9

Meaning in Acquisition
Semantic Structure, Lexical Organization, and Crosslinguistic Variation

PING LI
The Pennsylvania State University

INTRODUCTION

In this chapter, I examine three problems: the acquisition of tense and aspect, the acquisition of cryptotypes, and the development of lexical structure. The central issue in all three problems is how the child discovers word meanings in a lexical system, and what mechanisms are at work in the process of this discovery. In each case, the problem domain involves the learning of semantic structures that are important for the use of grammatical morphology or lexical categories. In examining these three cases, I argue that structured semantic representations of the lexicon can emerge as a natural outcome of the meaning-form and the meaning-meaning mappings in language acquisition.

Two major perspectives inform the examination of children's development of these structures (and have been key assumptions underlying Bowerman's own work). First, one cannot look at the acquisition of a given set of forms in isolation. In the study of the acquisition of morphology, for example, it is important to examine not only the morphological devices per se, but also the words (particularly verbs) with which these devices are used (Bowerman, 1982, 1983). For example, the acquisition of the prefix *un-* needs to be considered closely with the verbs that can take *un-*. In Whorf's (1956) view, *un-* marks a cryptotype, a covert semantic category that can only be implicitly defined by the devices it goes with. Thus, the understanding of *un-* in child language cannot be complete without a consideration of the cryptotype meanings that are characteristic of the *un*-able verbs. By the same token, the acquisition of tense-aspect suffixes such as *-ing* and *-ed* needs to be considered together with the types of verbs (atelic, telic, result, etc.) to which the suffixes are attached. In the empirical literature there is ample evidence that children pay attention to the inherent meanings of the verbs when they learn to use tense-aspect suffixes (Bloom, Lifter, & Hafitz, 1980; Brown, 1973; Slobin, 1985). In Li and Bowerman (1998) and Li and Shirai (2000), we examined specifically the interaction between grammatical aspect (expressed by morphological markers) and lexical aspect (expressed in inherent meanings of verbs) in both first and second language acquisition, in a systematic attempt to connect morphology and lexical semantics.

A second important perspective championed by Bowerman is that in formulating and testing theories of language acquisition, it is important to examine acquisition data from not only one language (usually English by default), but also from other languages. Crosslinguistic evidence is crucial in many cases because of the language-specific properties associated with individual languages. Often, a given hypothesis appears to be perfect for one language, but it may turn out to be inaccurate or incomplete upon a close look at data from other languages. For example, English uses the prepositions *on* and *in* to express spatial locations of objects in attachment and fitting situations. In Dutch, the orientation of attachment matters, such that vertical versus horizontal attachments involve the use of different prepositions; in Korean, the degree of fit matters, such that tight-fit versus loose-support situations involve the use of different locative verbs. Thus, how English-speaking children learn *on* and *in* will not be sufficiently informative as to how Dutch and Korean children learn the corresponding devices in their languages. A crosslinguistic perspective can be crucial in illuminating the extent to which language-specific input may play a role in shaping children's early grammar (e.g., Bowerman, 1985, 1989, 1996).

This chapter summarizes our research that aims at connecting morphology and lexical semantics and at connecting acquisition theories and crosslinguistic data. The discussion is organized by the three topics that I mentioned, namely, the acquisition of tense and aspect, the acquisition of cryptotypes, and the development of lexical structure. Through careful analysis of crosslinguistic data, as well as connectionist modeling of acquisition, I show that the semantic structure underlying each of these domains emerges from the associations between forms and meanings present in the input.

THE ACQUISITION OF TENSE AND ASPECT

Empirical Observations

The expression of time is one of the central conceptual domains of language, and the acquisition of the ability to talk about time through the use of tense-aspect markers is one of the earliest tasks in language acquisition. In the last 30 years researchers have devoted much attention to the acquisition of tense and aspect in various languages (for general reviews, see Li & Shirai, 2000; Shirai, Slobin & Weist, 1998; Slobin, 1985; Weist, 1986). An early observation of children's acquisition of tense-aspect markers came from Roger Brown (1973). Brown documented two interesting patterns: First, the earliest grammatical device in children's speech, the progressive aspect marker *-ing*, appears virtually always to be used correctly. In particular, children never use *-ing* incorrectly with state verbs; for example, they do not produce overgeneralizations like *knowing* or *wanting*. Second, English-speaking children first use past-tense forms with only a small, semantically coherent set of verbs, including *dropped*, *slipped*, *crashed*, and *broke*, verbs that indicate events that happen instantaneously but lead to clear end results. Some years after Brown's observations, Bloom, Lifter, and Hafitz (1980) provided further evidence that confirmed Brown's analyses (especially the second observation). They found that the inflections used by young English-speaking children correlated with the semantic types of verbs: *-ing* occurred almost exclusively with verbs such as *play*, *ride*, and *write* (durative, nonresultative), whereas past-tense forms occurred predominantly with verbs such as *find*, *fall*, and *break* (punctual, resultative). Together, Brown and Bloom et al.'s data suggested a picture of early "undergeneralization" in the acquisition of inflectional morphology: Rather than using tense-aspect markers with all types of verbs, as adults do, children use them more restrictively. Obviously, this more restrictive use of grammatical morphemes is associated with the inherent semantic differences of the verbs to which the progressive and past-tense forms are attached.

The strong associations between tense-aspect morphemes and verb semantics observed by Brown and Bloom et al. have generated a considerable amount of research. Many studies have examined the acquisition of tense and aspect in other languages, including Chinese (Erbaugh, 1982; Li, 1990; Li & Bowerman, 1998), French (Bronckart & Sinclair, 1973), Italian (Antinucci & Miller, 1976), Polish (Weist, Wysocka, Witkowska-Stadnik, Buczowska, & Konieczna, 1984), Turkish (Aksu, 1978; Aksu-Koç & Slobin, 1985), and other languages (see Li & Shirai, 2000 for a review of relevant crosslinguistic data). In general, data from these studies are robust and consistent with Brown and Bloom et al.'s observations, indicating that the use of imperfective or progressive aspect morphology in children's speech is first associated with atelic, activity verbs, whereas that of perfective aspect/past tense morphemes is associated with telic, resultative verbs.[1]

Subsequent discussions on these empirical data have stimulated intense debates on how children acquire lexical semantics and grammatical morphology and have motivated theoretical accounts of children's semantic and morphological development in general (see Li & Shirai, 2000, for an overview). Here I discuss two major contrasting proposals and then attempt to explain the data from a crosslinguistic developmental perspective.

---

[1] Not all acquisition researchers agree that these association patterns exist crosslinguistically. For example, Weist and his colleagues argued against the proposal of Bloom et al. (1980). They showed that Polish children are able to understand and produce the basic contrast between perfective and imperfective aspect as early as 2;6 (Weist et al., 1984). See Li and Shirai (2000, pp. 40–47) for a discussion of the relevant debate.

Contrasting Perspectives

The major divide in perspectives on the acquisition of tense and aspect falls between formalist-nativist and functionalist-cognitivist approaches. The formalist perspective is strongly associated with a nativist view in the tense-aspect acquisition literature. In its strongest form, Bickerton (1981, 1984) proposed the language bioprogram hypothesis, according to which specific semantic categories and concepts are biologically preprogrammed for the human language learner, and these categories or concepts will naturally unfold in the process of language acquisition. Because the categories and concepts are hard-wired ahead of time, the child simply needs to discover how they are instantiated in specific forms in the language to be learned. Two important innate distinctions in the domain of tense and aspect are between *state* and *process* and between *punctual* and *nonpunctual* categories. Given the innate nature of these distinctions, according to Bickerton, early on in language development *states* will be marked differently from *processes*, and *punctual* situations will be marked differently from *nonpunctual* situations, probably by the use of different tense-aspect markers.[2]

