

Connectionism and Language Learning

Brian MacWhinney

Franklin Chang

Carnegie Mellon University

In S. Kemmer (Ed.). 2000. Data-driven models of language learning. CSLI Press: Stanford.

1. Symbols and Connections

Linguistic behavior is governed by a rigid set of social conventions or “rules.” If we wake up one morning and decide to deliberately throw all these conventions to the wind, no one would understand us. Indeed, even our best friends might think we had gone quite insane. In everyday language, we could say that we had decided to “break the rules” of English grammar. Of course, force of habit inclines us against striking off on this iconoclastic course. Having spent so many years of our lives cooperatively following “the rules,” it is easier to continue to follow them than to wander off into new territory. This view of linguistic rules as social conventions and habits is grounded firmly on everyday experience and common sense. I think it is a view that virtually everyone accepts.

During the 60s and 70s, scientists took this common sense idea of a linguistic rule and reworked it into a basic principle underlying artificial intelligence (AI), Chomskyan theoretical linguistics, and cognitive psychology. By viewing the brain as a computer, they began to think of the mind as a system for transforming symbolic strings according to well-specified rules. The vision of human language as a system of formal rules was an important ingredient underlying two decades of work in linguistics and cognitive science. This work led to the emergence of complex and impressive systems of rules and symbols based on what I have called the “Big Mean Rules” (MacWhinney 1994) and the “Big Mean Flowcharts.”

In recent years, the biological and epistemological underpinnings of these great symbolic systems have become increasingly shaky and vulnerable. Two basic observational problems faced by all of these analyses are the fact that no developmental psychologist ever observed a child learning a rule and that no neuroscientist ever traced the neural substrate of either a rule or a symbol. Similarly, attempts in the 1970s to demonstrate the psychological reality of rules in adults (Fodor, Bever, and Garrett 1974; Jaeger 1984; Linell 1979; Ohala 1974a; Ohala 1974b; Ohala 1974c; Trammell 1978) yielded uniformly disappointing results. Of course, one could argue that the fact that no one has ever seen the top quark should not prevent us from constructing theories that rely on the reality of this subatomic particle. But analogies of this type are misleading. In fact, carefully controlled experiments with huge collectors of heavy water sunk deep in caves have provided solid tangible evidence for the reality of even this most elusive of physical entities. No such solid evidence has ever been provided for either linguistic rules or linguistic symbols.

Given these doubts and empirical failures, it made sense for researchers to begin to explore alternatives to symbols and rules. In the late 80s, work in connectionist modeling (Rumelhart and McClelland 1986a) began to challenge the necessity of linguistic rules and categories, focusing attention instead on models based on simple, observable cues

and connections between these cues. These new models correct a fundamental, fatal flaw inherent to symbolic models: the problem of excessive descriptive power.

The great power of AI systems derives from the computational architecture of the Von Neumann serial computer and the application of this architecture to human cognition by Simon, Newell, and their followers (Klahr and Kotovsky 1991; Newell and Simon 1972). This architecture provided unlimited symbol passing, full generativity, and unlimited scalability based on the system of data paths, memory addresses, and processing cycles that could be formalized in the logic of production systems. A modeler could take a few symbols, concatenate them into rules and, magically, the computer could conjure up a working model of mental processing. These models were at the same time both too powerful and too weak. They were too powerful in that they allowed one to model the learning of things that could never in reality be learned. At the same time, they were too weak in that they failed to generalize properly across language types and patterns. Moreover, attempts to identify a uniquely correct model without adding further constraints were shown to be impossible in principle (Anderson 1978). Neural nets (Grossberg 1987; Hopfield 1982; Kohonen 1982) limit this descriptive power by imposing two stringent limitations on computational models: a prohibition against **symbol passing** and an insistence on **self-organization**.

