even though the network learned only the association of *un-clench* and not *un-tighten* (i.e. at an earlier stage *tighten* was withheld from the training). This associative process of correlating semantic features, lexical forms, and morphological devices simulates the process of learning and generalisation in children's language acquisition, and shows that overgeneralisations can arise naturally from structured semantic representations (a result of self-organisation) and from associative learning of meanings and forms.

Finally, overgeneralisations were not limited to morphological generalisations. There were lexical generalisations similar to those reported by Bowerman (1982) and Miikkulainen (1997). For example, the 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 generalisations resemble lexical errors in surface dyslexia. Similarly, the network comprehended *see* as *speak*, *arm* as *clasp*, and *unscrew* as *hook*, due to its representation of these pairs of words in nearby regions in the semantic map, and these errors resemble lexical errors in deep dyslexia in reading comprehension. They demonstrate further the intimate relationship between semantic representation and generalisation. Again, self-organisation and Hebbian learning account for the origin of this type of generalisation errors.

Mechanisms of recovery from generalisations. Can our self-organising network recover from generalisations as children do? If so, what computational mechanisms permit its recovery?

Our network displayed significant ability to recover from overgeneralisations. When tested for generalisations at stage 3, no learning took place in the network for self-organisation or associative connection. When tested for recovery at stage 4, self-organisation and Hebbian learning resumed. Within 200 epochs of new learning during the last stage, the network recovered from the majority of the overgeneralisations tested at stage 3. Recovery in this case is a process of restructuring the mapping between phonological, semantic, and morphological patterns, and this restructuring is based on the network's ability to reconfigure the associative pathways through Hebbian learning, in our case, the ability to form new associations between prefixes and verbs and the ability to eliminate old associations that were the basis of erroneous overgeneralisations. When a given phonological unit and a given semantic unit have fewer chances to become coactivated, Hebbian learning decreases the strengths of their associative links. For example, *un-* and *tighten* were coactivated because of *un-* and *clench* at stage 3; at stage 4 *un-* and *clench* continue to be coactivated, but *un-* and *tighten* are not coactivated. Hebbian learning determines that the associative connection between *un-* and *clench* continues to increase as learning progresses, but that between *un-* and *tighten* gets decreased and eventually eliminated, thereby simulating what happens at the final phase of U-shaped learning. This result models the process in which children's overgeneralisations are gradually eliminated when there is no auditory support in the input about specific co-occurrences that they expect (MacWhinney, 2001). In the realistic learning situation, the strength of the connection between *un-* and inappropriate verbs may also be reduced by a competing form such as *loosen* that functions to express the meaning of **untighten*. This type of process is often discussed in the literature as the pre-emption mechanism (Clark, 1987) or the competition mechanism (Bates & MacWhinney, 1987; MacWhinney, 1987). Our model has not yet incorporated this type of mechanism.

Hebbian learning coupled with self-organisation provides a simple but powerful computational principle to account for the recovery process. Restructuring of associative links often goes hand-in-hand with the reorganisation of the 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 proposal that children recover from generalisations by recognising fine and subtle semantic and phonological properties of verbs (Pinker, 1989). Interestingly, in cases where it did not recover from overgeneralisations, the network had difficulty making fine semantic distinctions. For example, because it was unable to separate word pairs like *press* and *zip* in the semantic map, it continued to produce erroneous forms like **unpress*.

An additional parameter that we considered in the SOFM's error recovery was the size of the feature map (i.e. the number of units available for learning). The inability to further distinguish semantically similar words might be due to resource limitations. To verify this hypothesis, in a separate but otherwise identical simulation, we doubled the size of both maps (from 25×25 units to 50×50 units). In this new simulation, at stage 3 we continued to observe the same type of overgeneralisations as in the original simulations, but at stage 4 the network recovered completely from all the overgeneralisations. Thus, there is reason to believe that enough learning resource is needed for the network to further reorganise confused items that are due to great similarity. For the child, it is likely that the increasing capacity of memory and other cognitive abilities make resource limitation a nonproblem. We could model this type of resource increase with an architecture in which the number of neurons dynamically grows in response to the learning task (see Farkas & Li, 2002, for a recent implementation). This type of dynamic growth of SOFMs could be compared to the cascade correlation mechanism in back-propagation learning (Fahlman & Lebiere, 1990).