Bickerton first supported his hypothesis with evidence from Creole grammars, arguing that in the absence of relevant input (pidgins, the predecessors of Creoles, do not have tense-aspect markers), first-generation Creole speakers invent tense-aspect systems to mark the bioprogrammed distinctions. Drawing in addition on child language data, he argued that children first use the tense-aspect markers of their language to mark the distinctions between *state* and *process* and between *punctual* and *nonpunctual*. For example, Bickerton used Brown's (1973) observation in support of an innate specification of the distinction between *process* and *state*: Young English-speaking children never overgeneralize the progressive marker *-ing* to state verbs because they are sensitive to the bioprogrammed *state–process* distinction. Similarly, Bickerton argued for the existence of the *punctual–nonpunctual* distinction in the bioprogram, on the basis of interpreting the observations by Antinucci and Miller (1976), Bronckart and Sinclair (1973), and Aksu-Koç and Slobin (1985; originally Slobin & Aksu, 1980) that young children use the past or perfective morphemes to mark punctual events. Thus, although Bickerton's language bioprogram hypothesis originated from studies of Creole languages, many of its arguments rest on interpretations of data in children's acquisition of tense and aspect. Bickerton considered early child languages and Creoles to be ideal cases for observing the bioprogram, because the innate distinctions of the bioprogram are realized in these cases without being

[2] A somewhat different, but related view is advocated by Slobin (1985). Slobin proposed the basic child grammar, which contains a prestructured "semantic space" with universal semantic notions or categories. These semantic categories can act strongly to attract the morphological mapping from the input language. *Result* and *Process* are two such semantic categories that have to do with children's acquisition of tense-aspect morphology. However, because the issue of innateness is less fundamental to the basic child grammar than it is to the language bioprogram hypothesis, and because Slobin's (1997) recent reformulation is consistent with a cognitive–functional view, I do not consider the basic child grammar as a formalist–nativist theory in this debate.

contaminated by external factors subject to cultural evolution (e.g., individual linguistic variations that are idiosyncratic).

The language bioprogram hypothesis attempted to explain language acquisition by appealing to innately determined semantic distinctions. A contrasting perspective is the functionalist-cognitivist approach to tense-aspect acquisition, which has had many variants in the literature. One early explanation for the restricted tense-aspect uses in child language drew on the child's purportedly insufficient cognitive ability (Antinucci & Miller, 1976; Bronckart & Sinclair, 1973). Other investigators turned to the "input hypothesis" to explain the early associations between grammatical morphology and lexical aspect (e.g., Stephany, 1981 for Modern Greek; Li, 1990 for child Mandarin; Shirai, 1991 for English). Still others used the "aspect hypothesis" or the "prototype hypothesis" to explain similar patterns in L2 learning (Shirai, 1991; Shirai & Andersen, 1995). More recently, Li and Shirai (2000) presented an integrated view of the functional approach, drawing ideas from both the prototype hypothesis and connectionist networks. They argued that in both L1 and L2, the learner's early associations between lexical meanings of verbs and grammatical morphemes do not indicate innate specifications of semantic categories. Rather, these associations reflect the learner's sensitivity to (and recognition of) the statistical properties of the linguistic input, and in turn, statistical properties in the input may reflect inherent constraints on linguistic communication and event characteristics. The semantic categories believed to be innate can emerge naturally from the learning of the lexical characteristics of verbs in context. Finally, Li and Shirai argued that the associations between grammatical morphology and lexical aspect are probabilistic and not absolute, counter to what nativist proposals would assume. Depending on the structure of the target language, in some cases, learners retreat from the probabilistic associations and develop more flexible patterns of use; in other cases, they hold on to these associations even as they acquire adult patterns of language use.

Resolving the Conflict Crosslinguistically

It has become increasingly clear over the last few years that a strong nativist proposal like Bickerton's cannot account for the large body of crosslinguistic data. Instead, a functional, input-based, probabilistic learning mechanism seems to be most compatible with the way children approach the problem of tense-aspect acquisition. Even Bickerton (1999) himself has significantly weakened the strong predictions of the original language bioprogram hypothesis, assigning an increased role to input language patterns in acquisition. But how does crosslinguistic evidence help in resolving the conflict?

Li and Bowerman (1998) presented crosslinguistic data on the acquisition of aspect in Chinese. In three experiments we showed that young children learning Mandarin Chinese displayed the same type of strong associations between grammatical morphology and lexical aspect as in English and other languages. In particular, children comprehended the progressive marker *zai* better with atelic, activity verbs than with telic verbs, and, conversely, the perfective marker *-le* better with telic verbs than with atelic verbs. Children also produced the imperfective aspect markers *zai* and *-ne* mostly with atelic verbs and rarely with telic verbs, whereas they produced the perfective marker *-le* more frequently with telic verbs than with atelic verbs.

While the Chinese data would seem to be no surprise as compared with data from English and other languages, there were at least two important features that make them special.

1. Unlike English and other Indo-European languages, Mandarin Chinese has a special set of state verbs, the posture verbs, that cannot take the progressive marker *zai* (e.g., *zhan* 'stand,' *zuo* 'sit,' and *tang* 'lie'). Recall that Brown (1973) observed that children do not overgeneralize *-ing* to state verbs in English (posture verbs are not state verbs in English as they are in Chinese), and Bickerton interpreted this as indicative of children's innate sensitivity to the state-process distinction. In child Mandarin, however, children do overgeneralize the progressive marker *zai* to posture state verbs as revealed in our production experiment, and such errors are also observed in children's spontaneous speech. Two lessons can be drawn from this finding.
   a. First, the *state-process* distinction is not a universal semantic primitive as dictated by the language bioprogram hypothesis, and different languages may define the distinction differently. For example, posture verbs are state verbs in Chinese but not in English (in Chinese **ta zai zuo* is ungrammatical while in English *he is sitting there* is perfectly acceptable). The English-speaking child can use *-ing* with *sit* and *stand* to indicate the dynamic aspect of the action, treating sitting and standing as events, while the Chinese-speaking child uses *zai* with *zuo* and *zhan* in Chinese, treating them as states. Comrie (1976) discussed this phenomenon in connection with perception verbs (e.g., *see, hear*): English treats perception verbs as stative and these verbs consequently do not accept progressive marking, while Portuguese treats them as dynamic so they can naturally accept progressive marking.
   b. Even if the state-process distinction is neatly defined by language, children may not observe the distinction in their use of tense-aspect markers. This is consistent with Shirai's (1994) analysis that shows that even in English, children do occasionally generalize *-ing* to state verbs depending on the type of input they receive from maternal speech.
2. A second important feature is that Chinese has a special set of resultative verb compounds (RVCs), such as *ti-dao* 'kick-down,' *qie-kai* 'cut-open,' or *reng-diao* 'throw away,' which are telic verbs that encode a clear end result (Klein, Li, & Hendriks, 2000; Smith, 1997; Tai, 1984). These verbs accept only perfective marking, not progressive marking in Chinese, unlike resultative verbs in English, which can easily take the progressive *-ing* (and so the resultative *-ing* combinations are found in child English). In our study children rarely used the progressive aspect with the RVC verbs, showing that they respected the incompatibility between aspectual

imperfectivity and lexical resultativity from early on. Interestingly, these RVC verbs correspond to a set of verbs that "name events of such brief duration that the event is almost certain to have ended before one can speak" (Brown, 1973, p. 334)—i.e., verbs with which young English-speaking children only use the past-tense forms. Bickerton would have taken the exclusive perfective marking of RVCs as an indication of the *punctual–nonpunctual* distinction, just as he did with the English data. But the crucial evidence from semelfactive verbs (Smith, 1997; e.g., *tiao* 'jump,' *qiao* 'knock,' *ti* 'kick,' which have the punctuality but not the resultativity feature) indicated that it was not punctuality but resultativity that children pay attention to (Li & Bowerman, 1998): With these semelfactive verbs, Chinese-speaking children do use the progressive marker *zai*.