Neural networks require that the computations involved in the models echo the connectionist architecture of the brain. The basic constraint involved here is the prohibition against symbol passing. Neuroscience has shown that the brain cannot use memory addresses to bind variables; there is no neural mechanism that can assign an absolute “address” to a particular neuron (Squire 1987). Neurons do not send Morse code, symbols do not run down synapses, and brain waves do not pass phrase structures. Unlike the computer, the brain has no general scheme for register assignment, data pathing, or memory addressing. Moreover, the individual components of the neural system do not have the reliability of the electrical components of a standard digital computer (von Neumann 1956). In general, the brain provides no obvious support for the symbol passing architecture that provides the power underlying the von Neumann machine. Instead, computation in the brain appears to rely ultimately on the formation of redundant connections between individual neurons.

By itself, the requirement that computation be performed locally without symbol passing or homunculi is not enough to fully constrain descriptive power. One could still hand-wire a connectionist system to perform a specific function or to model a particular behavior. By detailed weight setting and the use of gating and polling neurons, virtually any function can be wired into a neural net (Hertz, Krogh and Palmer 1991). An early example of a fully hand-wired connectionist architecture was Lamb’s Stratificational Grammar (Lamb 1966). More recently, we have seen hand-wired connectionist models in areas such as speech errors (Dell 1986; MacWhinney and Anderson 1986; Stemberger 1985), ambiguity resolution (Cottrell 1985), and lexical activation (Marslen-Wilson 1987). The “implementational” approach to hand-wiring spares the modeler the tedium of hand-wiring by running the wiring procedure off symbolic templates. For example, Touretzky (1990) has shown that there are techniques for bottling the full power of a LISP-based production system architecture into a neural net. These demonstrations are important because they show how difficult it is to control excessive modelling power. However, they tell us little about how language is implemented in the brain.

In order fully to constrain descriptive power, modelers must match the constraint against symbol passing with the requirement that networks be **self-organizing**. This is to say, that models cannot be hand-wired and the connections between units must be developed on the basis of automatic learning procedures. It is this property of neural nets that makes them particularly interesting to the developmental psychologist and which also poses the greatest challenge to detailed modeling work. When the prohibition against symbol passing is combined with the demand for self-organization, the class of potential models of language learning becomes extremely limited. In fact, there is currently no detailed model of language acquisition that can satisfy these two criteria. Is this evidence that the criteria are too strict? I think not. Rather it is evidence that we can use these criteria to constrain our search for a truly plausible model of language acquisition. More importantly, it appears that those models which come closest to satisfying these criteria are also the same models that display further interesting and important properties, such as category leakage (McClelland and Kawamoto 1986), graceful degradation (Harley and MacAndrew 1992; Hinton and Shallice 1991; Marchman 1992), and property emergence (MacWhinney, Leinbach, Taraban, and McDonald 1989).

When these twin constraints are taken seriously, along with the standard conditions that must be imposed on any formal model (MacWhinney 1978b), building successful models becomes a tough job. When we add a third constraint—the need to demonstrate scalability—building powerful connectionist models becomes a nearly impossible task. Some modelers try to make headway against these odds by ignoring the scalability constraint and confronting only the first two constraints. This leads them to building models of very small pieces of the language acquisition puzzle. For example, some networks are constrained to well-defined topics such as the acquisition of the English past tense (Cottrell and Plunkett 1991) or German gender (MacWhinney et al. 1989). Other models have focused on small slices across larger problems such as question answering (St. John 1992) or word recognition (McClelland and Elman 1986). Some of these toy models may use only a few dozen sentences or a few dozen words. When one attempts to add additional words or sentences to these models, their performance often begins to degenerate. These problems with inadequate scalability are particularly serious in the study of language acquisition, since the move from a vocabulary of 500 words to a vocabulary of 700 words is a smooth accretional transition for the language-learning child. If connectionist models are to provide serious alternatives to symbolic models, it is crucial that they directly address each of these three issues: scalability, symbol passing, and self-organization. Any attempt to ignore one of these constraints detracts from the impact of the connectionist enterprise.