Results and discussion: Suffix simulations

The simulation procedures for the suffixes were similar to those for the prefixes except the training materials and stages. We also used larger maps (50 × 50 units) given the resource problem considered above and given that twice as many verbs were involved in the suffix simulations as in the prefix simulations. Below, I focus on three levels of analysis for the suffix simulations: the role of input, the emergence of lexical aspect categories, and the formation and relaxation of strong associations between lexical semantic categories and grammatical suffixes.

Role of input. One important rationale behind the current modelling effort is the understanding of the role of linguistic input in guiding children's acquisition of lexical and grammatical aspect. The relationship between patterns observed in children's speech and those in parental speech with respect to the interaction between verb semantics and aspect suffixes has been emphasized elsewhere (Li & Bowerman, 1998; Li & Shirai, 2000) but a simple correlation between children's and adults' patterns tells us only that the child is sensitive to the linguistic environment and is able to incorporate information from that environment into his or her own speech. It does not tell us how the child actually does the analysis, or what mechanisms allow the child to do the analysis. Thus, we need to test if a connectionist network—endowed with self-organisation and Hebbian learning principles—is able to display learning patterns as found in child language. If so, we can conclude that self-organisation and Hebbian learning may provide the necessary kinds of mechanisms that allow for the formation of patterns in language acquisition. In this way, our modelling enterprise provides insights into the mechanisms underlying the learning process.

Table 4.1 presents a summary of the major patterns from the network's learning according to the tense-aspect suffixes it produced at the different learning stages. It shows the results of the network's production of three suffixes, *-ing*, *-ed*, and *-s* with three types of verbs, processes, events, and states. The results are based on the unit activations on the phonological map that each verb in the semantic map activated, after the network had been trained for 200 epochs at each stage.

The results in this table are highly consistent with empirical patterns observed in early child language: the use of the progressive aspect (marked by *-ing* in English) is closely associated with process verbs that indicate ongoing processes, while the use of past-perfective aspect (marked by *-ed* in English) is closely associated with event verbs that indicate endpoints or end results. Some studies also suggest a strong association between the habitual *-s* and state verbs (Clark, 1996). Our network, having received input patterns based on parental data, behaved in the same way as children do. For example, at

TABLE 4.1
Percentage of use of tense-aspect suffixes with different verb types across input age groups in the network's production and in the parental input data*

	Tense-aspect suffixes											
	Age 1;6			*Age 2;0*			*Age 2;6*			*Age 3;0*		
Verbs	*-ing*	*-ed*	*-s*	*-ing*	*-ed*	*-s*	*-ing*	*-ed*	*-s*	*-ing*	*-ed*	*-s*
Network production												
Processes	75	18	0	66	16	0	64	26	0	52	9	10
Events	25	82	0	28	84	0	31	74	0	44	77	10
States	0	0	100	0	0	100	0	0	100	4	14	80
Item totals**	40	11	3	71	19	9	89	19	7	70	22	10
Parental input data												
Processes	69	22	0	74	15	17	67	23	20	67	23	23
Events	28	77	33	24	77	0	25	69	20	31	65	8
States	3	0	67	2	8	83	8	8	60	2	12	69
Item totals	29	3	9	54	13	12	40	13	10	60	26	13

* The table includes only verbs that could be uniquely assigned to one or the other suffixation pattern and does not include instances for which the network produced a given verb with multiple suffixes. See Table 4.2 for the latter.
** These are the total number of verbs that occurred with the given suffix. Note that the percentages within a given column do not always add up to 100, reflecting that some verbs could not be easily classified into one or the other category. This is also true for Table 4.2.

input age 1;6, the network produced *-ing* predominantly with process verbs (75 per cent), *-ed* overwhelmingly with event verbs (82 per cent), and *-s* exclusively with state verbs (100 per cent). Such associations remained strong at input age 2 but gradually became weaker (although still transparent) at later stages.