The language-specific properties with RVC verbs in Chinese not only influence the specific patterns of children's acquisition of aspect markers with these verbs, but also speak to several general crosslinguistic differences.

a. First, because the adult language has a constraint on the combination of RVCs with the progressive marker, the picture of lexical aspect and grammatical morphology appears much more absolute than probabilistic, as compared with that in other languages.
b. Second, as compared with children learning other languages, Chinese-speaking children display no developmental transition from prototypical associations (e.g., result verbs with perfective aspect) to nonprototypical associations (e.g., result verbs with imperfective aspect). The prototype hypothesis predicts such transitions as the child's linguistic experience enriches (Shirai & Andersen, 1995). Now, given that the prototypical association between RVCs and perfective marking is preserved in the adult language, there is no reason for children to retreat from this association and move to nonprototypical ones, as they would in other languages where the associations are more flexible.
c. Finally, the combinatorial constraints in adult Chinese might reflect a general constraint on linguistic communication and event characteristics. The prototypical association between RVCs and perfective aspect reflects one of the most natural combinations between verb semantics and grammatical morphology. In Comrie's (1976) view, certain aspect morphemes combine most naturally with certain verb types but not others (the "naturalness of combination" principle). As Brown (1973) had also pointed out, events denoted by verbs like *drop*, *fall*, and *crash* (events expressed by RVCs in Chinese) occur instantaneously; any comment on them will have occurred after their ending. Thus, it is only natural to describe these events with perfective aspect (i.e., to combine *-le* but not *zai* with RVCs in Chinese).

The above crosslinguistic analyses, along with results from studies of other languages indicate that children are highly sensitive to language-specific properties of the input, and are capable of extracting systematic patterns from the input (see Li & Bowerman, 1998; Li & Shirai, 2000 for detailed discussions). Given this, we do not need to presuppose, as nativists do, that certain semantic categories are innately specified and brought to bear on the language acquisition task. Rather, semantic categories can emerge from the learning of the statistical regularities in the input language. But what capacity allows the child to carry out the pattern extraction in learning? Put simply, if input is important to language acquisition, in which way does it play a causal role? I referred earlier to a functional, input-based, and probabilistic learning mechanism that could be responsible for the acquisition task. In the next section, I discuss how such a learning mechanism could work in terms of the operations of connectionist networks.

THE ACQUISITION OF A CRYPTOTYPE

Whorf's Cryptotypes

In one of the classic papers of early cognitive linguistics, Whorf (1956) presented the following puzzle. In English, the reversative prefix *un-* can be used productively with many verbs to indicate the reversal of an action, as in *uncoil, uncover, undress, unfasten, unfold, unlock,* or *untie* (the meaning of reversal can also be expressed by other prefixes such as *dis-* or *de-* in English). However, many seemingly parallel forms are not allowed, such as **unbury, *unfill, *ungrip, *unhang, *unpress, *unspill,* or **unsqueeze.* Why is *un-* prefixation allowed with some verbs but not others?

Whorf's puzzle was deeper than this simple discrepancy. He noted that *un-* is a productive device in English morphology, and that despite the difficulties linguists have in characterizing its use, native speakers do have an intuitive feel for which verbs can be prefixed with *un-* and which cannot. He presented the following thought experiment: If a new verb *flimmick* is coined to mean "to tie a tin can to something," then native speakers are willing to accept the sentence "He unflimmicked the dog" as expressing the reversal of the "flimmicking" action; if *flimmick* means "to take apart," then they will not accept "He unflimmicked the puzzle" as describing the act of putting a puzzle back together. The constrained productivity of *un-* prompted Whorf to conjecture that there is some underlying or covert semantic category, a *cryptotype*, that governs the productive use of *un-*. According to Whorf, cryptotypes only make their presence known by the restrictions they place on the possible combinations of overt forms. When the overt prefix *un-* is combined with the overt verb *tie*, there is a covert cryptotype that licenses the combination *untie*. This same cryptotype also blocks a combination such as **unmove*.

To Whorf, the deep puzzle was that while the use of the prefix *un-* is productive, the cryptotype that governs its productivity is unclear: "We have no single word in the language which can give us a proper clue to this meaning or into which we can compress this meaning; hence the meaning is subtle, intangible, as is typical of cryptotypic meanings" (Whorf, 1956, p. 71). Here we have a case for which language use is conditioned in a principled way, but the principles themselves are not clearly subject to linguistic analysis. At some point Whorf did propose that there was "a covering, enclosing, and surface-attaching meaning" that could be the basis of the cryptotype for *un*-. But this definition was still rather elusive, as it was not clear whether we should view this as a single unit, three separate meanings, or a cluster of related meanings. Subsequent analyses also suggested the existence of other important aspects in the use of *un*- (Clark, Carpenter, & Just, 1995; Marchand, 1969)—for example, that *un*- takes change of state verbs, and that these verbs involve a direct object (so intransitive verbs such as **unswim*, **unplay*, and **unsnore* are ill-formed).

Cryptotypes in Child Language

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

According to Bowerman (1982, 1983, 1988), children's acquisition of *un*- tends to follow a U-shaped pattern, a pattern found in other areas of morphological acquisition as well, such as the acquisition of the English past tense. Children initially produce *un*- verbs in appropriate contexts, treating *un*- and its base verb as an unanalyzed whole. This initial stage of rote control is analogous to the child's saying *went* without realizing that it is the past-tense form of *go*.

Productivity of *un*- comes at the next stage, when children realize that *un*- is independent of the verb in indicating the reversal of an action. This stage in the acquisition of *un*- begins at around age 3. At this stage, children start to produce overgeneralizations in spontaneous speech such as **unarrange,* **unbreak,* **unblow,* **unbury,* **unget,* **unhang,* **unhate,* **unopen,* **unpress,* **unspill,* **unsqueeze,* or **untake* (Bowerman, 1982, 1983). Such overgeneralizations have also been documented by Clark et al. (1995) in both experimental and naturalistic data with children from ages 3 to 5, and were found in the CHILDES database (Li & MacWhinney, 1996). During this period, children also make certain "overmarking" errors. For example, the child might say "unopen" but really only mean to say *open*, or "unloosen" to mean *loosen*. In such cases, the base forms *open* and *loosen* have a reversative meaning that triggers the attachment of the prefix, even when the action of the base meaning is not actually being reversed. These errors are analogous to redundant past-tense marking as in **camed* and redundant plural marking as in **feets*. Finally, at a third stage, overgeneralization and overmarking errors both disappear.

A critical factor that leads to children's overgeneralization of *un*- at the second stage of this U-shaped learning, according to Bowerman (1982), is that children have somehow discovered the inherent meaning common to the verbs that take *un*-. In other words, they have developed what Whorf called the native "intuitive feel" for English verbs with respect to whether they are *un*-able: They have acquired the representation for the *un*- cryptotype. Examining speech errors from a longitudinal dataset, Bowerman further suggested two possible roles for a cryptotype to influence the learning of *un*-: (1) "Generalization via cryptotype": The cryptotype triggers morphological productivity and leads to overgeneralizations. This occurs because, once children have identified the cryptotype, they will overgeneralize *un*- to all verbs that fit the cryptotype, irrespective of whether the adult language actually allows *un*- in these cases (e.g., *squeeze* fits in the cryptotype just as *clench* does, so say "unsqueeze"). (2) "Recovery via cryptotype": The cryptotype helps the child to overcome overgeneralizations made at an earlier stage, if these overgeneralizations involve verbs that fall outside the cryptotype (e.g., *hate* does not fit in the cryptotypic meaning, so stop saying "unhate").

While Bowerman correctly identified the important role of cryptotypes in child language acquisition, one issue remains unclear. According to Whorf, cryptotypes are covert semantic categories that are elusive, subtle, and intangible, and linguists have a hard time to pin down their precise meanings. How then could the child extract the cryptotype and use it as a basis for morphological generalization or recovery, if the *un*- cryptotype is intangible even to linguists like Whorf?

The answers to this question have significant implications for the issue concerning whether language acquisition is a rule-based process or a statistical learning process (Pinker, 1991; Pinker & Prince, 1988; Rumelhart & McClelland, 1986; Seidenberg, 1997). Using the acquisition of the English past tense as an example, researchers have debated whether the acquisition process should be characterized by dual mechanisms (an internalized linguistic rule for regulars and an associative learning process for irregulars) or by a single mechanism (connectionist learning with distributed knowledge representation and adaptive connection weights). Cryptotypes provide another test case for this debate. If the learning of *un*- and its governing cryptotype is a process of rule extraction (category identification), then the overgeneralization errors with *un*- are rule-governed, in the same way as are the overgeneralization errors with -*ed*. However, if the learning of *un*- and the cryptotype is a connectionist statistical process, then the overgeneralization errors are due to the system's computation of relevant semantic features, lexical forms, and prefixation patterns in the form-meaning mapping process.