2. Grand Pretensions, Modest Reality

Like the symbolic paradigm before it, the connectionist paradigm seeks to provide a general model of human cognition. Because it has staked out such a wide territory, connectionism is committed to providing an account of all of the core issues in language acquisition, including grammatical development, lexical learning, phonological development, second language learning, and the processing of language by the brain. Despite these grand pretensions, the reality of connectionist modeling is more sober and

modest. In fact, much of the work to date has focused on the learning of narrow aspects of inflectional morphology in languages like English and German. While limited, work in this area has taught us a great deal. This paper sketches out the achievements of connectionist models in this well-researched area and then examines how we can move from these preliminary achievements to a fuller, more explanatory, unified approach to all of the core issues facing language acquisition theory.

Let us begin by reviewing some recent connectionist models of the learning of inflectional morphology. The first study of this topic was a model of English past tense marking presented by Rumelhart and McClelland (1987; 1986b). A more fully elaborated version of this model was developed by MacWhinney and Leinbach (1991). The task of these models was to convert the stem of an English verb into another inflected form, such as the past tense. For example, given a stem such as *eat*, the model could produce *eats*, *eating*, *ate*, or *eaten*.

Like all connectionist models, this model based its performance on the development of the weights on the connections between a large collection of “units.” The pattern of inputs and the connections between units was designed to implement the pattern of an autosegmental grid that has been developed in phonological theory (Goldsmith 1976; Nespor and Vogel 1986). The idea is that each vowel or consonant sound is a bundle of features that sits inside a slot within the framework or grid of the syllable. Words, in turn, are formed from combinations of syllables in a metrical grid.

The MacWhinney-Leinbach model used 12 consonantal slots and 6 vowel slots and allowed for words of up to three syllables. The segments of the stem were filled into this grid in right-justified fashion (MacWhinney 1993), as in this example for the word *bet* :

CCC VV CCC VV CC_b V_E CC_t

A further syllable was reserved for the suffix in the output. Each of the slots was in turn composed of a group of feature units. Since each of these feature units was bound to its particular slot, we can think of each unit as a slot/ feature unit. For example, the first filled consonantal slot in the representation for *bet* had active units for the labial, consonantal, and voiced features required for the sound /b/. Each of the consonantal slots had ten units and each of the vowel slots had eight units, for a total of 168 sound units per word. The architecture of the network had the shape shown in Figure 1.

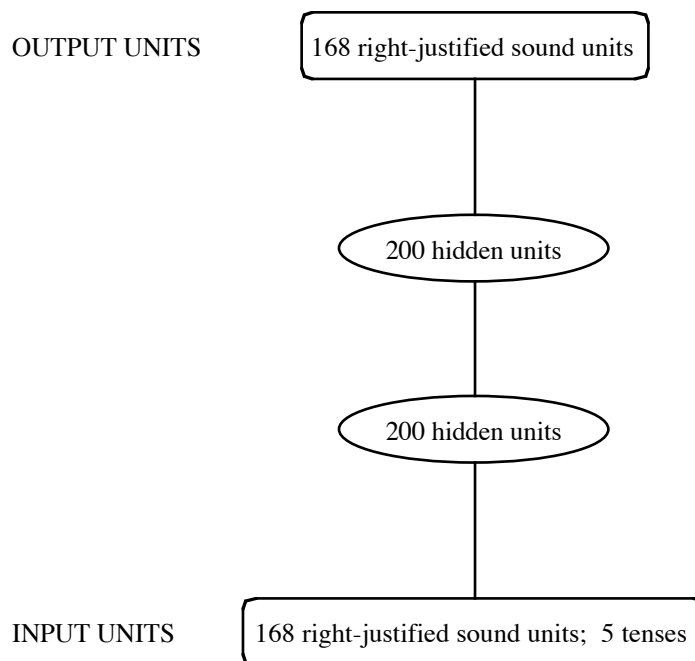


Figure 1. Architecture of the MacWhinney-Leinbach Past-Tense Network

The complete training corpus used 6949 different verb forms, derived from the 2161 highest frequency verbs in English (Francis and Kuc&era 1982). Of these 2161 verbs, 118 were irregulars and 2043 were regulars. The frequency with which a given form was included in the training epochs was determined by its frequency in the Francis and Kuc&era (1982) word frequency list. The highest frequency verbs were included most often. Learning in the model was controlled by the back-propagation algorithm (Rumelhart, Hinton and Williams 1986).

