MODELS OF THE EMERGENCE OF LANGUAGE

Brian MacWhinney
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213,
macw@cmu.edu
Appeared in: Annual Review of Psychology 1998 49: 199-227

Language is a uniquely human achievement. All of the major social achievements of human culture -- architecture, literature, law, science, art, and even warfare -- rely on the use of language. Although there have been attempts to teach language to primates (Allen & Gardner, 1969; Savage-Rumbaugh, Sevcik & Hopkins, 1988), the successful learning of human language seems to be a tightly copyrighted component of our basic human nature.

This view of language as a Special Gift has led some researchers (Bickerton, 1990) to hypothesize that some small set of evolutionary events may have triggered the emergence of language in the human species. Others (Chomsky, 1980; Fodor, 1983) have argued that the capacity to learn language is a unique property of the human mind that is represented neurologically in a separate cognitive module. These scholars believe that this modular architecture allows the shape and form of human language to be largely independent of other aspects of cognitive processing or social functioning. Studies of language learning stimulated by this nativist perspective have tended to focus attention onto a small set of syntactic structures that are thought to constitute the core of Chomsky's Universal Grammar (Chomsky, 1965). According to the "principles and parameters" model of language structure (Hyams & Wexler, 1993), the learning of particular languages occurs through a process called parameter-setting. During parameter setting, children identify the exact shape of their mother tongue by choosing the proper settings on a small set of binary oppositions. For example, a positive setting on the pronoun omission parameter will select for languages like Italian or Chinese, whereas a negative setting will select for English.

Recent studies of the neural basis of communication systems in organisms such as crickets (Wyttenback, May & Hoy, 1996), quail, and song birds (Marler, 1991) have emphasized the extent to which species-specific communication patterns are stored in highly localized hard-wired neurological structures. However, even these lower organisms display some developmental plasticity in the ways in which communication is supported by the brain. When we look at human language learning we see that children learn language gradually and inductively, rather than abruptly and deductively. There is little evidence for a tight biological timetable of developments of the type that we see in other species. In fact, children can learn language even when they have been isolated until an age of even 6 years (Davis, 1947). Throughout the protracted period of human language learning, it is impossible to find evidence for some discrete moment at which a child sets some crucial parameter (Hyams, 1995; MacWhinney & Bates, 1989) that can determine the shape of the native language. Moreover, it is very difficult to use standard experimental methods to prove that children have acquired some of the more abstract categories and structures required by Universal Grammar, such as argument chains, empty categories, landing sites, or dominance relations (Gopnik, 1990; van der Lely, 1994). Despite these empirical problems, the nativist approach remains dominant for studies that investigate the acquisition of formal linguistic structures. For a comprehensive survey of nativist approaches to the acquisition of grammar, the reader may wish to consult Atkinson (1992). Similarly, Markman (1989) summarizes evidence supporting a nativist approach to the acquisition of the lexicon.

NATIVISM AND EMERGENTISM

The inability of nativist accounts to provide accurate or testable accounts of the details of language acquisition has led many language development researchers to explore alternatives to genetically-wired modules. These alternative frameworks emphasize the ways in which the formal structures of language emerge from the interaction of social patterns, patterns implicit in the input, and pressures arising from the biology of the cognitive system. The emergentist approach to language acquisition views language as a structure arising from interacting constraints, much as the shape of the coastline arises from pressures exerted by ocean currents, underlying geology, weather patterns, and human construction. The formalisms that are used to express these nonlinear patterns of interaction include neural network modelling (Fausett, 1994), dynamic systems theory (Port & van Gelder, 1995), and structured approaches such as Optimality Theory {Tesar, in press #7502}. In this chapter, I examine the extent to which neural network models can account for what we currently know about the early states of language development.

THE EMERGENCE OF AUDITORY PATTERNS

During the first year of life, the child goes through a complex set of experiences that lays down an extensive perceptual and motor framework for the learning of the first words. On the perceptual side, the child actively encodes the raw sound patterns of her native language, organizing these patterns into types and sequences. At one time, researchers thought that the learning of perceptual contrasts, such as the ones that allow us to distinguish between "pin" vs "bin", occurred during the second year of life when words are being learned (Jakobson, 1968; Shvachkin, 1948). This picture changed radically when Eimas, Siqueland, Juszcyk, and Vigorito (1971) showed that the ability to detect the contrast between /b/ and /p/ is present soon after birth. Initially, it was thought that these abilities were innate components of a species-specific language gift. However, researchers soon showed that these abilities are shared with other mammals, such as chinchillas (Kuhl & Miller, 1975; Kuhl & Miller, 1978) and monkeys (Kuhl & Padden, 1982; Kuhl & Padden, 1983). It now appears that the ability to discriminate the sounds of language is grounded on raw perceptual abilities of the mammalian auditory system. The sharpness and accuracy of this ability declines during the first year, as children learn to lump together sounds that their language treats as equivalent (Polka & Werker, 1994). In effect, children spend much of the first year of life losing the ability to make contrasts that are not used in the speech they hear about them. Kuhl (1991) has interpreted these findings as evidence for a "preceptual magnet" effect. This effect can be understood by imagining that there is a magnet at the center of each phonemic category that tends to draw in the edges of the category, thereby shortening the distance and leading to an inability to make fine distinctions within this compressed region.

Given the fact that children do not yet understand the words they are hearing, their attentiveness to sound is all the more remarkable. Recent research shows that they are attending not just to the individual phonemes they hear, but even to longer range patterns, such as syllabic sequences. For example, Saffran, Aslin, and Newport (1996) have shown that, when eight-month-old children listen to long sound sequences such as "dabigogatanagotidabigo", they tend to pull out repeated sequences such as "dabigo". As a result, they tend to listen to these familiar sequences more than to similar new sequences.

Infants also demonstrate an early attentiveness to the prosodic characteristics of the language they are hearing. Soon after birth, infants tend to prefer sounds produced by their own mothers to those produced by other women (DeCasper & Fifer, 1980). They also prefer their native languages to other languages (Moon, Cooper & Fifer, 1993). These preferences are probably dependent both on the infant's ability to detect speaker-specific vocal characteristics and on the detection of language-specific prosodic patterns. Infants seem to be sensitive early on to the presence of intonational organization in the language they listen to. Using the sucking habituation technique, Mandel et al. (1994) showed that 2-month-olds tend to remember word strings better when they are presented with normal sentence intonation, than when they are presented as unintegrated lists of words with flat prosody. It appears that stressed intonation may have a particularly important role in picking up auditory strings. Jusczyk and Pisoni (1995) have shown that children tend to pick up and learn stressed syllables above unstressed syllables. However, it also appears that syllables which directly follow after a stressed syllable are also well encoded (Aslin, Jusczyk & Pisoni, 1997). As a result, many of the first sound sequences recorded by the child consist of a stressed peak followed by one or two further weak syllables. This pattern of sound learning has been discussed as a "trochaic bias". However, it can also be viewed as emerging from the combination of a bias to track stressed syllables together with a linear sequence recorder that fires when a stressed syllable is detected.

THE EMERGENCE OF ARTICULATORY PATTERNS

During the first year of life, the infant's articulatory abilities also progress through radical transformations. The basic shape of these changes has been documented since the beginning of the century. We know that children's first vocalizations include the birth cry, the pain cry, the hunger cry, and the pleasure cry. These cries are tightly linked to clear emotional states (Lewis, 1936). By the age of 3 months, children begin a type of social vocalization known as cooing. Around the age of 6 months, children begin a form of sound play that we call babbling. At first, babbling involves the sporadic production of a few simple sounds. These sounds include some strange sounds like clicks that are not found in the input. However, it is not true that each child babbles all the sounds of all the world's languages. Nor is there much evidence for any tight linkage before nine months between the form of the child's babbling and the shape of the input language (Atkinson, MacWhinney & Stoel, 1970; Boysson-Bardies & Vihman, 1991). However, around 11 months, there is increasing evidence for a drift toward the segments and prosody of the target language (Levitt, Utman & Aydelott, 1993), as the child begins to move into the period of the first words.

Initially, it appears that auditory and articulatory development proceed as if largely decoupled. The fact that deaf children babble normally at the age of six months is particularly strong evidence for this conclusion. Given the fact that the brain areas subserving audition (inferior parietal, superior temporal) and articulation (motor cortex) are distant neurologically, this initial decoupling is not too surprising. By the age of nine months, evidence starts to emerge of a connection between babbling and audition. By this age, deaf children, who are not receiving adequate auditory feedback, cease babbling. Normal children start to show the first movement in the prosodic shape of their babbling toward the forms of the input language.

THE EMERGENCE OF THE FIRST WORDS

One of the most active areas of current research in the child language is the study of early word learning. Philosophers like Quine (1960) have emphasized the extent to which word learning needs to be guided by ideas about what might constitute a possible word. For example, if the child were to allow for the possibility that word meanings might include disjunctive Boolean predicates (Hunt, 1962), then it might be the case that the word "grue" would have the meaning "green before the year 2000 and blue thereafter". Similarly, it might be the case that the name for any object would refer not to the object itself, but to its various undetached parts. When one thinks about the word learning task in this abstract way, it appears to be impossibly hard.