Interestingly, when we analysed the actual input to our network (based on parental speech), we found similar patterns. Table 4.1 also presents the percentages of the use of suffixes with different verb types in the input data. The degree to which the network's production matches up with the input patterns (Table 4.1) indicates that our network was able to learn on the basis of the information of the co-occurrences between lexical aspect (verb types) and grammatical aspect (use of suffixes). This learning ability was due to the network's use of Hebbian associative learning in computing: (1) when the semantic, phonological, and morphological properties of a verb co-occur; and (2) how often they do so.

The results in Table 4.1 also match up nicely with several empirical studies that have examined the correspondence between children's speech and adult input in the acquisition of tense–aspect suffixes, in English (Shirai & Andersen, 1995), Japanese (Shirai, 1998), modern Greek (Stephany, 1981), and Turkish (Aksu-Koç, 1998). Note that the patterns in the input, as discussed by Li and Shirai (2000), are usually less absolute or restrictive than in children's early productions, showing that adults are more flexible in associating various types of grammatical morphology with various types of verbs. Indeed, the patterns in Table 4.1 show that the associations between verb types and suffixes are weaker in the input to the network than they were in the network's production. This is important, because if the learner—child and network alike—simply mimicked what is in the input, the learner would have no productive control over the relevant linguistic problem and would simply produce the patterns verbatim. The modelling results further confirm the hypothesis that a probabilistic pattern in the input can lead to more absolute patterns in the learner's output, because the learner initially capitalises on the prototypical representations of the verb–suffix association (see Li & Shirai, 2000, for the role of input in inducing prototypes).

Emergence of semantic categories of lexical aspect. Figure 4.4 presents a snapshot of the network's self-organisation of the semantic representations of verbs at input age 1;6 (from the semantic map). The network clearly developed structured semantic representations that correspond to categories of lexical aspect such as processes, events, and states. For example, towards the lower right-hand corner, state verbs like *feel*, *know*, *think*, *remember*, *wonder*, *love*, and *like* were mapped onto the same region of the map. Event verbs can be found in the middle-to-left portion of the map, including verbs like *catch*, *fix*, *break*, *knock*, *grab*, and *throw*, all of which indicate actions that lead to clear end results. Process verbs can be found spanning the upper end of the map, including (from left to right) *rub*, *scrub*, *sleep*, *shout*, *laugh*, *drink*, *walk*, *kiss*, *cry*, *swim*, *dance*, and so on.

The above patterns of semantic neighbourhood bear close similarity with the formation of mini-cryptotypes in the case of prefixes. As discussed earlier, the formation of semantic categories goes hand-in-hand with the acquisition of grammatical morphology. On the one hand, similar verbs form concentrated patterns of activity, providing the basis for semantic categories, and on the other hand, they also form focused associative pathways to the phonological and morphological representations of verbs in the other map. When concentrated activities occur both horizontally (within a two-dimensional map) and vertically (across the maps), the semantic categories of lexical aspect will behave like magnets to connect the lexicon to morphology. Thus, when a new input has semantic overlap with verbs of an existing lexical category and resembles members of that category, its mapping to corresponding morphemes will be readily done through the existing associative pathways going from verb semantics to suffixes; that is, no additional learning will be needed for the new mapping. This analysis provides a mechanistic account for