The network did an excellent job learning its input corpus, producing the correct output forms for 97% of the forms. At the end of 24,000 epochs of training, the only forms that it was still missing were low-frequency irregulars such as *bled* or *underwent*. Generalization testing showed that most new verbs were produced in the regular past, but that a few forms were treated as irregulars. Additional generalization testing is reported in MacWhinney (1993) and Ling and Marinov (1993).

English is a relatively poor language, at least in regard to its system of inflectional morphology. It has virtually no marking of case or gender. Nouns have only a single basic suffix for plurality and virtually the same suffix for the possessive. Although there are a few irregular past tense verbs, even the system of verbal morphology is fairly simple. Fortunately, we do not have to look far afield for a more challenging problem. Even a closely related language like German presents us with a far richer system of inflectional morphology. So rich, indeed, that Mark Twain (1935[1880]) once complained that:

a person who has not studied German can form no idea of what a perplexing language it is.... Every noun has a gender, and there is no sense or system in the distribution;

so the gender of each must be learned separately and by heart. There is no other way. To do this, one has to have a memory like a memorandum book. In German, a young lady has no sex, while a turnip has. Think what overwrought reverence that shows for the turnip, and what callous disrespect for the girl.

Any English speaker who has studied German, be it in the context of the classroom or in the country itself, has probably reached a very similar conclusion.

The vagaries of German gender are compounded by the fact that written German still clings to a system of case-marking only slightly simpler than that found in Classical Latin. For example, the definite article is declined through all four cases and all three genders in the singular and across all four cases with gender neutralized in the plural. The result of these various obligatory markings is the paradigm for the definite article in Table 1.

	Masc.	Fem.	Neut.	Plural
Nom.	der	die	das	die
Gen.	des	der	des	der
Dat.	dem	der	dem	der
Acc.	den	die	das	die

Table 1. The German Definite Article Paradigm

This paradigm is rife with homonymy. Only six distinct forms of the definite article occur in the 16 cells in the paradigm. This means that many of the forms cover several meanings. For example, the article *der* can mean either ‘masculine singular nominative,’ ‘feminine singular genitive,’ ‘feminine singular dative,’ or ‘plural genitive.’

In order to select the correct form of the definite article, the language learner has to know three things about the noun—its case, its number, and its gender. Number is the easiest category, since it bears a fairly straightforward relation to real-world properties. Case is somewhat more abstract, but it can generally be figured out through a combination of cues from the verb, related prepositions, and some word order patterns. However, there is little in the external situation that can help the child figure out the gender of a noun (Maratsos and Chalkley 1980). It is possible that the noun’s gender could be simply memorized or even inferred on the basis of its use within the paradigm. For example, knowing that the novel noun *Gann* is treated as *den Gann* in the accusative, the child can infer that it takes the form *der Gann* in the nominative. However, recent work by Köpcke and Zubin (Köpcke 1982; Köpcke 1988; Köpcke and Zubin 1983; Köpcke and Zubin 1996; Köpcke and Zubin 1984; Zubin and Köpcke 1981; Zubin and Köpcke 1986) has shown that Mark Twain’s view of gender as arbitrary and unpredictable is incomplete and partially incorrect.

In fact, Köpcke and Zubin have shown that there are dozens of phonological cues that can be used to predict the gender of a German noun. For example, almost all nouns ending in *-e* are feminine, as in *die Sonne*, *die Ente*, and *die Tante*. Almost all nouns beginning with *dr-*, *tr-*, and *kn-* are masculine, as in *der Knecht*, *der Trieb*, and *der*

Drang. There are dozens of other cues like these. In addition to these purely phonological cues, there are derivational endings such as *-chen*, *-lein*, *-ett*, *-tum*, *-ei*, and so on, each of which reliably specifies a particular gender.