Lexical principles

Markman (1989) and Golinkoff, Mervis, and Hirsh-Pasek (1994) have proposed that Quine's problem can be solved by imagining that the child's search for word meanings is guided by lexical principles. For example, children assume that words refer to whole objects, rather than parts of objects. Thus, a child would assume that the word "rabbit" refers to the whole rabbit and not just some parts of the rabbit. However, there is reason to believe that such principles are themselves emergent properties of the cognitive system. For example, Merriman and Stevenson (1997) have argued that the tendency to avoid learning two names for the same object emerges naturally from the competition (MacWhinney, 1989) between closely-related lexical items.

Another proposed lexical principle is the tendency to focus on object names and nominal categories over other parts of speech. Gentner (1982) compared the relative use of nominal terms, predicative terms, and expressive terms in English, German, Japanese, Kaluli, and Turkish. She found that, in all five languages, words for objects constituted the largest group of words learned by the child. Like Gentner, Tomasello (1992) has argued that nouns are easier to "package" cognitively than verbs. Nouns refer to objects that can be repeatedly touched and located in space, whereas verbs refer to transitory actions that are often hard to repeat and whose contour varies markedly for different agents. However, Gopnik and Choi (1990) and Choi and Bowerman (1991) have reported that the first words of Korean-speaking children include far more verbs than do those of English-speaking children. Findings of this type indicate that the nominal bias emerges only in languages that tend to emphasize nouns.

Even in English, we know that children will often treat a new word as a verb or an adjective (Hall, Waxman & Hurwitz, 1993), because words like "run", "want", "hot" and "good" are included in some of the child's first words. Children are also quick to pick up socially-oriented words such as "hi" and "please". As Bloom, Tinker, and Margulis (1993) and Vihman and McCune (1994) have argued, the nominal bias is far from a predominant force, even in English.

Social support

The idea that early word learning depends heavily on the spatio-temporal contiguity of a novel object and a new name can be traced back to Aristotle, Plato, and Augustine. Recently, Baldwin (1991; 1989) has shown that children try to acquire names for the objects that adults are attending to. Similarly, Akhtar, Carpenter, and Tomasello (1996) and Tomasello and Akhtar (1995) have emphasized the crucial role of mutual gaze between mother and child in the support of early word learning. Moreover, Tomasello has argued that human mothers differ significantly from primate mothers in the ways that they encourage mutual attention during language. While not rejecting the role of social support in language learning, Samuelson and Smith (in press) have noted that one can also interpret the findings of Akhtar, Carpenter, and Tomasello in terms of low-level perceptual and attentional matches that help focus the child's attention to novel objects to match up with new words.

Child-based meanings

Several researchers have emphasized the extent to which the shape of the meanings of the first words is governed by a "child-based agenda" (Mervis, 1984; Slobin, 1985). Children seem to be particuarly interested in finding ways of talking about their favorite toys, friends, and foods (Dromi, 1997). They also like to learn words to discuss social activities and functions. In fact, Ninio and Snow (1988) have argued that the basic orientation of the child's first words and early grammar is not towards some objective, nominal, cognitive reality, but towards the interpersonal world involving people and social roles.

Overgeneralization and undergeneralization

We can refer to the formation of a link between a particular referent and a new name as "initial mapping". This initial mapping is typically fast, sketchy, and tentative. Most lexical learning occurs after the formation of this initial mapping. As the child is exposed repeatedly to new instances of an old word, the semantic range of the referent slowly widens. Barrett (1995), Huttenlocher (1974) and others have viewed this aspect of meaning growth as "decontextualization". Harris, Barrett, Jones, and Brookes (1988) have shown that the initial representations of words contain components that are linked to the first few contacts with the word in specific episodes or specific contexts. Gradually, the process of generalization leads to a freeing of the word from irrelevant aspects of the context.

Over time, words develop a separation between a "confirmed core" (1984; 1989) and a peripheral area of potential generalization. As long as the child sticks closely to attested instances of the category inside the confirmed core, she will tend to undergeneralize the word "car". Anglin (1977) and Dromi (1987) have argued that the frequency of such undergeneralizations is typically underestimated, because undergeneralizations never lead to errors. If one does a careful analysis of the range of uses of new words, it appears that undergeneralization is closer to the rule than the exception. As the confirmed core of the meaning of a word widens and as irrelevant contextual features are pruned out, the word begins to take on a radial or prototype form (Lakoff, 1987; Rosch & Mervis, 1975). In the center of the category, we find the best instances that display the maximum category match. At the periphery of the category, we find instances whose category membership is unclear and which compete with neighboring categories (MacWhinney, 1989).

According to the core-periphery model of lexical structure, overgeneralizations arise from the pressures that force the child to communicate about objects that are not inside any confirmed core. Frequently enough, children's overgeneralizations are corrected when the parent provides the correct name for the object (Brown & Hanlon, 1970). The fact that feedback is so consistently available for word learning increases our willingness to believe that the major determinants of word learning are social feedback, rather than innate constraints or even word learning biases.

The shape of vocabulary growth

Researchers have often noted that the growth of the overall size of the lexicon does not follow a smooth linear trend. After the child has acquired an initial vocabulary of about 100 words, the learning of new words seems to progress more and more rapidly. This rapid rise in the size of the vocabulary, which has been called the "vocabulary spurt" (Bates & Carnevale, 1993; Bloom, 1993), is more evident in some children than in others. However, Mervis and Bertrand (1994) and Dromi (1997) have shown that accurate detection of the timing of the vocabulary spurt may require following children well past the first 100 words. Mervis and Bertrand (1995) argue that the timing of the vocabulary spurt is dependent on the rate of cognitive development, with slower developers having a later spurt. They further claim that, before the beginning of the vocabulary spurt, children cannot pick up words through a few brief exposures. However, recent experimental work by Woodward, Markman, and Fitzsimmons (1994) and Schafer and Plunkett (in press) has indicated that infants who have not yet gone through the vocabulary shift are still capable of quick learning of new words in an experimental context.

Three accounts have been offered for the timing of the vocabulary burst and the causes of the burst. One account attributes the burst to the development of control over articulatory representations. Schwartz (1988) and Schwartz and Leonard (1981) have shown that young children tend to avoid producing difficult phonological forms. Once these output limitations are surmounted, the child is free to produce words that had been difficult to produce during earlier periods.

A second account (MacWhinney, 1982) focuses on the role of syntactic patterns in the learning of new words. Often parents make extensive use of stable syntactic frames such as “Here’s the nice (toy name)” or “Show me your (body part name)”. Having learned these frames, children can quickly pick up a large quantity of new words in the context of each frame. In this way, the vocabulary spurt could be dependent upon syntactic development. In fact, Bates et al, (1988) reported a correlation of between .70 and .84 between lexical size at 20 months and syntactic abilities at 28 months. This level of correlation is exactly what is predicted by a model that views lexical learning as facilitated by the appearance of words in the context of well-understood syntactic frames.

In accord with the Piagetian emphasis on cognitive determination of developmental stages, a third group of authors has attributed the vocabulary spurt to the underlying growth in those cognitive capacities (Bloom, 1970; Gopnik & Meltzoff, 1987) that allow children to understand the meanings of new words. For example, one could argue that 14-month-olds are not yet ready conceptually to acquire the meanings of comparative adjectives, conjunctions, abstract nouns, speech act verbs, and superordinates. To be sure, very young children have not yet acquired complex relational concepts, such as the ones required to support the learning of form like “nonetheless”, “preamble”, or “next Thursday” (Kenyeres, 1926). However, attempts to relate overall aspects of linguistic development to fundamental changes in cognitive development have seldom demonstrated strong linkages (Corrigan, 1978; Corrigan, 1979). Instead, it appears that the links between cognitive and lexical development are fragmentary and specific to particular lexical fields (Gopnik & Meltzoff, 1986).

Each of these three accounts is compatible with attempts (Bates & Carnevale, 1993; van Geert, 1991) to model vocabulary growth as a dynamic system using logistic growth functions. The nonlinear effects that emerge during the vocabulary spurt can be viewed as arising from the dynamic coupling of the lexical system with a quickly developing system of syntactic patterns, phonological advances, or cognitive advances. As these various patterns develop, they feed into vocabulary growth in a nonlinear and interactive fashion, as growth in vocabulary leads to further growth in syntactic structures, at least during the several months of the vocabulary spurt.