A Connectionist Model of the un- Cryptotype

Connectionist networks are dynamic learning systems that explore the regularities in the input-output mapping processes through the adjustment of connection weights and the activation of processing units. To answer the above questions, Li (1993) and Li and MacWhinney (1996) built a connectionist model to learn the reversative cryptotype associated with the use of *un*-. Our model was a standard feedforward network consisting of three layers of processing units (input, output, and hidden units). The network was trained with the backpropagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). In our simulations, we used as input to our network 49 verbs that can take *un*-, 19 verbs that can take the competing prefix *dis*-, and 92 randomly selected verbs that can take neither prefix. Each verb was represented by a semantic pattern (a vector) that consisted of 20 semantic features that were selected in an attempt to capture basic linguistic and functional properties inherent in the semantic range of these verbs (see Li, 1993 for details). The task of the network was to take the semantic vectors of verbs as input and map them onto different prefixation patterns in the output: *un-*, *dis-*, or zero.

To analyze how our network developed internal representations, we used the hierarchical cluster analysis (Elman, 1990) to probe into the activation of the hidden units at various points in time during the network's learning. The network received input verbs one by one and determined if each verb should be mapped to *un-*, *dis-*, or no prefix. Each time the network received some feedback as to whether the mapping was correct. After learning, the averaged representations of the verbs at the hidden-unit level were clustered. Figure 9.1 presents a snapshot of the network's hidden-unit representations when the network had learned 50 verbs cumulatively. This graph shows two general clusters: one for the *un-* verbs, and the other for the zero verbs, verbs that cannot be prefixed with *un-* or *dis-*. Our interpretation of these clusters is that the network has acquired a distinct representation for the *un-* verbs by identifying the relevant semantic features shared by these verbs, and this representation corresponds most closely to Whorf's *un-* cryptotype. For example, most of the verbs in the *un-* cluster share the meaning of binding or locking: *bind, chain, fasten, hitch, hook, latch,* etc. Not all meanings relevant to the cryptotype are identified at this early stage in Figure 9.1. For example, the verbs *ravel* and *coil* were correctly categorized into the *un-* cluster, but the verb *roll* was incorrectly classified into the zero-verb cluster.

Note that our network received no discrete label of the semantic category associated with *un-* (the labels in Figure 9.1 were there simply to indicate which prefix the verb is supposed to take), nor was there a single categorical feature that tells which verb should take which prefix (hence Whorf's problem). All that the network received was the semantic feature information distributed across input patterns. Over time, however, the network was able to identify the regularities that hold between distributed semantic patterns and patterns of prefixation, and developed a structured representation that corresponded to Whorf's cryptotype. The cryptotype representations in the network thus emerged as a function of the network's learning of the association between form and meaning, not as a property that was given ad hoc to the network by the modeler.

Our simulation results provide support for Bowerman's (1982) hypothesized role of cryptotypes in inducing overgeneralizations. In Bowerman's data, most of the errors fell within the realm of Whorf's cryptotype (e.g., *squeeze* is similar to *clench*, so *squeeze* can also take *un-*). In Clark et al.'s (1995) data, the child's innovative uses of *un-* also respected the cryptotype from the beginning: They matched the semantic characteristics of the cryptotype even when the conventional meanings of the verb in the adult language did not. For example, **unbuild* was used to describe the action of detaching Lego-blocks, **undisappear* was used to describe the releasing of the child's thumbs from inside his fists.[3] Our simulations also indicate how the network generalized *un-* on the basis of the cryptotype. In Figure 9.1,

[3] Diary notes of my daughter's speech also include similar uses: "<u>unbuild</u> the snowman" was used to refer to the detachment of decorative pieces from the snowman, and "<u>untape</u>" to refer to the removal of scotch tape from a piece of paper (age of child was 6;9).

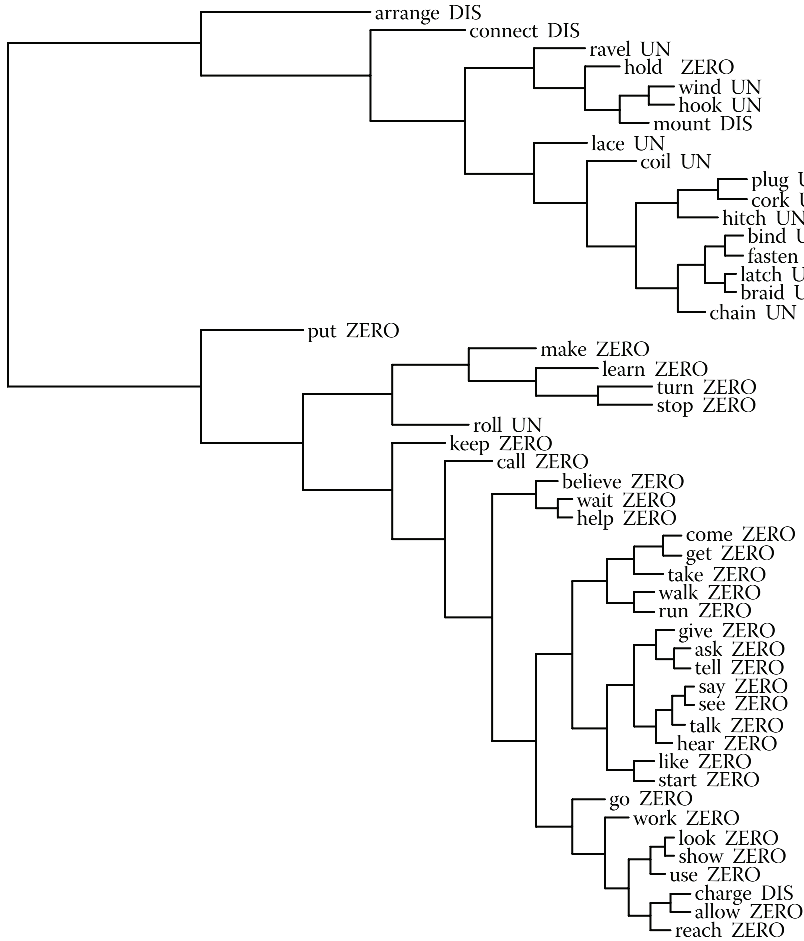

Figure 9.1 A hierarchical cluster analysis of the network's hidden-unit representations at the 50-word stage. The similarity function is determined by the Euclidean distances of the items on the cluster tree. The capitalized marker after each verb indicates the prefixation pattern of the verb, but the network did not receive these labels during training.

the network included both *hold* and *mount* (which should not take *un-*) in the *un-* category. These verbs were included apparently because of their semantic similarity with members of the cryptotype (e.g., *bind, chain, fasten, hitch, hook, latch*). Examining *hold* and *mount* in the network output patterns, we found that *un-* was overgeneralized to these verbs at the end of the 50-word stage. Similar errors produced by the network included **unbury, *uncapture, *unfill, *unfreeze, *ungrip, *unhold, *unloosen, *unmelt, *unpeel,*unplant, *unpress, *unsplit, *unsqueeze, *unstrip, *untack,* and **untighten*. In contrast to these cases, the network produced few errors that constitute flagrant violations of the cryptotype; hence there was no basis for the model to verify Bowerman's (1982) second hypothesis that the cryptotype can serve to eliminate overgeneralizations. Thus, in our simulations, overgeneralizations went hand-in-hand with the network's representation of the cryptotype. This is clearly another case (in addition to tense-aspect acquisition discussed earlier) where the understanding of morphological behavior requires the examination of the semantic structure of the words with which the grammatical morphemes are used.

Implications of the Connectionist Model

A connectionist perspective as described above provides us with a natural way of capturing Whorf's insights into cryptotypes as well as a formal mechanism to mimic their acquisition. In our view, there can be several "mini-cryptotypes," each of which represents some underlying semantic features that work together as interactive "gangs" (McClelland & Rumelhart, 1981). For example, "enclosing" verbs, such as *coil, curl, fold, reel, roll, screw, twist,* and *wind*, all seem to share a meaning of circular movement. Similarly, "attaching" verbs, such as *clasp, fasten, hook, link, plug,* and *tie*, all involve hand movements. Still another cluster of verbs such as *cover, dress, mask, pack, veil,* and *wrap* forms the "covering" mini-cryptotype. These mini-cryptotypes or mini-gangs interact collaboratively to support the formation of the larger cryptotype that licenses the use of *un-* in terms of summed activation in a connectionist network.