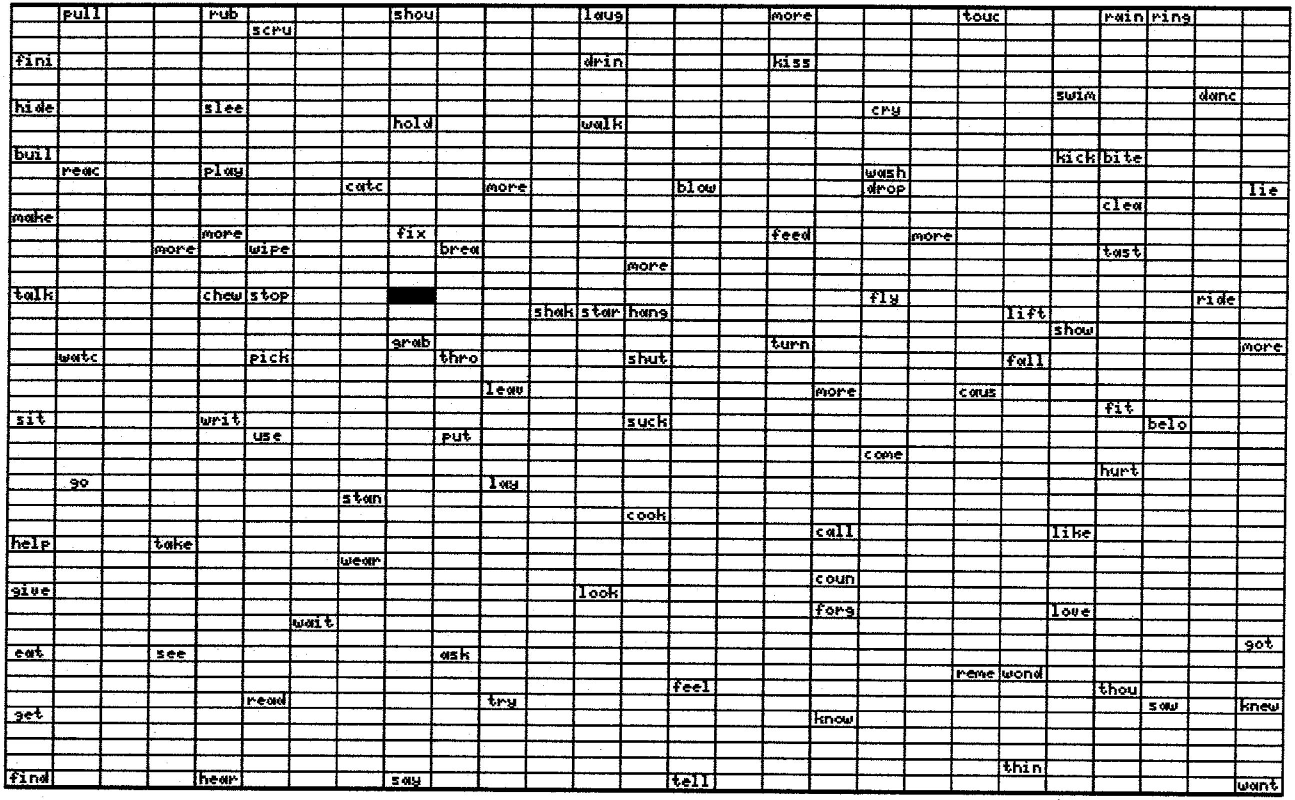

Figure 4.4. The representation of verb categories in a SOFM (semantic map only). Only the left portion of the complete map is shown due to space limit. Words longer than four letters are truncated.

Slobin's (1985) basic child grammar hypothesis that the initial semantic categories act as magnets to attract grammatical mappings in the input language.

From strong associations to diverse mappings.. As with the prefix simulations, the associative pathways between forms and meanings are established via Hebbian learning across learning stages. Depending on how often forms and meanings co-occur, Hebbian learning establishes either stronger or weaker associations. Thus, when the network has a focused pathway, for example, between *-s* on the lexical map and state verbs on the semantic map, it can readily "comprehend" new state verbs at no additional learning, promoting an even stronger state-to-*s* association (a prototypical association). However, as learning progresses, *-s* may be used more diversely with other verb types in the input, so that the prototypical association weakens over time. The fact that a given suffix occurs with multiple verbs, and a given verb occurs with multiple suffixes in the input tells the system that it should no longer be restricted to the prototypical associations, but develop new nonprototypical mappings between lexicon and morphology.

Table 4.2 presents the same simulation results as in Table 4.1, except that multiple suffixation patterns are included here—a given verb was counted for multiple number of times in the table depending on the number of suffixes with which it co-occurred (Table 4.1 included only verbs that could be uniquely assigned to one suffixation pattern; see Li & Shirai, 2000, for the rationale behind this treatment).