Components of a model of word learning

We are now ready to explore ways in which these facts about lexical development can be captured in an emergentist model based on neural network theory. The preceding sections indicate that a neural network model of lexical learning will need six components. First, it must provide a system for representing auditory contrasts. Second it must be able to use this system to store frequently heard auditory sequences. Third, the model must be able to account for the development from unconstrained babbling to the controlled articulation of real words. Fourth, the model must be able to account for both social and child-based influences on the meanings underlying the first words. Fifth, the model has to account for the ways in which parents can provide social scaffolding that focuses children’s attention on referents. Sixth, the model must be able to account for both fast initial learning and slow subsequent tuning of the meaning of new words. Seventh, the model must be able to capture facts about the induction of word meanings from syntactic frames.

Neural network models are systems based on the use of a common language of units, connections, weights, and learning rules. Within this common language of connectionism, architectures differ markedly both in their detailed patterns of connectivity and in the specific rules used for activation and learning. There are now many excellent readable introductions to the theory and practice of neural network modeling. The reader who is interested in learning more about the mechanics of this framework may wish to consult Bechtel and Abrahamsen (1991) or Fausett (1994).

Lexical learning as self-organization

One emergentist framework that allows us to model many of these forces is the self-organizing feature map (SOFM) architecture of Kohonen (1982) and Miikkulainen and

Dyer (1990; 1991). These self-organizing networks treat word learning as occurring in maps of connected neurons in small areas of the cortex. Three local maps are involved in word learning: an auditory map, a concept map, and articulatory maps. Emergent self-organization on each of these three maps uses the same learning algorithm. Word learning involves the association of elements between these three maps. What makes this mapping process self-organizing is the fact that there is no pre-established pattern for these mappings and no preordained relation between particular nodes and particular feature patterns.

Evidence regarding the importance of syllables in early child language (Bijeljac, Bertoncini & Mehler, 1993; Jusczyk, Jusczyk, Kennedy, Schomberg & Koenig, 1995) suggests that the nodes on the auditory map may best be viewed as corresponding to full syllabic units, rather than separate consonant and vowel phonemes. The recent demonstration by Saffran et al. (1996) of memory for auditory patterns in four-month-old infants indicates that children are not only encoding individual syllables, but are also remembering sequences of syllables. In effect, prelinguistic children are capable of establishing complete representations of the auditory forms of words. Within the self-organizing framework, these capabilities can be represented in two alternative ways. One method uses a slot-and-frame featural notation from MacWhinney, Leinbach, Taraban, and McDonald (1989). An alternative approach views the encoding as a temporal pattern that repeatedly accesses a basic syllable map. A lexical learning model developed by Gupta and MacWhinney (1997) uses serial processes to control word learning. This model couples a serial order mechanism known as an “avalanche” (Grossberg, 1978) with a lexical feature map model. The avalanche controls the order of syllables within the word. Each new word is learned as a new avalanche.

The initial mapping process involves the association of auditory units to conceptual units. Initially, this learning links concepts to auditory images (Naigles & Gelman, 1995; Reznick, 1990). For example, the 14-month-old who has not yet produced the first word, may demonstrate an understanding of the word “dog” by turning to a picture of a dog, rather than a picture of a cat, when hearing the word “dog”. It is difficult to measure the exact size of this comprehension vocabulary in the weeks preceding the first productive word, but it is probably at least 20 words in size.

In the self-organizing framework, the learning of a word is viewed as the emergence of an association between a pattern on the auditory map and a pattern on the concept map through Hebbian learning (Hebb, 1949; Kandel & Hawkins, 1992). When the child hears a given auditory form and sees an object at the same time, the coactivation of the neurons that respond to the sound and the neurons that respond to the visual form produces an association across a third pattern of connections which maps auditory forms to conceptual forms. Initially, the pattern of these interconnections is unknown, because the relation between sounds and meanings is arbitrary (de Saussure, 1966). This means that the vast majority of the many potential connections between the auditory and conceptual maps will never be used, making it a very sparse matrix (Kanerva, 1993). In fact, it is unlikely that all units in the two maps are fully interconnected (Shrager & Johnson, 1995). In order to support the initial mapping, some researchers (Schmajuk & DiCarlo, 1992) have suggested that the hippocampus may provide a means of maintaining the association until additional cortical connections have been established. As a result, a single exposure to a new word is enough to lead to one trial learning. However, if this initial association is not supported by later repeated exposure to the word in relevant social contexts, the child will no longer remember the word.

Word learning and working memory

The account of word learning we have been examining so far has focused on the learning of the auditory form of the word. In the infant, learning of the articulatory form is typically more delayed. For adults, the task of articulating a newly perceived word is a simple one. However, for the child in the second year of life, matching up articulations to auditions is a major challenge. The simple control of the articulatory system is still a major challenge for the two-year-old. Apart from this, the child have must acquire a mapping from individual auditory features to articulatory gestures, and must also encode the sequence and prosodic contour of each of the syllables in the word. Just like the learning of auditory sequences requires the mediation of memory systems, the learning of articulatory sequences may involve support from rehearsal loops or hippocampal systems.

Models of word learning in adults (Burgess & Hitch, 1992; Grossberg, 1978; Grossberg, 1987; Houghton, 1990)} have tended to emphasize the role of working memory. Gupta and MacWhinney (1997) have shown that a model based on the encoding of syllable strings for output phonology in avalanches does a good job of accounting for a wide variety of well-researched phenomena in the literature on word learning, immediate serial recall, interference effects, and rehearsal in both adults and children (Gathercole & Baddeley, 1993).

The organization of semantic fields

Parallel with the growth of the auditory map, the child is working on the development of an extensive system for conceptual coding. As we have noted, studies of concept development in the preverbal infant (Piaget, 1954; Stiles-Davis, Sugarman & Nass, 1985; Sugarman, 1982) indicate that the child comes to the language learning task already possessing a fairly well-structured coding of the basic objects in the immediate environment. Children treat objects such as dogs, plates, chairs, cars, baby food, water, balls, and shoes as fully structured separate categories (Mervis, 1984). They also show good awareness of the nature of particular activities such as falling, bathing, eating, kissing, and sleeping.

Like auditory categories, these basic conceptual categories can be represented in self-organizing feature maps. Schyns (1991) applied a self-organizing feature map to the task of learning three competing categories with prototype structures. The individual exemplars of each category were derived from geometric patterns that were blurred by noise to create a prototype structure, although the actual prototypes were never displayed. The simulations showed that the network could acquire human-like use of the categories. When presented with a fourth new word that overlapped with one of the first three words, the system broke off some of the territory of the old referent to match up with the new name. This competitive behavior seems to reflect the process of competition between old words and new words discussed for children's word learning by Markman (1989), Clark (1987), and MacWhinney (1989).

Another simulation of meaning development by Li and MacWhinney (1996) used a standard backpropagation architecture to model the learning of reversive verbs that used the prefix "un-" as in "untie" or "dis-" as in "disavow". The model succeeded in capturing the basic developmental stages for reversives reported by Bowerman (1982) and Clark, Carpenter, and Deutsch (1995). In particular, the model was able to produce overgeneralization errors such as "*unbreak" or "*disbend". The network's performance was based on its internalization of what Whorf (1938; 1941) called the "cryptotype" for the reversive which involved a "covering, enclosing, and surface-attaching meaning" that is present in a word like "untangle", but absent in a form such as "*unbreak". Whorf viewed this category as a prime example of the ways in which language reflects and possibly shapes thought.

A similar neural network model of the learning of fine differences in the meaning of the word "over" was developed by Harris (1990; 1994). The Harris model is capable of taking new input test sentences of the type "the pin rolled over the table" and deciding on the basis of past learning that the meaning involved is "across", rather than "covering" or "above". It does this only on the basis of the cooccurrence patterns of the words involved, rather than on information from their individual semantics. Thus, it learns that combinations like "ball", "roll", and "table" tend to activate "across" without regard to facts such as knowing that balls are round and can roll or knowing that tables are flat and that rolling involves movement.

THE EMERGENCE OF INFLECTIONAL MARKING

One of the most active areas in recent work on language acquisition has been the study of the child's learning of inflectional marking. In English, inflections are short suffixes that occur at the ends of words. For example, the word "dogs" has a final /s/ suffix that marks the fact that it is plural. There are now well over 30 empirical studies and simulations investigating the learning of inflectional marking. The majority of work on this topic has examined the learning of English verb morphology with a particular focus on the English past tense. These models are designed to learn irregular forms such as "went" or "fell", as well as regular past tense forms such as "wanted" and "jumped". Other areas of current interest include German noun declension, Dutch stress placement, and German participle formation. Although the learning of inflectional markings is a relatively minor aspect of language learning, our ability to quantify this process has made it an important testing ground not only for the study of child language, but for developmental psychology and cognitive science more generally.