Note that the mini-gangs collaborate rather than compete because their members are closely related by the overlap of semantic features. For example, the verb *screw* in *unscrew* may be viewed as having both a meaning of circular movement and a meaning of binding or locking, *zip* in *unzip* may be viewed as sharing both the "binding/locking" meaning and the "covering" meaning, and both *screw* and *zip* involve hand movements. Moreover, a feature may also vary in the strength with which it is represented in different verbs. For example, circular movement is an essential part of the meaning of *screw*, but less so for *wrap* (one can wrap a small ball with a tissue paper without turning around either the object or the wrapping paper). In this way, the meaning of the *un-* cryptotype constitutes a complex semantic network, in which verbs can differ in (1) how many features are relevant to each verb, (2) how strongly each feature is activated in the representation, and (3) how strongly features overlap with each other across category membership (all true with the input to our network). It is these complex relationships that give rise to the meaning of the cryptotype. It is also these complex relationships that gave trouble to traditional rule-based linguistic analyses (hence Whorf's statement regarding the elusive and intangible character of its semantic content).

While these complex structural properties render a symbolic analysis less effective if not impossible, they are accessible to native intuition, according to Whorf. Native intuitions are clearly implicit representations of the complex semantic relationships among verbs and morphological markers, and connectionist networks provide a formal mechanism to capture these intuitions through weighted connections, distributed representations, and statistical learning. In our simulations, the network was able to explore and identify these relationships through the input-output mapping process. The network computed the combinatorial constraints on the co-occurrences of the prefix *un-* and the distributed semantic features of verbs. The result of this process was that new representations that developed at the hidden layer of the network differed from each other in the number of features they shared and in the strengths with which the features were activated, as revealed in Figure 9.1.

Our network's behavior suggests that children, in learning to use the reversative prefix *un-*, may also abstract the semantic regularities from the *un-* verbs through combinatory restrictions that the prefix places on these verbs. In this perspective, children's learning of *un-* is not the learning of a symbolic rule for the use of the prefix with a class of verbs, but rather the accumulation of the connection strength that holds between a particular prefix and a set of weighted semantic features in verbs. The learner groups together those verbs that share the largest number of features and take the same prefixation patterns. Over time, the verbs gradually form coherent classes, with respect to both meaning and prefixation. This learning process can best be described as a statistical procedure in which the child implicitly tallies and registers the frequencies of co-occurrences of distributed semantic features, lexical items, and morphological markers. Not surprisingly, the same process would apply equally well to the acquisition of lexical aspect categories and tense-aspect morphology (see Li & Shirai, 2000 for an analysis; Zhao & Li, in press, for a recent connectionist formalization).

The formation of a cryptotype as in the case of *un-* is not an isolated linguistic phenomenon. It can be observed in many domains in which the problem is primarily semantically motivated. For example, the use of classifiers is one of the hardest problems for second language learners of Chinese, as well as a major challenge for linguistic description (cf. Chao, 1968; Lakoff, 1987). Each noun in Chinese has to be preceded by a classifier that categorizes the object of the noun in terms of its shape, orientation, dimension, texture, countability, and animacy. The appropriate use of classifiers by native speakers is largely automatic, yet it is difficult for linguists to come up with a clear description of rules that govern their use (Erbaugh, 2006). We can probably assume that native speakers have acquired a representation that is cryptotype-like, in which multiple semantic features connected in a network jointly support the use of classifiers.

In short, our connectionist model provides significant insights into the understanding of Whorf's cryptotype—in particular, the understanding of complex structural relationships in lexical semantics and the role of a structured semantic representation in the overgeneralization of morphology in language acquisition. Our model demonstrates how cryptotype representations can emerge in connectionist networks as a natural result of the meaning-form mapping processes. These findings help us to better understand the processes underlying important phenomena such as U-shaped behavior in language acquisition.

LEXICAL ORGANIZATION IN DEVELOPMENT

Lexical Categories and Lexical Organization

The previous discussion examined two morphological cases, tense-aspect suffixes and reversative prefixes, and in both cases our focus has been on the emergence of the corresponding lexical semantic categories associated with the uses of the morphology. But lexical categories develop over time in children, and the structural relationships between words can change as learning progresses. In a seminal paper on the acquisition of word meaning, Bowerman (1978) discussed the issue of semantic organization in child language (the essence of the issue was reflected clearly in the title: "Systematizing Semantic Knowledge: Changes over Time in the Child's Organization of Word Meaning"). Bowerman's discussion focused on reorganization (see also Bowerman, 1982): Reorganization in the child's mental lexicon occurs when word pairs that are apparently not initially recognized as semantically related move closer together in meaning. The reorganization is often signaled by speech errors, where semantically similar words compete for selection in production in particular speech contexts—for example, substitutions of *put* for *give* ("put me the bread") or *fall* for *drop* ("I falled it").

Such a reorganization view has important implications for current theories of the mental lexicon. It suggests that we need to take a developmental, dynamic perspective on the lexical system (with respect to both semantic and grammatical properties of the lexicon). A popular trend in cognitive neuroscience today is the attempt to localize various "lexicon modules" in the brain. Using neuroimaging techniques, researchers have identified specific areas in the brain that respond to different lexical categories, such as nouns and verbs, concrete words and abstract words, content words and function words, and words for animals, persons, and tools (e.g., Caramazza & Hillis, 1991; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996, Pulvermueller, 1999). The underlying hypothesis is that different linguistic categories are subserved by different neural substrates, thus supporting the modularity of the mind/brain hypothesis (Fodor, 1983). Neuroscience research in this direction, along with its companion theory of modularity in psychology and philosophy, echoes a long historical tradition in brain localization (Bates, 1999; Gardner, 1986; Uttal, 2001). A fundamental problem with this approach, however, is that it ignores the fact that lexical modules, if they exist, need not be in the brain from the beginning. By assigning an undue weight to a static structure that is "in there," it fails to address the origin of the representation of linguistic categories in the brain.

Taking a developmental, dynamic perspective on this issue, research in my lab attempts to (1) identify crosslinguistic differences in lexical organization and, (2) capture structured, localized, representations of lexical categories as a function of learning and development. First, to identify crosslinguistic differences, in a recent neuroimaging study we found that Mandarin speakers show no distinct neural responses to nouns and verbs in Chinese (both the frontal and the temporal regions were activated by both nouns and verbs), in contrast to findings that distinct cortical regions are involved with nouns versus verbs in other languages (Li, Jin, & Tan, 2004). Second, assuming that these distinct lexical modules do exist in other languages, we have attempted to describe how they could emerge from the learner's organization and reorganization in response to characteristics of the learning environment, in both monolingual and bilingual contexts (see Hernandez, Li, & MacWhinney 2005, for a review; Chan et al., in press for the neural representation of nouns and verbs in bilinguals). To this end, we have developed DevLex, a self-organizing connectionist model of the development of the lexicon. Our model is designed to achieve two goals at the same time: to model the acquisition of lexical organization as a developmental process in child language, and to examine the emergence of categorical representations in our network as a possible explanation of neural representations.

The DevLex Model

Current connectionist models of language acquisition have focused on the examination of phonological patterns rather than meaning structure of words, on the use of artificially generated input rather than realistic linguistic data, and on the use of supervised learning algorithms rather than unsupervised learning (see Li, 2003 for discussion). These limitations led us to consider a model that deals with the acquisition of semantics, with exposure to realistic child-directed parental input, and in self-organizing neural networks with unsupervised learning. Our primary concern has been the development of a psycholinguistically plausible model that can handle realistic linguistic data in the domain of lexical acquisition.