MacWhinney, Leinbach, Taraban, and McDonald (1989) constructed a series of models of the acquisition of German gender. The first model dedicated a series of nodes to the cues enumerated by Köpcke and Zubin, along with a series of nodes for case and number cues. The second model made no explicit coding of the Köpcke-Zubin cues, instead simply encoding the phonological form of the base in the manner of the MacWhinney-Leinbach model for English. Much to our surprise, the network with no hand-coding of features outperformed the hand-crafted network in terms of both learning and generalization. These results provide nice support for the view of connectionist networks as providing emergent self-organizing characterizations of linguistic systems. Similar results for hand-wired vs. emergent solutions are reported by Daelemans, Gillis, and Durieux (1994) for the learning of Dutch stress by a connectionist network.

The architecture of the successful non-handcrafted German simulation had the form in Figure 2. The input to the network was a pattern across 143 phonological units representing the noun stem and 11 phonological units representing suffixes attached to the noun. In addition, there were 5 semantic units representing inherent gender and 17 cues that provided a distributed pattern of surface structure information helpful in determining the case for the noun. However, the actual identity of the case was not given. This network was trained with 2000 German nouns from all cells in the paradigm. It learned the training set completely. When tested with 200 new nouns, the system was able to guess the gender of the new words with 70% accuracy. This compares with a level of 80% accuracy that could be expected from a native German speaker (Köpcke 1982).

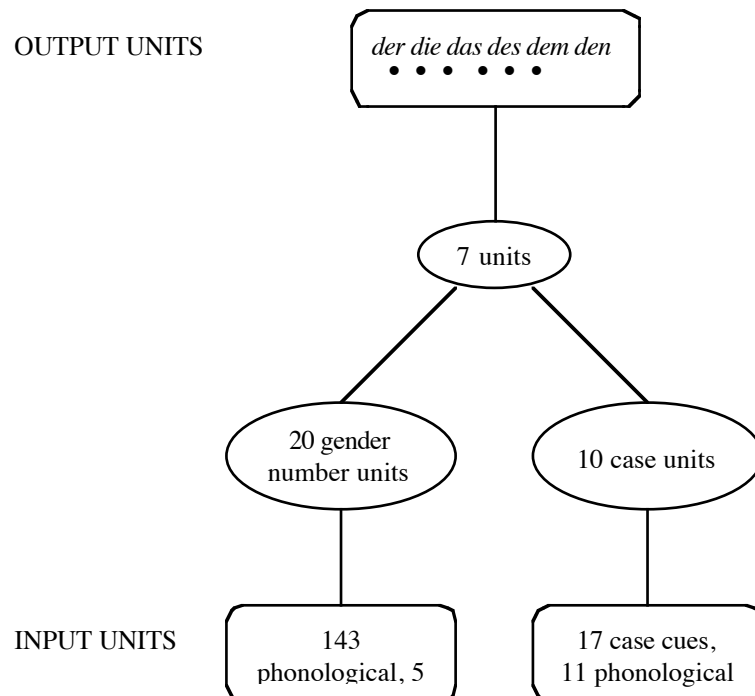


Figure 2. Network Architecture for German Stress

The model also succeeded in capturing a variety of important developmental phenomena. Like the children studied by MacWhinney (1978a) and Mills (1986), the model showed early acquisition of the nominative and delayed acquisition of the genitive. These acquisitional order effects are undoubtedly due to the fact that the frequencies of the four cases in the training corpus were based on their actual distribution in German corpora. Also, like German children, the model made good use of reliable cues to gender such as final *-e* or some of the derivational markers. Like children, the model was able to use the paradigm to infer word class. For example, given the accusative form *den Bauer*, the model could produce the genitive singular form *des Bauers*. Native speakers can do this on the basis of only one exposure to the word and the model displays similar behavior. Like children, the model frequently omitted the article. This occurred when the output units did not reach threshold. Finally, the model demonstrated the same tendency toward overgeneralization of the feminine gender often found in children. This is apparently due to the fact that the similarity of the feminine to the plural lends it enough frequency and paradigmatic support to tend to overcome the effects of the other two genders.