A sample model for inflectional learning

To illustrate how connectionist networks can be used to study the learning of inflectional morphology, let us take as an example the model of German gender learning developed by MacWhinney, Leinbach, Taraban, and McDonald (1989). This model was designed to explain how German children learn to select one of the six different forms of the German definite article. In English we have a single word "the" that serves as the definite article. In German, the article can take the form "der", "die", "das", "des", "dem", or "den". Which of the six forms of the article should be used to modify a given noun in German depends on three additional features of the noun: its gender (masculine, feminine, or neuter), its number (singular or plural), and its role within the sentence (subject, possessor, direct object, prepositional object, or indirect object). To make matters worse, assignment of nouns to gender categories is often quite nonintuitive. For example, the word for "fork" is feminine, the word for "spoon" is masculine, and the word for "knife" is neuter. Acquiring this system of arbitrary gender assignments is particularly difficult for adult second language learners. Mark Twain expressed his consternation at this aspect of German in a treatise entitled "The aweful German language" (Twain, 1935) in which he accuses the language of unfairness and capriciousness in its treatment of young girls as neuter, the sun as feminine, and the moon as masculine. Along a similar vein, Maratsos and Chalkley (1980) argued that, because neither semantic nor phonological cues can predict which article accompanies a given noun in German, children could not learn the language by relying on simple surface cues.

Although these relations are indeed complex, MacWhinney et al. show that it is possible to construct a connectionist network that learns the German system from the available cues. The MacWhinney et al. model, like most current connectionist models, involves a level of input units, a level of hidden units, and a level of output units (Figure 1). Each of these levels or layers contains a number of discrete units or nodes. For example, in the MacWhinney et al. model, the 35 units within the input level represent features of the noun that is to be modified by the article. Each of the two hidden unit levels includes multiple units that represent combinations of these input-level features. The six output units represent the six forms of the German article.

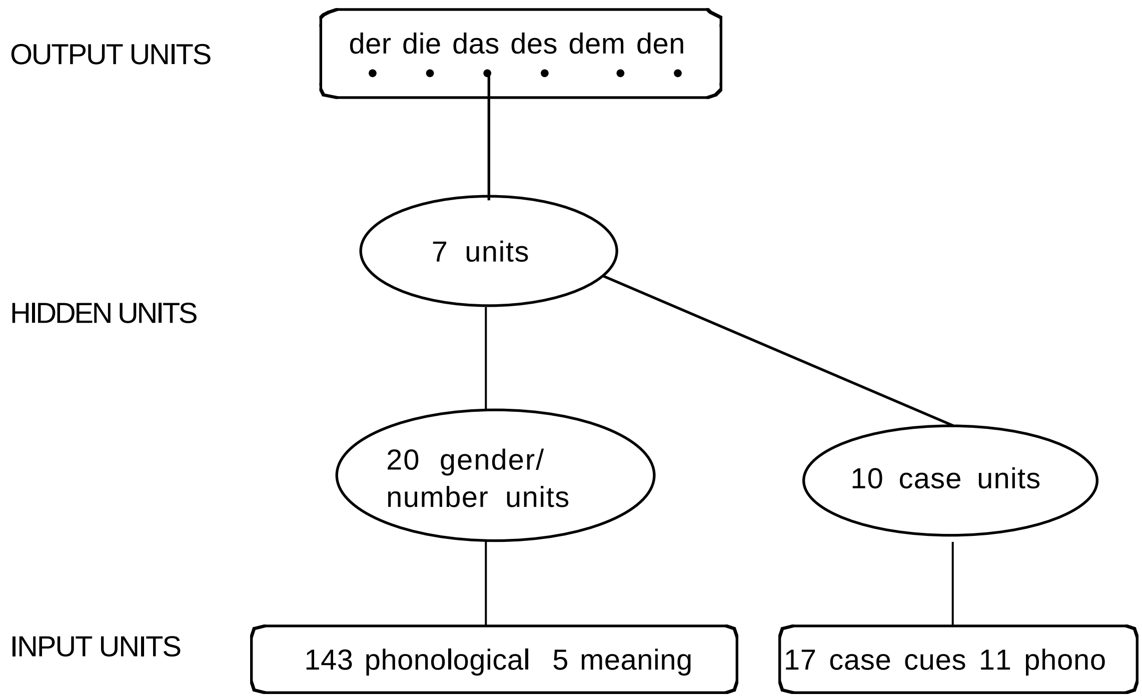

Figure 1: A network model of the acquisition of German declensional marking

As noted above, a central feature of such connectionist models is the very large number of connections among processing units. As shown in Figure 1, each input-level unit is connected to first-level hidden units; each first-level hidden unit is connected to second-level hidden units; and each second-level hidden unit is connected to each of the six output units. None of these hundreds of individual node-to-node connections is illustrated in Figure 1, since graphing each individual connection would lead to a blurred pattern of connecting lines. Instead a single line is used to stand in place of a fully interconnected pattern between levels. Learning is achieved by repetitive cycling through three steps. First, the system is presented with an input pattern that turns on some, but not all of the input units. In this case, the pattern is a set of sound features for the noun being used. Second, the activations of these units send activations through the hidden units and on to the output units. Third, the state of the output units is compared to the correct target and, if it does not match the target, the weights in the network are adjusted so that connections that suggested the correct answer are strengthened and connections that suggested the wrong answer are weakened.

MacWhinney et al. tested this system's ability to master the German article system by repeatedly presenting 102 common German nouns to the system. Frequency of presentation of each noun was proportional to the frequency with which the nouns are used in German. The job of the network was to choose which article to use with each noun in each particular context. After it did this, the correct answer was presented, and the simulation adjusted connection strengths so as to optimize its accuracy in the future. After training was finished, the network was able to choose the correct article for 98 percent of the nouns in the original set.

To test its generalization abilities, we presented the network with old nouns in new case roles. In these tests, the network chose the correct article on 92 percent of trials. This type of cross-paradigm generalization is clear evidence that the network went far beyond rote memorization during the training phase. In fact, the network quickly succeeded in learning the whole of the basic formal paradigm for the marking of German case, number, and gender on the noun. In addition, the simulation was able to generalize its internalized knowledge to solve the problem that had so perplexed Mark Twain -- guessing at the gender of entirely novel nouns. The 48 most frequent nouns in German that had not been included in the original input set were presented in a variety of sentence contexts. On this completely novel set, the simulation chose the correct article from the six possibilities on 61 percent of trials, versus 17 percent expected by chance. Thus, the system's learning mechanism, together with its representation of the noun's phonological and semantic properties and the context, produced a good guess about what article would accompany a given noun, even when the noun was entirely unfamiliar.

The network's learning paralleled children's learning in a number of ways. Like real German-speaking children, the network tended to overuse the articles that accompany feminine nouns. The reason for this is that the feminine forms of the article have a high frequency, because they are used both for feminines and for plurals of all genders. The simulation also showed the same type of overgeneralization patterns that are often interpreted as reflecting rule use when they occur in children's language. For example, although the noun Kleid (which means clothing) is neuter, the simulation used the initial "kl" sound of the noun to conclude that it was masculine. Because of this, it invariably chose the article that would accompany the noun if it were masculine. Interestingly, the same article-noun combinations that are the most difficult for children proved to be the most difficult for the simulation to learn and to generalize to on the basis of previously learned examples.

How was the simulation able to produce such generalization and rule-like behavior without any specific rules? The basic mechanism involved adjusting connection strengths between input, hidden, and output units to reflect the frequency with which combinations of features of nouns were associated with each article. Although no single feature can predict which article would be used, various complex combinations of phonological, semantic, and contextual cues allow quite accurate prediction of which articles should be chosen. This ability to extract complex, interacting patterns of cues is a characteristic of the particular connectionist algorithm, known as back-propagation, that was used in the MacWhinney et al. simulations. What makes the connectionist account for problems of this type particularly appealing is the fact that an equally powerful set of production system rules for German article selection would be quite complex (Mugdan, 1977) and learning of this complex set of rules would be a challenge in itself.

Cues vs. rules

The central issue being addressed in the study of the learning of inflectional markings is whether one can model this process without using formal rules. Rumelhart and McClelland (1986) were the first to provide a demonstration of how rules could emerge from the behavior of neural networks without being explicitly learned. Conceding that irregular forms are produced by connectionist networks, Pinker (1991) nonetheless argues that regular forms are produced by a regular rule. This dual-route model echoes an earlier account by MacWhinney (1978) and related dual-route models in the study of reading by Coltheart, Curtis, Atkins, and Waller (1993)

These attempts to preserve a role for rules in human cognition have run into problems with the fact that even the most regular patterns or "rules" display phonological conditioning and patterns of gradience (Bybee, in press) of the type that are well captured in a connectionist network. Moreover, the existence of differences between regular and irregular processing does not, in itself, provide strong evidence for the existence of rules. Kawamoto (1994) has shown that regular and irregular forms display quite different activation patterns, even within a homogeneous neural network. Therefore, differences in the processing of regular and irregular verbs that have recently been demonstrated through neural imaging work (Jaeger, Lockwood, Kemmerer, Van Valin & Murphy, 1996; Weyerts, Penke, Dohrn, Clahsen & Münte, 1996) do not provide strong evidence for the separate existence of rule system.