Like most previous connectionist models of language acquisition, the *un*-model I presented earlier was based on the back-propagation learning algorithm. Although significant progress has been made with models based on back-propagation, such models have serious limitations with regard to their neural and psychological plausibility as models of human learning. In particular, "back-propagation networks" are known to suffer from catastrophic forgetting (inability to remember old information with new learning), from scalability (inability to handle realistic, large-scale problems), and above all, from error-driven learning, which adjusts weights according to the error signals from the discrepancy between desired and actual outputs. Some of these problems become most transparent in the context of language acquisition. For example, it would take a strong argument if one claimed that the feedback process used in back-propagation resembles processes of child language learning. Children do not receive constant feedback about what is incorrect in their speech, nor do they get the kind of error corrections on a word-by-word basis that is provided to the network (cf. the "no negative evidence problem" in language acquisition; Baker, 1979; Bowerman, 1988). Instead, much of language acquisition in the natural setting, especially the organization of the mental lexicon, is a self-organizing process that proceeds without explicit teaching.

The DevLex model (Li, Farkas, & MacWhinney, 2004; Li, Zhao, & MacWhinney, 2007) is a type of self-organizing neural network. In contrast to networks with back-propagation learning, self-organizing networks do not require the presence of an explicit teaching signal; learning is achieved entirely by the system's self-organization in response to the input. Self-organization in these networks

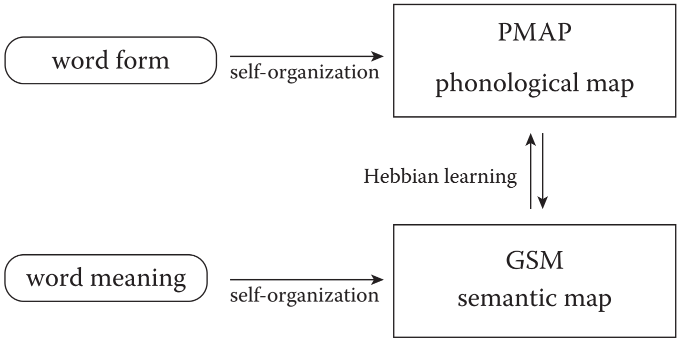

Figure 9.2 DevLex: A self-organizing neural network model of the development of the lexicon.

typically occurs in a two-dimensional map—a self-organizing map (SOM; Kohonen, 1982, 2001). Each processing unit in the network 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 a set of units that surround the best matching unit (the "winner"). Once these units become active in response to a given input, the weights of the winner and those of its neighboring units are adjusted such that they become more similar to the input, and these units will therefore respond to the same or similar inputs more strongly the next time. This process continues until all the inputs can elicit specific response patterns in the network. As a result of this self-organizing process, the network gradually develops concentrated areas of units on the map (the "activity bubbles") that capture input similarities, and the statistical structures implicit in the high-dimensional space of the input are preserved on the 2-D space in the map.

The self-organizing 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 organizes information first in large areas of the map and gradually zeros in onto smaller areas; this zeroing-in is a process from diffuse patterns to focused ones, as a function of the network's continuous adaptation to the input structure. This process allows us to model the emergence of linguistic categories as a gradual process of lexical development. It also has the potential of explaining language disorders that result from the breakdown of focused activation or the inability to form focused representations (Miikkulainen, 1997; Spitzer, 1999).

Figure 9.2 presents a diagrammatic sketch of the DevLex model (for technical details, see Farkas & Li, 2002a; Li, Farkas, & MacWhinney, 2004; see also Li, Zhao, & MacWhinney, 2007 for DevLex-II, an extension of the original DevLex model). It consists of a lexical (phonological) map that processes phonological information of words (PMAP), and a growing semantic map that processes semantic information (GSM). An important feature of the model is that the GSM network can automatically extract semantic and grammatical features of each word by computing the transitional probabilities of co-occurring words in speech context (Farkas & Li, 2001). The size of GSM can also grow along with a growing lexicon in incremental vocabulary learning (Farkas & Li, 2002b; Li et al., 2004). GSM is connected to PMAP via Hebbian learning (Hebb, 1949), according to which the associative strength between two units is increased if the units are both active at the same time. Upon training of the network, a phonological representation of the word is presented to the network, and simultaneously, the semantic representation of the same word is also presented to the network. Through self-organization the network forms 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. At the same time, through Hebbian learning the network forms associations between the two maps for all the active units that respond to the input. The combination of Hebbian learning with self-organization in this way can account for the process of how the learner establishes relationships between semantic features, lexical forms, and morphological markers, on the basis of how often they co-occur and how strongly they are coactivated in the representation.

DevLex based on the above characteristics (1) allows us to track the development of the lexicon clearly as an emergent property in the network's self-organization (from diffuse to focused patterns or from incomplete to complete associative links), (2) allows 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) provides us with a set of biologically plausible and computationally relevant principles to study language acquisition—biologically plausible because the human cerebral cortex can be considered as essentially a self-organizing map (or multiple maps) that compresses information on a 2-D space (Kohonen, 2001; Spitzer, 1999), and computationally relevant because language acquisition in the natural setting (especially organization and reorganization of the lexicon) is largely a self-organizing process (MacWhinney, 1998, 2001). In what follows, I focus on the model's first property, that is, how it is able to model the development of lexical organization. Other aspects of the model, including simulations of the acquisition of cryptotypes and tense-aspect markers and an account of the vocabulary spurt, can be found in Li (2003), Li et al. (2004), and Li et al. (2007).

Lexical Organization in Development

Because DevLex was designed to model a realistic lexicon, we used two sets of child language corpora as the basis of our modeling: the vocabulary from the MacArthur-Bates Communicative Development Inventories (the CDI; Dale & Fenson, 1996) and the parental speech from the CHILDES database (MacWhinney, 2000). From CDI's Toddler List (680 words) we extracted 500 words, excluding homographs, word phrases, and onomatopoeias in the original list. The 500 words were sorted according to their order of acquisition, determined by the CDI lexical norms at the 30th month. In the CDI, early words can be divided into four major categories: (1) nouns, including animals, body, clothing, food, household, outside, people, rooms, toys, and vehicles; (2) verbs; (3) adjectives; and (4) closed-class words, including auxiliary verbs, connecting words, prepositions, pronouns, quantifiers, and question words.

To represent these 500 words as input to DevLex, we first produced the CHILDES parental corpus, which contains the speech transcripts from child-directed adult speech in the CHILDES database (Li, Burgess, & Lund, 2000). Next we presented the sentences (word by word) in the parental corpus to the growing semantic map (GSM). The GSM then computed the lexical co-occurrence statistics for each of the 500 words in terms of the transitional probabilities of successive words, and used these statistics as values in a vector to represent word meaning. One might wonder how much these co-occurrence statistics can capture the semantic as well as the grammatical information of words, but a series of previous experiments indicate the validity of this method in representing the meaning of words (Li, Burgess, & Lind, 2000; Li, Farkas, & MacWhinney, 2004; Farkas & Li, 2001, 2002a, 2000b). To model lexical organization over time, we divided the 500 words into 10 growth stages, each comprising 50 words (cumulatively). Thus, lexical representations may vary (become enriched) from stage to stage, as more and more words are added into the target lexicon.

Figure 9.3 presents snapshots of the GSM at four different stages of learning in the network. These snapshots illustrate the process of lexical organization and reorganization in the network, as a result of the growing lexicon and the development of enriched lexical representations over time. In particular, one can see how the major categories of nouns, verbs, adjectives, and closed-class items start to form coherent classes. At the beginning of learning (stage 1), due to a strong bias in favor of nouns in the CDI vocabulary (and perhaps in child English in general), nouns spread all over the map. A few verbs present in the lexicon are scattered but have not formed any compact clusters. As learning progresses, more verbs, adjectives, and closed-class words enter the vocabulary, and the noun area starts to give way to words in other categories. The developing categories are also clearly reflected in the formation of smaller clusters over several areas on the map—for example, verbs at stage 2. More important for our discussion here, these results not only show the development of major grammatical categories, but also the emergence of semantic categories within the major categories. Most noticeably, within the boundary of nouns a number of compact clusters emerge toward the end of learning. For example, words representing animals, people, household items, food, and body parts (the CDI semantic subcategories) are correspondingly clustered within the nearest neighborhoods on the map. By contrast, the clustering of semantic categories for these words is much less clear at the early stages (i.e., words that belong to the same category are spread farther apart on the map). These results illustrate clearly how lexical organization may change and develop over time, with respect to both grammatical and semantic features of words.