A comparison of this table with Table 4.1 reveals that, for the early stages (1;6 and 2;0), the two tables are very similar; for the later stages, however, they become more distinct, mainly with respect to the uses of *-ed* and *-s*. Detailed analyses show that over 50 per cent of all suffixed verbs had more than one

TABLE 4.2
Percentage of use of tense-aspect suffixes with different verb types across input age groups in the network's production (multiple suffixations)*

	Tense-aspect suffixes											
	Age 1;6			*Age 2;0*			*Age 2;6*			*Age 3;0*		
Verbs	*-ing*	*-ed*	*-s*	*-ing*	*-ed*	*-s*	*-ing*	*-ed*	*-s*	*-ing*	*-ed*	*-s*
Processes	72	16	0	62	29	6	64	40	44	52	38	30
Events	28	75	0	32	66	31	32	60	12	43	53	26
States	0	8	100	0	4	63	0	0	44	5	9	44
Item totals*	43	12	5	81	24	16	114	35	16	121	64	27

* Because the same verb could occur in more than one column, the sum of the item totals across columns does not equal the total word types. This differs from the interpretation of item totals in Table 4.1.

suffix at input age 3;0, compared with only 5 per cent at input age 1;6. These results suggest that multiple suffixations might be the driving force for the learner to break from the strong associations to more diverse mappings. There was relatively little change with the *-ing* verbs, because the majority of the early verbs were process verbs that take *-ing*. Overall, these results indicate that increasing associative links between verbs and suffixes (along with incremental vocabulary growth) lead to diverse mappings, first with some words and then spreading to others, thus accounting for how the strong associations weaken over time in children's language.

Our simulation results also shed some light on the acquisition of the English past tense. First, given that children's early use of *-ed* is restricted to specific lexical meanings, overgeneralisations of *-ed* would not occur across the board for all types of verbs but will rather be restricted to event verbs initially. Second, overgeneralisations of *-ed* not only may be semantically restricted, but also sometimes semantically motivated. In our network, semantic pathways formed via Hebbian learning provide the basis for the production of overgeneralisation errors. For example, *knock* and *break* share semantic similarities and were mapped onto nearby regions in the semantic map. During learning, the semantics of *knock* and *knocked* were coactivated, and the phonological forms of *knock*, *knocked*, and *-ed* were also coactivated. When the semantics and the phonology of these items were associated via Hebbian learning, the network would connect the semantics of *break* with the phonology of *-ed* because of *knock*, even though it learned only the association for *knock-ed* and not *break-ed* (i.e. when *break* was withheld from training initially). This result parallels the overgeneralisation errors on prefixes such as **un-tighten* (due to *un-clench*) that we discussed earlier.

CONCLUSIONS

In this chapter, I started by reviewing some of the problems in previous connectionist models of language acquisition. I pointed out that previous models have been largely restricted to the examination of phonological patterns (in contrast to semantic structures), to the use of artificially generated input (in contrast to realistic linguistic data), and to the use of supervised learning algorithms (in contrast to unsupervised learning). I proposed a new connectionist model of language acquisition that is based on the examination of the acquisition of semantics, with exposure to realistic child-directed parental data, and in self-organising networks with unsupervised learning. I showed how SOFMs can be used successfully to model the acquisition of lexicon and morphology. In particular, I applied SOFMs to examine two linguistic domains where the development of lexicon and morphology is crucial: (1) the acquisition of derivational prefixes with respect to semantic cryptotypes of verbs; and (2) the acquisition of inflectional suffixes with respect to grammatical and lexical aspect of verbs. The new model sheds light on issues of semantic representation, morphological overgeneralisation and undergeneralisation, and recovery from erroneous generalisations in humans and networks. I argue that self-organising neural networks coupled with Hebbian learning provide computationally relevant and psychologically plausible principles for modelling language development.