U-shaped learning

A major shortcoming of nearly all connectionist models of inflectional learning has been their inability to capture the patterns of overgeneralization and recovery from overgeneralization that have been called "u-shaped" learning. In u-shaped learning, the child begins by correctly producing an irregularly inflected form such as "went". Next, under the pressure of the general pattern, the child produces the overgeneralized form "goed". Finally, the child recovers from overgeneralization and returns to saying "went". Some writers have mistakenly assumed that this type of u-shaped learning applies across all verbs to create three major periods in language learning. However, empirical work by Marcus, Ullman, Pinker, Hollander, Rosen, and Xu (1992) has shown that strong u-shaped learning patterns occur only for some verbs and only for some children.

The modelling of even these weaker u-shaped patterns has proven difficult for neural networks. In order to correctly model the child's learning of inflectional morphology, models must go through a period of virtually error free learning of irregulars, followed by a period of learning of regulars accompanied by the first overregularizations (Marcus et al., 1992). No current model consistently displays all of these features in exactly the right combination. MacWhinney (1997) has argued that models that rely exclusively on backpropagation will never be able to display the correct combination of developmental patterns and that a two-process connectionist approach may be needed (Kawamoto, 1994; Stone, 1994). The basic process is one that learns new inflectional formations, both regular and irregular, as items in self-organizing feature maps. The secondary process is a network that generalizes the information inherent in feature maps to extract secondary productive generalizations. Unlike Pinker's dual-route account, this proposed account works on a uniform underlying connectionist architecture without relying on formal, symbolic linguistic rules.

The role of semantic factors

The first attempts to model morphological learning focused exclusively on the use of phonological features as both input and output. However, it is clear that the formation of past tense forms must also involve semantic factors. In English, the use of semantic information is associated with the irregular patterns of inflection. The idea is that, because we cannot access "went" by combining "go" and "-ed", it might be that we can access it directly by a semantic route. Of course, this idea is much like that underlying the dual-route theory. In German gender, the role of semantic information is much clearer. Köpcke and Zubin (Köpcke, 1994; Köpcke & Zubin, 1983; Köpcke & Zubin, 1984; Zubin & Köpcke, 1981; Zubin & Köpcke, 1986) have shown that a wide variety of both phonological and semantic factors are used in predicting the gender of German nouns and their plural. Some of the features involved include: alcoholic beverages, superordinates, inherent biological gender, gem stones, body parts, rivers inside Germany, and light vs heavy breezes. Simulations by Cottrell and Plunkett (1991) and Gupta and MacWhinney (1992) have integrated semantic and phonological information in various ways. However, a better understanding of the ways in which semantic factors interact during word formation will require a more extensive modeling of lexical items and semantic features.

Extensions of irregular patterns to new words

Extending earlier work by Bybee and Slobin (1982) with older children, Prasada and Pinker (1993) examined the abilities of adult native English speakers to form the past tense for nonsense words like "plink", "plup", or "ploth". They found that, the further the word diverged from the standard phonotactic rules for English verbs, the more likely the subjects were to form the past tense by just attaching the regular "-ed" suffix. Ling and Marinov (1993) noted that the original verb-learning model developed by Rumelhart and McClelland (1987) failed to match these new empirical data, largely because of its tendency to overapply irregular patterns. To correct this problem, Ling and Marinov created a nonconnectionist symbolic pattern associator which did a better job modeling the Prasada and Pinker data. However, MacWhinney (1993a) found that the network model of MacWhinney and Leinbach (1991) worked as well as Ling and Marinov's symbolic model in terms of matching up to the Prasada and Pinker generalization data.

Inflections and the logical problem of language acquisition

In the network we have been discussing, a single lexical feature map can produce both a rote form like "went" and a productive form like "*goed". The fact that both can be produced in the same lexical feature map allows us to develop a general solution to the "logical problem of language acquisition" (Baker & McCarthy, 1981; Gleitman, 1990; Gleitman, Newport & Gleitman, 1984; Morgan & Travis, 1989; Pinker, 1984; Pinker, 1989; Wexler & Culicover, 1980). The logical problem of language acquisition arises from the (incorrect) assumption that recovery from overgeneralization must depend on corrective feedback. Because corrective feedback is seldom available for grammatical patterns (as opposed to lexical and semantic patterns) it can be shown that language learning from input data is impossible. Therefore, it is argued, the acquisition of grammar constitutes a logical problem and requires the postulation of innate constraints on the form of language. The solution to this problem proposed by MacWhinney (1993b) focuses on the competition between regular and irregular forms. In the case of the competition between "went" and "*goed", we expect "went" to become solidified over time because of its repeated occurrence in the input. The form "*goed", on the other hand, is supported only by the presence of the -ed form. Figure 2 illustrates this competition.

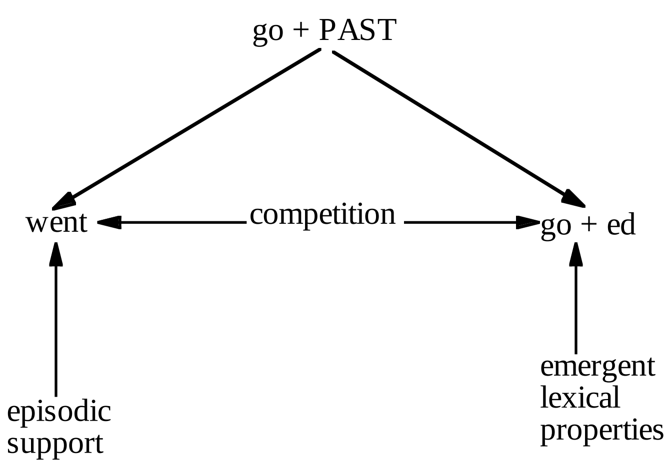

Figure 2: Competition between episodic and combinatorial knowledge

This particular competition is an example of what Baker (1979) calls a "benign exception to the logical problem". The exception is considered benign because the child can learn to block overgeneralization by assuming that there is basically only one way of saying "went". This Uniqueness Constraint is thought to distinguish benign and non-benign exceptions to the logical problem. However, from the viewpoint of the Competition Model account we are constructing here, all exceptions are benign.

The basic idea here is that, when a child overgeneralizes and produces "*goed", the system itself contains a mechanism that eventually forces recovery. Thus, the solution to the logical problem of language acquisition emerges from the competition between alternative competing expressions. One of these forms receives episodic support from the actual linguistic input. This episodic support grows slowly over time. The other form arises productively from the operation of analogistic pressures. When episodic support does not agree with these analogistic pressures, the episodic support eventually comes to dominate and the child recovers from the overgeneralization. This is done without negative evidence, solely on the basic of positive support for the form receiving episodic confirmation.

THE EMERGENCE OF SYNTACTIC PATTERNS

Induction from syntactic frames

Many aspects of word meaning can be acquired from individual words without relying on the role that the word plays in sentences. However, other aspects of meaning require close attention to the ways in which words are combined. In an early demonstration of these effects, Katz, Baker, and Macnamara (1974) gave children a human and a non-human figure and asked them either to “Show me the zav” or to “Show me Zav.” When “zav” was treated syntactically as a proper noun by omitting the definite article, two-year-olds tended to hand the experimenter the figure of a doll. When “zav” was treated syntactically as a common noun by use of the definite article, children tended to hand the experimenter the non-human figure. In this way, even children as young as 20 months of age showed how syntactic context can serve as a powerful guide to word learning.

Similar effects have now been demonstrated for a wide variety of syntactic constructions. Brown (1957) found that children could use a sentence frame like “in this picture you see sibbing” to infer that “sib” is a verb. Carey (1978) and Landau, Smith, and Jones (1992) found that when asked to choose “not the red one, but the xerillium one”, children would assume that “xerillium” was a color name. Golinkoff, Hirsh-Pasek, Cauley, and Gordon (1987) showed children movies in which Big Bird and Cookie Monster were either turning separately or turning each other. When 27-month-old children heard “Big Bird is gorping with Cookie Monster”, they tended to look at the video with both characters turning separately. However, when they heard “Big Bird is gorping Cookie Monster”, they tended to look at the video with Big Bird turning Cookie Monster. These results indicate that children can use the transitive syntactic frame to induce some aspects of the meaning of the new word “gorp”. Similarly, if we ask children to “please repulsate Big Bird the banana”, they will assume that “repulsate” is a verb of transfer that permits a double-object construction. However, if we tell children to “please repulsate the tub with water”, they will assume that “repulsate” is a verb like “fill” that takes a goal as direct object and a transferred object in an instrumental phrase.

Gleitman (1990) has argued that the meanings of words can be induced largely on the basis of this syntactic information. In addition, she has argued that certain aspects of argument structure can only be reliably induced from syntactic frames. However, P. Bloom (1994) has argued that representations acquired in this way would be incomplete. Since children have access to both semantic and syntactic information, it seems likely that both types of information are used whenever they are reliable. In a detailed computational model of verb argument frame induction, Siskind (1996) has shown that, if the child has access to a basic situational representation along with surface cooccurence information, the argument frames of verbs, which are in fact the backbone of the language (Goldberg, 1995; MacWhinney, 1988; Pinker, 1989), can be learned easily even from fairly noisy input data.