Implications of the Model

As mentioned earlier, DevLex is designed to achieve two goals: to model the acquisition of lexical organization as a developmental process in child language and to examine the emergence of categorical representations in the brain.

With respect to the first goal, our hypothesis is that there are inherent statistical (e.g., distributional) differences between nouns, verbs, adjectives, and closed-class

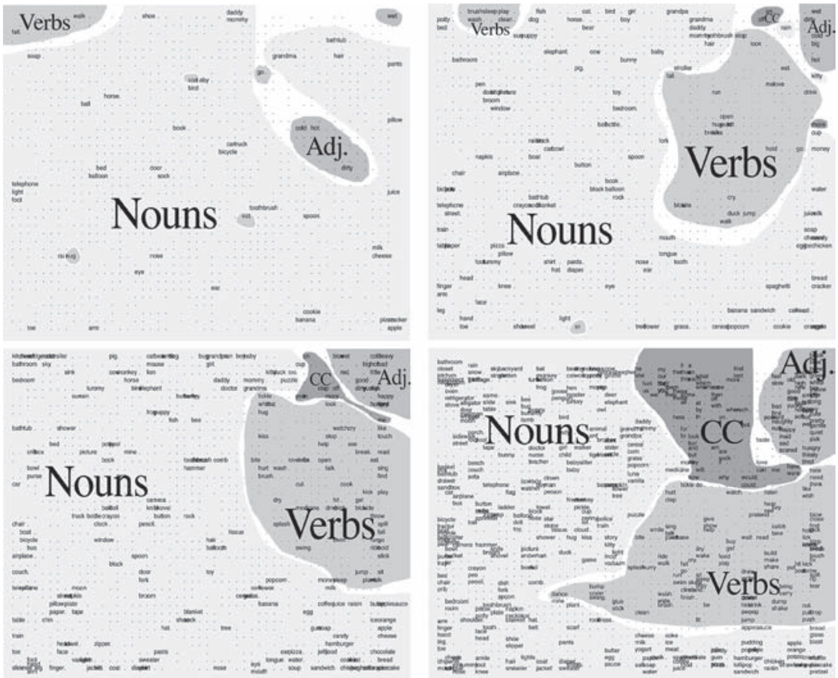

Figure 9.3 Snapshots of GSM at the end of different stages of development: stage 1 (50 words—upper left), 3 (150 words—upper right), 5 (250 words—lower left) and 10 (all 500 words—lower right). The sequence of images illustrates the nature of changes underlying the developmental process, as a result of a developing vocabulary and the changing/enriched word representations. GSM clearly separates the four major categories (and the semantic subcategories within each category) toward the final stage. Because of the large number of words involved in each map, the individual words are not legible in this figure. (Reprinted from *Trends in Cognitive Sciences,* Vol. 9, Hernandez, A., Li, P., & MacWhinney, B: *The emergence of computing modules in bilingualism*, pp. 220–225, Copyright (2005), with permission from Elsevier.)

words in English and that the network can identify these differences through the analysis of lexical co-occurrences. Traditional linguistic analyses (e.g., structuralism) already had ample evidence on the distributional differences between lexical categories (de Saussure, 1916). The question for acquisitionists is how the child can make use of these statistical differences in learning the functions of words, and how this learning can result in different organizational patterns over time.

The DexLex model shows the developmental pathways to lexical representation and organization, in that early on the category memberships are rather diffuse and distributed, and later on with learning, they become more focused and localized. This development from diffuse to focused patterns is consistent with recent findings by Schlaggar et al. (2002) showing that children and adults display different patterns of neural activities in language processing. For children, the activation pattern is diffuse and unfocused, whereas for adults, the activation pattern is more focused, and dedicated to specific cortical regions. Note that this early-diffuse-late-focused pattern is not an artifact of the modulation of the network's training parameters—critical parameters such as learning rate and network size were kept constant across stages in our simulations. Our simulation results also match up with Bowerman's (1978) analyses of the organization and reorganization of the lexicon in development: Many words that are far apart on the map at an earlier stage become grouped together at a later stage. This type of reorganization, within and across categories, clearly serves to structure lexical domains as a whole (Carey, 1978). In addition, according to Bowerman (1978), such reorganization is often signaled by speech errors, where semantically similar words compete for selection in production. In a separate study where production was modeled in DevLex, we found that the organizational structure of the categories has a significant impact on the number of naming errors as well as word confusions in the network (Farkas & Li, 2002b).

With respect to the second goal, our model shows that a developmental connectionist perspective can yield significant insights into the nature and origin of categorical representation in the brain. Cognitive neuroscientists have identified various "brain centers" of language for nouns, verbs, tools, fruits, animals, and so on. An important assumption in many of their studies is that the brain is a highly modularized system, with different cognitive functions localized to different cerebral regions, perhaps from the beginning (the classical "modularity of mind" hypothesis; Fodor, 1983). The ability of DevLex to represent major linguistic categories distinctively without distinct representational modules attests further to the process of "emergent organization" through which localized representations or modules can arise as a function of the developmental processes during ontogenesis, confirming Elizabeth Bates' motto "modules are made, not born" (see Hernandez et al., 2005 for a discussion). Our model shows how the organization of local maps can give rise to emergent categories across stages of learning, with substantial early plasticity and early competition for the organization and reorganization of category members. In our model, the GSM has no pool of units dedicated to any specific category at the outset of learning, but as it progresses through learning, certain groups of units start to develop sensitivity to only the same kinds of words that form coherent categories.

Although there is an obvious difference between "neuronal" activities in our model and neural activities in the brain, DevLex can illustrate the functional mechanisms for the emergence of categorical representations through learning and development. In the spirit of Bowerman's research philosophy, one can often find alternative explanations to a strong brain localizer's account. For example, Farah (1994) pointed out that the well-known deficit with some patients at processing closed-class words might be due to speech stress patterns that differentiate closed-class words from open-class words—that is, patients have trouble dealing with certain words because of the unstressed pattern and short duration, as if damage occurs only to the closed-class words. Similarly, Shi (2006) pointed out that newborn infants are able to categorically discriminate lexical content words from grammatical function words on the basis of multiple acoustic and phonological cues. Functional mechanisms offered by connectionist models such as DevLex are consistent with such explanations, but are often at odds with brain localization accounts in current cognitive neuroscience.

CONCLUSIONS

In this chapter I have discussed the acquisition of word meaning in three domains: tense-aspect acquisition, cryptotype formation, and lexical organization. I have attempted to show that the semantic structure of the lexicon, whether it be overt or covert, can emerge from the interaction between the learner and the linguistic input in the form-meaning mapping process, and we need no a priori assumptions about the innate status or the symbolic nature of semantic categories in the learner's representational system.

The search for meaning is characteristic of children's early lexical and morphological development. It is also characteristic of Bowerman's research emphasis. The three examples that I have discussed here illustrate the importance of semantic learning, the role of linguistic input, and the role of crosslinguistic data, all of which have been carefully examined by Bowerman. In my discussion, I have highlighted how connectionist networks can help us understand the acquisition of lexical semantic representations—in particular, how semantic representations may emerge through the statistical analysis of the linguistic input. Thus, before attributing semantic representations in children as "pre-linguistic" or innate, it is better to first consider mechanisms of learning and the learning environment as potential sources of solution (consistent with Bowerman's approach in "considering the alternatives first"). Our discussions show that the linguistic input contains rich information that the child can exploit in the acquisition of the lexicon, and that modular lexical categories that have often been considered innate (e.g., Bickerton, 1981) may emerge from the learning of statistical properties in language use. Clearly, the data-rich, highly consistent, and statistically regular input in the child's learning environment is at odds with the "poverty of stimulus" hypothesis that the child is exposed to error-laden, random, and inconsistent input. The view that children are statistical learners for language is gaining popularity in recent years since the publication of a number of important volumes and articles (e.g., Elman et al., 1996; MacWhinney, 1998; Saffran, Aslin, & Newport, 1996).