One of our major tasks in modelling the acquisition of semantics is to see how structured semantic representations could emerge from the network's self-organisation of lexical features of verbs. This task is designed to answer two challenges: (1) how can neural networks capture the formation of covert semantic categories, categories that have been traditionally thought elusive, subtle, or even intangible (e.g. the Whorfian cryptotype)? and (2) how can neural networks capture the emergence of lexical aspect categories, categories that have been believed to be innate or otherwise universal (e.g. Bickerton's bioprogram or Slobin's semantic space)? Our SOFM network, through the self-organisation of multiple semantic features, develops concentrated patterns of activity that correspond to cryptotypes (in the case of prefix acquisition) and verb categories (in the case of suffix acquisition). Note that the actual identity of each of the semantic features is unknown to the modeller, because the features encode lexical co-occurrence constraints in a high-dimensional space (see the discussion of input representations on pp. 129–130; see also p. 124). This contrasts with traditional hand-crafted or *ad hoc* features. Our simulation results suggest that connectionist mechanisms as implemented in our model can indeed capture critical aspects of semantic organisation and category formation in language acquisition, without making *a priori* assumptions about the intangibility or the innate nature of lexical semantics.

One might argue that our input representations already contain a rich set of semantic information (as in the HAL semantic vectors), and so it is misleading to claim that the network is acquiring semantic categories. This argument should be considered in at least two perspectives. First, our network takes in only individual verbs as input, in no structured order, but with each verb having information of lexical co-occurrences. What the network needs to do is to re-represent the lexical co-occurrence information in such a way that the resulting two-dimensional map can maximally preserve the similarity of verbs in the original high-dimensional space. This is a process in which the network attempts to discover the underlying structure or organisation for all the verbs in question. None of this structural or organisational information is labelled in the individual verbs, but derived only by the statistical procedure of the network.

A second, and perhaps more important perspective, is to consider that the learner has two simultaneous processes, one that organises the lexical co-occurrence information into meaningful structures (as in SOFMs), and another that extracts the co-occurrence information from the corpus (part of the language experience). In fact, we have recently built a model that does just that. In Farkas and Li (2001, 2002), we developed a connectionist model that acquires lexical knowledge from the learning of distributional characteristics of words. The model consists of two subnetworks: one learns word transitional probabilities in sentence contexts, and the other—a SOFM—reads these probabilities as distributed representations and self-organises them. We applied the model to a CHILDES parental data set and found that the model is able to acquire grammatical and semantic categories through learning in the corpus. In addition, the network demonstrates ability to develop rather accurate representations even with sparse training data, contrary to what is commonly expected of large-scale statistical learning models that typically compute tens or hundreds of millions of lexical items in the corpus.

Thus, the argument here is that the linguistic input to the learner contains very rich distributional information, and our network as well as the child can explore and extract from the input the necessary semantic categories (see Rohde & Plaut, 1999, and Seidenberg & MacDonald, 1999, for similar arguments in the case of grammar induction). Instead of assuming that certain semantic categories are available ahead of time for the child, we need only to make a few simple assumptions about what the child can do: (1) the child has the ability to track continuous speech with some limitation on working memory; and (2) the child is sensitive to lexical co-occurrence probabilities during language comprehension. Such statistical abilities seem to be readily available to the child at a very early age, as studies of statistical learning in infants have revealed (Saffran, Aslin, & Newport, 1996; Saffran et al., 1997). Note that such assumptions differ from the empiricist *tabula rasa* approach to the learning problem, as illustrated clearly by Elman, Bates, Johnson, Karmiloff-Smith, Parisi, and Plunkett (1996) on connectionist learning. Along the arguments of Elman et al., I suggest that specific linguistic categories (e.g. semantic categories discussed here) are not innate; rather, the learner has available a set of statistical mechanisms (which can be operationalised as connectionist principles), and these mechanisms, when applied onto the linguistic input, can yield relevant semantic or grammatical categories. Our modelling results show exactly how such categories can emerge naturally from connectionist learning of the statistical properties of lexical and morphological uses.