The emergence of parts of speech

Psycholinguists working in the standard symbolic tradition (Chomsky, 1965; Fodor & Pylyshyn, 1988; Lachter & Bever, 1988) have pointed to the learning of syntax as a quintessential problem for connectionist approaches. One of the key abilities involved in the learning of syntax is the abstraction of syntactic classes or “parts of speech”, such as nouns, verbs, or prepositions. In the theory of universal grammar, these categories are innately given. However, their actual realization differs so much from language to language that it makes sense to explore accounts that induce these categories from the input data.

Bates and MacWhinney (1982) and MacWhinney (1988) emphasize the extent to which the assignment of words to syntactic classes is heavily dependent upon semantic category structure. Although not all nouns are objects, the best or most prototypical nouns all share this feature. As the category of “noun” radiates out (Lakoff, 1987), non-central members start to share fewer of the core features of the prototype. Maratsos and Chalkley (1980) point out that words like “justice” and “lightning” are so clearly non-objects that their membership in the class of nouns cannot be predicted from their semantic status and can only be inferred from the fact that the language treats them as nouns. Although Bates and MacWhinney (1982) and Maratsos and Chalkley (1980) staked out strongly contrasting positions on this issue, each of the approaches granted the possibility that both cooccurence and semantic factors play a major role in the emergence of the parts of speech.

At this point, language researchers are primarily interested in exploring detailed models that show exactly how the parts of speech and argument frames can be induced. Elman (1993) has presented a connectionist model that does just this. The model relies on a recurrent architecture of the type presented in Figure 3. This model takes the standard three-layer architecture of pools A, B, and C and adds a fourth input pool D of context units which has recurrent connections to pool B. Because of the recurrent or bidirectional connections between B and D, this architecture is know as “recurrent backpropagation”.

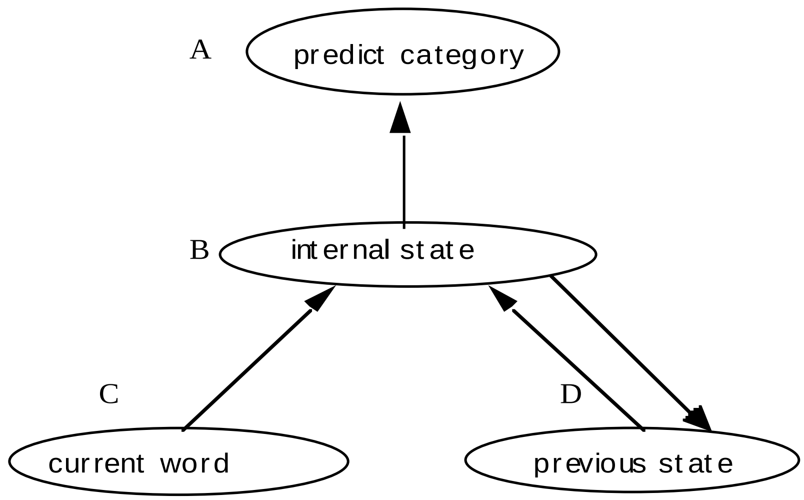

Figure 3: A recurrent network

A recurrent backpropagation network encodes changes over time by storing information regarding previous states in the pool of units labeled as D. Consider how the network deals with the processing of a sentence such as “Mommy loves Daddy”. When the first word comes in, pool C is activated and this activation is passed on to pool B and then pools A and D. The complete state of pool B at Time 1 is stored in pool D. The activation levels in pool D are preserved, while pools A, B, and C are set back to zero. At time 2 the networks hears the word “love” and a new pattern of activations is established on pool C. These activations are passed on to pools B, C, and D. However, because pool D has stored activations from the previous word, the new state is blended with the old state and pool C comes to represent aspects of both “Mommy” and “love”.

Processing in a network of this type involves more than just storage of a superficial sequence of words or sounds. For example, in the simulations of sentence processing developed by Elman (1993), the output units are trained to predict the identity of the next word. In order to perform in this task, the network needs to implicitly extract part-of-speech information from syntactic cooccurrence patterns. Alternatively, the output units can be used to represent comprehension decisions, as in the model of MacWhinney (1997). In that model, part-of-speech information is assumed and the goal of the model is to select the agent and the patient using a variety of grammatical and pragmatic cues.

The training set for the model consists of dozens of simple English sentences such as “The big dog chased the girl.” By examining the weight patterns on the hidden units in the fully trained model, Elman showed that the model was conducting implicit learning of the parts of speech. For example, after the word “big” in our example sentence, the model would be expecting to activate a noun. The model was also able to distinguish between subject and object relative structures, as in “the dog the cat chased ran” and “the dog that chased the cat ran”.

The emergence of argument structures

The machinery governing word combinations depends not only on part of speech information, but also on information regarding detailed aspects of argument structure. Consider the use of verbs like “pour” and “fill”. Bowerman (1988) discusses cases in which the child says “I poured the tub with water” instead of “I pouredwater into the tub” and “I filled water into the tub” instead of “I filled the tub with water.” We can describe these errors by saying that the child has overgeneralized the “pour” pattern to the word “fill” or overgeneralized the “fill” pattern to the word “pour”. In order to avoid these overgeneralizations and to recover from them once they are made, children have to organize verbs into semantic fields. Extending the self-organizing topological network approach we examined earlier, we can model this process by building a network in which the use of the pattern “V N with N” is correlated with the semantic features of words like “fill”, “stuff”, and “load”, and in which the use of the pattern “V N into N” is correlated with the semantic features of words like “fill”, “paint”, “cover”, and “load”. Because of network of this type uses semantic features to achieve a separation on the argument frame map, it is able to implement both the semantic proposals of Bates and MacWhinney (1982) and the cooccurrence proposals of Maratsos and Chalkley (1980).

CONCLUSION

In this chapter, we have seen how neural network models can help us organize our growing understanding of auditory, articulatory, lexical, inflectional, and syntactic development. There are many aspects of language development to which these models have not yet been applied. We do not yet have models that can learn to control sociolinguistic relations, conversational patterns, narrative structures, intonational contours, and gestural markings. Even in the areas to which they have been applied, emergentist models are limited in many ways. The treatment of the more complex aspects of syntax remains unclear, the modelling of lexical extensions is still quite primitive, and the development of the auditory and articulatory systems is not yet sufficiently grounded in physiological and neurological facts. Despite these limitations, we can see that, by treating language learning as an emergent process, these models have succeeded in providing an exciting new perspective on questions about language learning that have intrigued scholars for centuries.

LITERATURE CITED

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. Child Development, 62, 635-645.
Allen, R., & Gardner, B. (1969). Teaching sign language to a chimpanzee. Science, 165, 664-672.
Anglin, J. M. (Ed.). (1977). Word, object, and conceptual development. New York: Norton.
Aslin, R., Jusczyk, P., & Pisoni, D. (1997). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), Handbook of child psychology. Volume 2, . New York: Wiley.
Atkinson, K., MacWhinney, B., & Stoel, C. (1970). An experiment on the recognition of babbling. Papers and Reports on Child Language Development, 5, 1-8.
Atkinson, M. (1992). Children's syntax. Oxford: Blackwells.
Baker, C. L. (1979). Syntactic theory and the projection problem. Linguistic Inquiry, 10, 533-581.
Baker, C. L., & McCarthy, J. J. (Eds.). (1981). The logical problem of language acquisition. Cambridge: MIT Press.
Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. Child Development, 62, 875-890.
Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. Child Development, 60, 381-398.
Barrett, M. (1995). Early lexical development. In P. Fletcher & B. MacWhinney (Eds.), Handbook of Child Language, . Oxford: Basil Blackwell.
Bates, E., Bretherton, I., & Snyder, L. (1988). From first words to grammar: Individual differences and dissociable mechanisms. Cambridge, MA: Cambridge University Press.
Bates, E., & Carnevale, G. (1993). New directions in research on language development. Developmental Review, 13, 436-470.
Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), Language acquisition: The state of the art, (pp. 173-218). New York: Cambridge University Press.
Bechtel, W., & Abrahamsen, A. (1991). Connectionism and the mind: An introduction to parallel processing in networks. Cambridge, MA: Basil Blackwell.
Bickerton, D. (1990). Language and species. Chicago: Chicago University Press.
Bijeljac, B., R., Bertoncini, J., & Mehler, J. (1993). How do four-day-old infants categorize multisyllabic utterances? Developmental Psychology, 29, 711-721.
Bloom, L. (1970). Language development: Form and function in emerging grammars. Cambridge, MA: MIT Press.
Bloom, L. (1993). The transition from infancy to language: Acquiring the power of expression. Cambridge: Cambridge University Press.
Bloom, L., Tinker, E., & Margulis, C. (1993). The words children learn: Evidence against a noun bias in early vocabularies. Cognitive Psychology, 8, 431-450.
Bloom, P. (1994). Overview: Controversies in language acquisition. In P. Bloom (Ed.), Language acquisition: Core readings, (pp. 5-48). Cambridge, MA: MIT Press.
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). New York: Cambridge University Press.
Bowerman, M. (1988). The "no negative evidence" problem. In J. Hawkins (Ed.), Explaining language universals, (pp. 73-104). London: Blackwell.