Current debates in cognitive science and cognitive neuroscience revolve around the issue of the nature of linguistic representation. Models based on classical linguistic theories construe linguistic representations in terms of rules in symbol systems. A child is said to have internalized a general rule in her mental representation, "adding *-ed* to make the past tense," at some stage of language acquisition. This kind of description seems intuitively clear, and the rules offer great descriptive power. However, connectionist models provide alternative explanations to this perspective, explanations that place strong emphasis on the statistical learning processes that lead to rule-like behaviors. These models are especially suited for solving problems with which traditional symbolic analyses have difficulty, such as the cryptotype problem discussed here that was once thought "subtle" and "intangible" by Whorf. The distributed representations and adaptive weights used in connectionist models provide mechanisms to capture the complex semantic relationships among words and between words and their morphological markers.

In sum, my conclusion is that structured semantic representations can emerge through statistical computations of the various constraints among lexical items, semantic features, and morphological markers, and the development and organization of the representations are due to basic probabilistic procedures of the sort embodied in connectionist networks in the learning of form-to-form and form-to-meaning mappings. We only need to make a few simple assumptions for this type of probabilistic procedures to work for young children—for example, a working memory that can hold items in sequence and the ability to track distributional regularities (e.g., co-occurrences) in the sequence. Such abilities, as recent studies of statistical learning in infants have revealed, seem to be readily available to the young child at a very early stage (Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997).

ACKNOWLEDGMENTS

Preparation of this article was supported by grants from the National Science Foundation (#BCS-9975249 and #BCS-0131829). I would like to thank Elizabeth Bates, Melissa Bowerman, Brian MacWhinney, and Risto Miikkulainen for their comments and insights on the ideas presented here.

REFERENCES

Aksu, A. (1978). Aspect and modality in the child's acquisition of the Turkish past tense. Unpublished doctoral dissertation, Department of Psychology, University of California, Berkeley.

Aksu-Koç, A., & Slobin, D. (1985). The acquisition of Turkish. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 1, pp. 839–878). Hillsdale, NJ: Erlbaum.

Antinucci, F., & Miller, R. (1976). How children talk about what happened. *Journal of Child Language, 3*, 167–189.

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

Bates, E. (1999). Plasticity, localization and language development. In S. Broman & J. M. Fletcher (Eds.), *The changing nervous system: Neurobehavioral consequences of early brain disorders* (pp. 214–253). New York: Oxford University Press.

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

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences, 7*, 173–188.

Bickerton, D. (1999). Creole languages, the language bioprogram hypothesis, and language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of child language acquisition.* San Diego, CA: Academic Press.

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). Systematizing semantic knowledge: Changes over time in the child's organization of word meaning. *Child Development, 49*, 977–987.

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 319–346). Cambridge, UK: 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. (1985). What shapes children's grammars? In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2, pp. 1257–1319). Hillsdale, NJ: Erlbaum.

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.

Bowerman, M. (1989). Learning a semantic system: What role do cognitive predispositions play? In M. Rice & R. L. Schiefelbusch (Eds.), *The teachability of language* (pp. 133–169). Baltimore, MD: Paul H. Brookes.

Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 385–436). Cambridge, MA: MIT Press.

Bronckart, J., & Sinclair, H. (1973). Time, tense and aspect. *Cognition, 2,* 107–130.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349*, 788–790.

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.

Chan, A.. Luke, K., Li, P., Li, G., Yip, V., Weekes, B. et al. (in press). Neural correlates of nouns and verbs in early bilinguals. *Annals of the New York Academy of Sciences*.

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

Clark, E., 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, UK: Cambridge University Press.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments,* & *Computers, 28*, 125–127 (Information about CDI is available at http://www.sci.sdsu.edu/cdi/#lexnorms).

Damasio, H., Grabowski, T., Tranel, D., Hichwa, R., & Damasio, A. (1996). A neural basis for lexical retrieval. *Nature, 380*, 499–505.

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

Elman, J., Bates, A., Johnson, A., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Erbaugh, M. (1982). *Coming to order: Natural selection and the origin of syntax in the Mandarin speaking child*. Unpublished doctoral dissertation, University of California, Berkeley.

Erbaugh, M. (2006). Chinese classifiers: Their use and acquisition. In P. Li, L. Tan, E. Bates, & O. Tzeng (Eds.), *Handbook of East Asian psycholinguistics: Vol.1. Chinese* (pp. 39–51). Cambridge, UK: Cambridge University Press.

Farah, M. (1994). Neuropsychological inference with an interactive brain: A critique of the "locality" assumption. *Behavioral and Brain Sciences, 17*, 43–104.

Farkas, I., & Li, P. (2001). A self-organizing 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: Erlbaum.

Farkas, I., & Li, P. (2002a). Modeling the development of lexicon with a growing self-organizing map. In H. J. Caulfield et al. (Eds.), *Proceedings of the 6th Joint Conference on Information Sciences* (pp. 553–556). JCIS/Association for Intelligent Machinery, Inc.

Farkas, I., & Li, P. (2002b). DevLex: A self-organizing neural network of the development of lexicon. In *Proceedings of the International Conference on Neural Information Processing*.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press

Gardner, H. (1986). *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

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

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9,* 220–225.

Klein, W., Li, P., & Hendriks, H. (2000). Aspect and assertion in Mandarin Chinese. *Natural Language and Linguistic Theory*, *18*, 723–770.

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

Kohonen, T. (2001). *The self-organizing maps* (3rd ed.). Berlin: Springer.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Li, P. (1990). *Aspect and aktionsart in child Mandarin*. Unpublished doctoral dissertation, Leiden University, the Netherlands.

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

Li, P. (2003). Language acquisition in a self-organising neural network model. In P. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 115–149). Hove, UK & New York: Psychology Press.

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. Clark (Ed.), *Proceedings of the Thirtieth Stanford Child Language Research Forum* (pp. 167–178). New York: Cambridge University Press.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, *17*, 1345–1362.

Li, P., Jin, Z., & Tan, L. (2004). Neural representations of nouns and verbs in Chinese: An fMRI study. *NeuroImage*, *21,* 1533–1541.

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

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

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science, 31,* 531–612.

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

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Erlbaum.

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.

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

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

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

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

Pulvermueller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–336.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart, J. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland, D. Rumelhart, & the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, 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.

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

Schlaggar, B., Brown, T., Lugar, H., Visscher, K., Miezin, F., & Petersen, S. (2002). Functional neuroanatomical differences between adults and school-age children in the processing of single words. *Science, 296*, 1476–1479.

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

Shi, R. (2006). Basic syntactic categories in early language development. In P. Li, L. Tan, E. Bates, & O. Tzeng (Eds.), *Handbook of East Asian psycholinguistics*: *Vol.1. Chinese* (pp. 90–102). Cambridge, UK: Cambridge University Press.

Shirai, Y. (1991). *Primacy of aspect in language acquisition: Simplified input and prototype*. Unpublished doctoral dissertation, University of California, Los Angeles.

Shirai, Y. (1994). On the overgeneralization of progressive marking on stative verbs: Bioprogram or input? *First Language, 14*, 67–82.

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

Shirai, Y., Slobin, D., & Weist, R. (1998). Introduction: The acquisition of tense/aspect morphology. *First Language, 18*, 245–253.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 2, pp. 1158–1249). Hillsdale, NJ: Erlbaum.

Smith, C. (1997). *The parameter of aspect* (2nd ed). Dordrecht: Kluwer.

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.

Tai, J. (1984). Verbs and times in Chinese: Vendler's four categories. In *Papers from the parasession on lexical semantics of the Chicago Linguistic Society* (pp. 289–296).

Uttal, W. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, MA: MIT Press.

Weist, R. (1986). Tense and aspect: Temporal systems in child language. In P. Fletcher & M. Garman (Eds.), *Language acquisition*. Cambridge, UK: Cambridge University Press.

Weist, R., Wysocka, H., Witkowska-Stadnik, K., Buczowska, E., & Konieczna, E. (1984). The defective tense hypothesis: On the emergence of tense and aspect in child Polish. *Journal of Child Language, 11*, 347–374.

Whorf, B. (1956). Thinking in primitive communities. In J. B. Carroll (Ed.), *Language, thought, and reality.* Cambridge, MA: MIT Press. (Original work published 1936)

Zhao, X., & Li, P. (in press). The acquisition of lexical and grammatical aspect in a developmental lexicon model. *Linguistics.*