Finally, our modelling endeavour has also attempted to make a connection between structured semantic representations and the acquisition of morphology. Our SOFM network, when coupled with Hebbian learning, produces developmental patterns of both overgeneralisation (in prefix acquisition) and undergeneralisation (in suffix acquisition) that mirror empirical data in child language. Our analyses of the simulations indicate that these generalisation errors naturally result from the structure of the network's semantic representations (a result of self-organisation) and from the focused associative pathways in the mappings between semantic features, lexical forms, and morphological markers (a result of Hebbian learning). Further analyses also show that our network is able to recover from the generalisation errors as learning progresses, achieved by the readjustment of the associative weights between forms and meanings via Hebbian learning. These analyses suggest that the learning of a morphological affix is not simply the learning of a rule (leaving alone the fact that it is unclear what the rule is, as per Whorf on the use of *un-*), but the accumulation of associative strengths that hold between a particular affix and a complex set of semantic features distributed across verbs. This learning process can be best described as a statistical process in which the learner implicitly tallies and registers the frequency of co-occurrences (strengthening what goes with what) and identifies the co-occurrence constraints (inhibiting what does not go with what) among the semantic features, lexical forms, and morphological markers.

To conclude, our self-organising neural network model of language acquisition provides significant insights into the mechanisms and processes of language acquisition. It may also serve to stimulate further empirical research, because the model often generates detailed patterns that are not yet available from empirical studies. Future research in our laboratory involves the development of models that tie even more closely to realistic language learning, for example, in the dynamic growth of networks' processing resources, automatic extraction of contextual constraints, and the dynamic representation of lexical–semantic information.

ACKNOWLEDGEMENTS

This research was supported by a grant from the National Science Foundation (#BCS-9975249). I am very grateful to Brian MacWhinney and Risto Miikkulainen for their comments and discussions at various stages of the project. I would also like to thank Curt Burgess and Kevin Lund for making the HAL vectors available. The simulations were run on a SUN Sparc, using the DISLEX code configured by Miikkulainen (1999). Some of the preliminary results are reported in Li (1999), Li (2000), and Li and Shirai (2000).

REFERENCES

Aksu-Koç, A. (1998). The role of input vs. universal predispositions in the emergence of tense-aspect morphology: Evidence from Turkish. *First Language, 18*, 255–280.

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

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences, 7*, 173–188.
Bloom, L., Lifter, K., & Hafitz, J. (1980). Semantics of verbs and the development of verb inflection in child language. *Language, 56*, 386–412.
Bowerman, M. (1978). Systematising semantic knowledge: Changes over time in the child's organisation of word meaning. *Child Development, 49*, 977–987.
Bowerman, M. (1982). Reorganisational processes in lexical and syntactic development. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 319–346). Cambridge: Cambridge University Press.
Bowerman, M. (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* (pp. 445–470). New York: Plenum.
Bowerman, M. (1988). The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). New York: Blackwell.
Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*, 1–34.
Burgess, C., & Lund, K. (1999). The dynamics of meaning in memory. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 17–56). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Carey, S. (1978). The child as word learner. In M. Halle, G. Miller, & J. Bresnan (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.
Chomsky, N. (1957). *Syntactic structures*. The Hague, North Holland: Mouton.
Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–34). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Clark, E. V. (1996). Early verbs, event-types, and inflections. In C. E. Johnson & J. H. V. Gilbert (Eds.), *Children's language, 9*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Clark, E. V., Carpenter, K., & Deutsch, W. (1995). Reference states and reversals: Undoing actions with verbs. *Journal of Child Language, 22*, 633–662.
Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems.* Cambridge: Cambridge University Press.
Cottrell, G., & Plunkett, K. (1994). Acquiring the mapping from meaning to sounds. *Connection Science, 6*, 379–412.
Elman, J., Bates, A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
Fahlman, S., & Lebiere, C. (1990) The cascade-correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in neural information processing systems 2* (pp. 524–532). Los Altos, CA: Morgan Kaufmann Publishers.
Farkas, I., & Li, P. (2001). A self-organising neural network model of the acquisition of word meaning. In E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (Eds.), *Proceedings of the Fourth International Conference on Cognitive Modeling* (pp. 67–72). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Farkas, I., & Li, P. (2002). Modeling the development of lexicon with a growing self-organising map. In H. J. Caulfield et al. (Eds.), *Proceedings of the Sixth Joint Conference on Information Science* (pp. 553–556). Association for Intelligent Machinery, Inc.
Francis, W., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston, MA: Houghton Mifflin.
Hebb, D. (1949). *The organisation of behavior: A neuropsychological theory*. New York: Wiley.