Boysson-Bardies, B., & Vihman, M. M. (1991). Adaption to language: Evidence from babbling and first words in four languages. Language, 67, 297-320.
Brown, R. (1957). Linguistic determinism and the part of speech. Journal of Abnormal and Social Psychology, 55, 1-5.
Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), Cognition and the development of language, (pp. 11-54). New York: Wiley.
Burgess, N., & Hitch, G. (1992). Toward a network model of the articulatory loop. Journal of Memory and Language, 31, 429-460.
Bybee, J. (in press). Regular morphology and the lexicon. Cognition.
Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past. Language, 58, 265-289.
Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), Linguistic theory and psychological reality, (pp. 264-293). Cambridge, MA: MIT Press.
Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. Cognition, 41, 83-121.
Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.
Chomsky, N. (1980). Rules and Representations. New York: Columbia University Press.
Clark, E. (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.
Clark, E., Carpenter, K., & Deutsch, W. (1995). Reference states and reversals: Undoing actions with verbs. Journal of Child Language, 22, 633-652.
Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel distributed processing approaches. Psychological Review, 100, 589-608.
Corrigan, R. (1978). Language development as related to stage 6 object permanence development. Journal of Child Language, 5, 173-189.
Corrigan, R. (1979). Cognitive correlates of language: Differential criteria yield differential results. Child Development, 50, 617-631.
Cottrell, G., & Plunkett, K. (1991). Learning the past tense in a recurrent network: Acquiring the mapping from meaning to sounds. In K. Hammond & D. Gentner (Eds.), Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, (pp. 109-115). Hillsdale, NJ: Lawrence Erlbaum.
Davis, K. (1947). Final note on a case of extreme social isolation. American Journal of Sociology, 52, 432-437.
de Saussure, F. (1966). Course in general linguistics. New York: McGraw-Hill.
DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. Science, 208, 1174-1176.
Dromi, E. (1987). Early lexical development. New York: Cambridge University Press.
Dromi, E. (1997). Early lexical development. In M. Barrett (Ed.), The development of language, . London: UCL Press.
Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. Science, 171, 303-306.
Elman, J. (1993). Incremental learning, or the importance of starting small. Cognition, 48, 71-99.
Fausett, L. (1994). Fundamentals of neural networks. Englewood Cliffs, NJ: Prentice Hall.
Fodor, J. (1983). The modularity of mind: An essay on faculty psychology. Cambridge, Mass.: M. I. T. Press.
Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. Cognition, 28, 3-71.
Gathercole, V., & Baddeley, A. (1993). Working memory and language. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), Language development: Language, culture, and cognition, (pp. 301-334). Hillsdale, NJ: Lawrence Erlbaum.
Gleitman, L. (1990). The structural sources of verb meanings. Language Acquisition, 1, 3-55.
Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. Journal of Child Language, 11, 43-79.
Goldberg, A. (1995). Constructions. Chicago: University of Chicago Press.
Golinkoff, R., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. Journal of Child Language, 14, 23-46.
Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. Journal of Child Language, 21, 125-155.
Gopnik, A., & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A crosslinguistic study of semantic and cognitive development. First Language, 10, 199-215.
Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to the other cognitive and linguistic developments. Child Development, 58, 1523-1531.
Gopnik, A., & Meltzoff, A. N. (1986). Relations between semantic and cognitive development in the one-word stage: The Specificity Hypothesis. Child Development, 57, 1040-1053.
Gopnik, M. (1990). Feature blindness: A case study. Language Acquisition, 1, 139-164.
Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. Progress in Theoretical Biology, 5, 233-374.
Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. Cognitive Science, 11, 23-63.
Gupta, P., & MacWhinney, B. (1992). Integrating category acquisition with inflectional marking: A model of the German nominal system, Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, (pp. 253-258). Hillsdale, NJ: Lawrence Erlbaum Associates.
Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. Brain and Language, 59, 267-333.
Hall, D., Waxman, S., & Hurwitz, W. (1993). How two- and four-year-old children interpret adjectives and count nouns. Child Development, 64, 1651-1664.
Harris, C. (1990). Connectionism and cognitive linguistics. Connection Science, 2, 7-33.
Harris, C. L. (1994). Back-propagation representations for the rule-analogy continuum. In J. Barnden & K. Holyoak (Eds.), Analogical connections, (pp. 282-326). Norwood, NJ: Ablex.
Harris, M., Barrett, M. D., Jones, D., & Brookers, S. (1988). Linguistic input and early word meaning. Journal of Child Language, 15, 77-94.
Hebb, D. (1949). The organization of behavior. New York: Wiley.
Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), Current research in natural language generation, (pp. 287-319). London: Academic.
Hunt, E. (1962). Concept learning: an information processing approach. New York: Wiley.
Huttenlocher, J. (1974). The origins of language comprehension. In R. Solso (Ed.), Theories in cognitive psychology: The Loyola symposium, (pp. 331-388). Potomac, Maryland: Lawrence Erlbaum.
Hyams, N. (1995). Nondiscreteness and variation in child language: Implications for Principle and Parameter modesl of language development. In Y. Levy (Ed.), Other children, other languages, (pp. 11-40). Hillsdale, NJ: Lawrence Erlbaum.
Hyams, N., & Wexler, K. (1993). On the grammatical basis of null subjects in child language. Linguistic Inquiry, 24(3), 421-459.

Jaeger, J. J., Lockwood, A. H., Kemmerer, D. L., Van Valin, R. D., & Murphy, B. W. (1996). A positron emission tomographic study of regular and irregular verb morphology in English. Language, 72, 451-497.
Jakobson, R. (1968). Child language, aphasia and phonological universals. The Hague: Mouton.
Jusczyk, P., & Aslin, R. (1995). Infants' detection of the sound patterns of words in fluent speech. Cognitive Psychology, 29, 1-23.
Jusczyk, P. W., Jusczyk, A. M., Kennedy, L. J., Schomberg, T., & Koenig, N. (1995). Young infants' retention of information about bisyllabic utterances. Journal of Experimental Psychology: Human Perception and Performance, 21, 822-836.
Kandel, E. R., & Hawkins, R. D. (1992). The biological basis of learning and individuality. Scientific American, 266, 40-53.
Kanerva, P. (1993). Sparse distributed memory and related models. In M. Hassoun (Ed.), Associative neural memories: Theory and implementation, . New York: Oxford University Press.
Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? A study of how children learn common and proper names. Child Development, 45, 469-473.
Kawamoto, A. (1994). One system or two to handle regulars and exceptions: How time-course of processing can inform this debate. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), The reality of linguistic rules, (pp. 389-416). Amsterdam: John Benjamins.
Kenyeres, E. (1926). A gyermek elsö szavai es a szófajók föllépése. Budapest: Kisdednevelés.
Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43, 59-69.
Köpcke, K.-M. (1994). Funktionale Untersuchungen zur deutschen Nominal- und Verbalmorphologie. Linguistiche Arbeiten, 319, 81-95.
Köpcke, K. M., & Zubin, D. A. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. Zeitschrift für germanistische Linguistik, 11, 166-182.
Köpcke, K. M., & Zubin, D. A. (1984). Sechs Prinzipien fur die Genuszuweisung im Deutschen: ein Beitrag zur natürlichen Klassifikation. Linguistische Berichte, 93, 26-50.
Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Perception and Psychophysics, 50, 93-107.
Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonsants. Science, 190, 69-72.
Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 63, 905-917.
Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. Perception and Psychophysics, 32, 542-550.
Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. Journal of the Acoustical Society of America, 73, 1003-1010.
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.
Lakoff, G. (1987). Women, fire, and dangerous things. Chicago: Chicago University Press.
Landau, B., Smith, L., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. Journal of Memory and Language, 31, 807-825.
Levitt, A. G., Utman, J., & Aydelott, J. (1993). From babbling towards the sound systems of English and French: A longitudinal two-case study. Journal of Child Language, 19, 19-49.