Hertz, J., Krogh, A., & Palmer, R. (1991). *Introduction to the theory of neural computation.* Redwood City, CA: Addison-Wesley.

Joanisse, M., & Seidenberg, M. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, 96*, 7592–7597.

Klahr, D., & MacWhinney, B. (2000). Information processing. In D. Kuhn, R. S. Siegler, & W. Damon (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 631–678). New York: Wiley.

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

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

Kohonen, T. (1995). *Self-organising 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., & Bever, T. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. *Cognition, 28*, 195–247.

Li, P. (1993). Cryptotypes, form-meaning mappings, and overgeneralisations. In E. V. Clark (Ed.), *The Proceedings of the 24th Child Language Research Forum* (pp. 162–178). Center for the Study of Language and Information, Stanford University, CA.

Li, P. (1999). Generalisation, representation, and recovery in a self-organising feature-map model of language acquisition. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 308–313). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Li, P. (2000). The acquisition of lexical and grammatical aspect in a self-organising feature-map model. In L. Gleitman & Aravind K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 304–309). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Li, P., & Bowerman, M. (1998). The acquisition of lexical and grammatical aspect in Chinese. *First Language, 18*, 311–350.

Li, P., Burgess, C., & Lund, K. (2000). The acquisition of word meaning through global lexical co-occurrences. In E. V. Clark (Ed.), *Proceedings of the 30th Child Language Research Forum* (pp. 167–178). Center for the Study of Language and Information, Stanford University, CA.

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

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments and Computers, 34*, 408–415.

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Berlin: Mouton de Gruyter.

Ling, C., & Marinov, M. (1993). Answering the connectionist challenge. *Cognition, 49*, 267–290.

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

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk* (3rd ed). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

MacWhinney, B. (2001). Lexicalist connectionism. In P. Broeder & J. M. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (pp. 9–32). Oxford: Oxford University Press.

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

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

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T., & Xu, F. (1992). Overregularisation in language acquisition. *Monographs of the Society for Research in Child Development, 57* (Serial No. 228).

McClelland, J., & Rumelhart, D. (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.

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

Miikkulainen, R. (1999). The DISLEX simulator (new version). Available online at http://www.cs.utexas.edu/users/nn/pages/software/dislex.2.1.tar.Z/

Mourelatos, A. (1981). Events, processes, and states. In P. Tedeschi & A. Zaenen (Eds.), *Syntax and semantics: Vol. 14. Tense and aspect*. New York: Academic Press.

Parsons, T. (1990). *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT Press.

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.

Pinker, S. (1999). Out of the minds of babes. *Science, 283*, 40–41.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73–193.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, 23, 463–490.

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.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21–69.

Ritter, H., & Kohonen, T. (1989). Self-organising semantic maps. *Biological Cybernetics, 61*, 241–254.

Rohde, D., & Plaut, D. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition, 72*, 67–109.

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, pp. 216–271). Cambridge, MA: MIT Press.

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

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

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

Seidenberg, M., & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569–588.

Shirai, Y. (1998). The emergence of tense-aspect morphology in Japanese: Universal predisposition? *First Language, 18*, 281–309.

Shirai, Y., & Andersen, R. (1995). The acquisition of tense-aspect morphology: A prototype account. *Language, 71*, 743–762.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2, pp. 1157–1249). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Smith, C. (1983). A theory of aspectual choice. *Language, 59*, 479–501.

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

Stephany, U. (1981). Verbal grammar in modern Greek early child language. In P. S. Dale & D. Ingram (Eds.), *Child language: An international perspective*. Baltimore, MD: University Park Press.

Whorf, B. (1956). A linguistic consideration of thinking in primitive communities. In J. B. Carroll (Ed.), *Language, thought, and reality* (pp. 65–86). Cambridge, MA: MIT Press.