Lewis, M. M. (1936). Infant speech: A study of the beginnings of language. New York: Harcourt, Brace and Co.
Li, P., & MacWhinney, B. (1996). Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. Connection Science, 8, 3-30.
Ling, C., & Marinov, M. (1993). Answering the connectionist challenge. Cognition, 49, 267-290.
MacWhinney, B. (1978). The acquisition of morphophonology. Monographs of the Society for Research in Child Development, 43, Whole no. 1, pp. 1-123.
MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), Language acquisition: vol 1. Syntax and semantics, (pp. 73-136). Hillsdale, NJ: Lawrence Erlbaum.
MacWhinney, B. (1984). Where do categories come from? In C. Sophian (Ed.), Child categorization, (pp. 407-418). Hillsdale, N.J.: Lawrence Erlbaum.
MacWhinney, B. (1988). Competition and teachability. In R. Schiefelbusch & M. Rice (Eds.), The teachability of language, (pp. 63-104). New York: Cambridge University Press.
MacWhinney, B. (1989). Competition and lexical categorization. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), Linguistic categorization, (pp. 195-242). New York: Benjamins.
MacWhinney, B. (1993a). Connections and symbols: Closing the gap. Cognition, 49, 291-296.
MacWhinney, B. (1993b). The (il)logical problem of language acquisition, Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, (pp. 61-70). Hillsdale, NJ: Lawrence Erlbaum Associates.
MacWhinney, B. (1997). Lexical connectionism. In P. Broeder & J. Murre (Eds.), Models of language acquisition: Inductive and deductive approaches, . Cambridge, MA: MIT Press.
MacWhinney, B., & Bates, E. (Eds.). (1989). The crosslinguistic study of sentence processing. New York: Cambridge University Press.
MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. Cognition, 29, 121-157.
MacWhinney, B. J., Leinbach, J., Taraban, R., & McDonald, J. L. (1989). Language learning: Cues or rules? Journal of Memory and Language, 28, 255-277.
Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. G. (1994). Does sentence prosody help infants to organize and remember speech information? Cognition, 53, 155-180.
Maratsos, M., & Chalkley, M. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), Children's language: Volume 2, (pp. 127-214). New York: Gardner.
Marcus, G., Ullman, M., Pinker, S., Hollander, M., Rosen, T., & Xu, F. (1992). Overregularization in language acquisition. Monographs of the Society for Research in Child Development, 57(4), 1-182.
Markman, E. (1989). Categorization and naming in children: Problems of induction. Cambrdige, MA: MIT Press.
Marler, P. (1991). Song-learning behavior: the interface with neuroethology. Trends in Neuroscience, 14, 199-206.
Merriman, W. E., & Stevenson, C. M. (1997). Restricting a familiar name in response to learning a new one: Evidence for the mutual exclusivity bias in young 2-year-olds. Child Development, 68, 211-258.
Mervis, C. (1984). Early lexical development: The contributions of mother and child. In C. Sophian (Ed.), Origins of cognitive skills, (pp. 339-370). Hillsdale, N.J.: Lawrence Erlbaum.
Mervis, C., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (NC3) principle. Child Development, 65, 1646-1662.

Mervis, C., & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: a response to Goldfield and Reznick. Journal of Child Language, 22, 461-468.
Miikkulainen, R. (1990). A distributed feature map model of the lexicon, Proceedings of the 12th Annual Conference of the Cognitive Science Society, . Hillsdale, NJ: Lawrence Erlbaum Associates.
Miikkulainen, R., & Dyer, M. (1991). Natural language processing with modular neural networks and distributed lexicon. Cognitive Science, 15, 343-399.
Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day infants prefer their native language. Infant Behavior and Development, 16, 495-500.
Morgan, J., & Travis, L. (1989). Limits on negative information in language input. Journal of Child Language, 16, 531-552.
Mugdan, J. (1977). Flexionsmorphologie und Psycholinguistik. Tübingen: Gunter Narr.
Naigles, L. G., & Gelman, S. A. (1995). Overextensions in comprehension and production revisited: Preferential looking in a study of dog, cat, and cow. Journal of Child Language, 22, 19-46.
Ninio, A., & Snow, C. (1988). Language acquisition through language use: The functional sources of children's early utterances. In Y. Levy, I. Schlesinger, & M. Braine (Eds.), Categories and processes in language acquisition, (pp. 11-30). Hillsdale, NJ: Lawrence Erlbaum.
Piaget, J. (1954). The construction of reality in the child. New York: Basic Books.
Pinker, S. (1984). Language learnability and language development. Cambridge, Mass: Harvard University Press.
Pinker, S. (1989). Learnability and cognition: the acquisition of argument structure. Cambridge: MIT Press.
Pinker, S. (1991). Rules of Language. Science, 253, 530-535.
Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20, 421-435.
Port, R. F., & van Gelder, T. (Eds.). (1995). Mind as motion. Cambridge, MA: MIT Press.
Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. Language and Cognitive Processes, 8, 1-56.
Quine, W. V. O. (1960). Word and object. Cambridge, MA: MIT Press.
Reznick, S. (1990). Visual preference as a test of infant word comprehension. Applied Psycholinguistics, 11, 145-166.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.
Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland & D. E. Rumelhart (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition, (pp. 216-271). Cambridge: MIT Press.
Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processes? In B. MacWhinney (Ed.), Mechanisms of Language Acquisition, (pp. 195-248). Hillsdale, N.J.: Lawrence Erlbaum.
Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.
Samuelson, L. K., & Smith, L. B. (in press). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. Child Development, xx, xx.
Savage-Rumbaugh, S., Sevcik, R. A., & Hopkins, W. D. (1988). Symbolic cross-modal transfer in two species of chimpanzees. Child Development, 59, 617-625.
Schafer, G., & Plunkett, K. (in press). Rapid word learning by 15-month-olds under tightly controlled conditions. Child Development, xx, xx-xx.
Schmajuk, N., & DiCarlo, J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. Psychological Review, 99, 268-305.

Schwartz, R. G. (1988). Phonological factors in early lexical acquisition. In M. D. Smith & J. L. Locke (Eds.), The emergent lexicon: The child's development of a linguistic vocabulary, . New York: Academic Press.
Schwartz, R. G., & Leonard, L. B. (1981). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. Journal of Child Language, 9, 319-336.
Schyns, P. (1991). A modular neural network model of concept acquisition. Cognitive Science, 15, 461-508.
Shrager, J. F., & Johnson, M. H. (1995). Waves of growth in the development of cortical function: A computational model. In B. Julesz & I. Kovacs (Eds.), Maturational windows and adult cortical plasticity, (pp. 31-44). New York: Addison-Wesley.
Shvachkin, N. (1948). Razvitiye fonematicheskogo vospriyatiya rechi v rannem vozraste. Izvestiva Akademii Pedagogicheskikh Nauk RSFSR, 13, 101-132.
Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. Cognition, 61, 39-91.
Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), The crosslinguistic study of language acquisition. Volume 2: Theoretical issues, (pp. 1157-1256). Hillsdale, N. J.: Lawrence Erlbaum.
Stiles-Davis, J., Sugarman, S., & Nass, R. (1985). The development of spatial and class relations in four young children with right-cerebral-hemisphere damage: Evidence for an early spatial constructive deficit. Brain and Cognition, 4(4), 388-412.
Stone, G. (1994). Combining connectionist and symbolic properties in a single process. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), The reality of linguistic rules, (pp. 417-444). Amsterdam: John Benjamins.
Sugarman, S. (1982). Developmental change in early representational intelligence: Evidence from spatial classification strategies and related verbal expressions. Cognitive Psychology, 14, 410-449.
Tomasello, M. (1992). First verbs: A case study of early grammatical development. Cambridge: Cambridge University Press.
Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. Cognitive Development, 10, 201-224.
Twain, M. (1935). The aweful German language, The family Mark Twain, . New York: Harper & Brothers.
van der Lely, H. (1994). Canonical linking rules: Forward vs. reverse linking in normally developing and Specifically Language Impaired children. Cognition, 51, 29-72.
van Geert, P. (1991). A dynamic systems model of cognitive and language growth. Psychological Review, 98, 3-53.
Vihman, M. M., & McCune, L. (1994). When is a word a word? Journal of Child Language, 21, 517-542.
Wexler, K., & Culicover, P. (1980). Formal principles of language acquisition. Cambridge, Mass.: MIT Press.
Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., & Münte, T. (1996). Brain potentials indicate differences between regular and irregular German noun plurals. Essex Research Reports in Linguistics, 13, 54-67.
Whorf, B. (1938). Some verbal categories of Hopi. Language, 14, 275-286.
Whorf, B. (1941). The relation of habitual thought and behaviour to language. In L. Spier (Ed.), Language, culture, and personality: Essays in memory of Edward Sapir, (pp. 75-93). Ogden, Utah: University of Utah Press.
Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. Developmental Psychology, 30, 553-566.
Wyttenback, M., May, M., & Hoy, D. (1996). Categorical perception of sound frequency by crickets. Science, 273, 1542-1544.

Zubin, D. A., & Köpcke, K. M. (1981). Gender: A less than arbitrary grammatical category. In R. Hendrick, C. Masek, & M. Miller (Eds.), Papers from the Seventeenth Regional Meeting, (pp. 439-449). Chicago: Chicago Linguistic Society.

Zubin, D. A., & Köpcke, K. M. (1986). Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. In C. Craig (Ed.), Noun classes and categorization, (pp. 139-180). Amsterdam: John Benjamins.