When evaluating the success of these connectionist models of language acquisition, it is important to consider the extent to which symbolic models are able to address similar problems. For the learning of English verb morphology, Ling (1994) presents a model that performs about as well as the MacWhinney-Leinbach model for English. Although Ling's model is based on a conventional symbol passing architecture, it uses an input-driven induction algorithm, thereby avoiding problems with hand-wiring. However, the detailed feature combinations constructed by Ling's pattern associator provide no clear representation of rules and would probably not be accepted as a full symbolic model by many linguists and psycholinguists. Nonetheless, the head-on comparison of models conducted by Ling is quite interesting and further connectionist work by Bullinaria (1997) shows that the competitive testing of symbolic and connectionist models can be quite instructive.

3. Lexical Items: An Achilles Heel?

Despite their basic successes, there are several properties of the MacWhinney models that should give us serious cause for worry. In this sense, these weaknesses are actually more instructive than their successes. These weaknesses play a parallel role in both the model for English and the model for German.

3.1 Problem 1: Homophony

Because these models convert from phonology to phonology without using discrete representations for lexical items, they both run into serious problems with homophonous

forms. Consider what happens to the three homophones of the word *ring* in English. We can say *the maid wrung out the clothes*, *the soldiers ringed the city*, or *the choirboy rang the bell*. These three different words all have the same sound /rɪŋ/ in the present, but each takes a different form in the past.

A similar problem arises in German. The stem *Bund* can be either *der Bund* or *das Bund*, depending on whether it is an ‘alliance’ or a ‘bundle or sheaf of wheat.’ And the stem *Band* can be either *der Band* or *das Band* depending on whether it means a ‘volume of a book’ or a ‘rubber band.’ The problem here is that, in order to control this variation, one needs to distinguish the meanings of the two homophonous lexical items involved. If the network has no concept of “lexical item” this is difficult to do. These problems also affect the formation of the plural. For example, the singular form *das Wort* has two plural forms *die Wörter* (several isolated words) and *die Worte* (connected words or speech).

3.2 Problem 2: Compounds

A parallel problem crops up in the formation of the past tense of compound words. The English training set included several compounds based on irregular verbs such as *undergo*, *rethink*, and *undo*. The fact that the past tense of *undergo* is *underwent* depends on the fact that *undergo* is a variant of the stem *go*. When the compound itself is high enough in frequency, the network can learn to treat it as an irregular. However, the network had a hard time learning the past tense of low frequency irregular compounds. At the end of training, the model was still not producing *underwent* correctly, even though it had learned *went* early in training. It is clear that the model was not able to use its learning about *go ~ went* to facilitate learning of the far less frequent form *undergo ~ underwent*.

A similar problem emerged in the learning of the gender of compounds in German. The model quickly learned that *Mutter* (‘mother’) was feminine, because the noun was so frequent. However, there is a competing tendency to treat words with final *-er* as masculine. And this tendency led the model to treat the less frequent form *Grossmutter* (‘grandmother’) as masculine, although it is clearly a variant of *Mutter* and should be feminine.

3.3 Problem 3: Derivational Status

The model was also not capable of utilizing information regarding the derivational status of lexical items. As Kim, Pinker, Prince, and Prasada (1990) have noted, the past tense forms of denominal verbs are uniformly regular. For example, the word *ring* can be used as a verb in a sentence such as *the groom ringed her finger* and we would never say *the groom rung her finger*. However, as we noted earlier, the network of the MacWhinney-Leinbach simulation cannot use the derivational status of the verb *ring* to make this distinction.

German provides even clearer examples of the importance of derivational status. All German nouns that derive from verbs are masculine. For example, the noun *der Schlag* (‘blow’; ‘cream’) derives from the verb *schlagen* (‘to hit’). However, there is no motivated way of indicating this in the model. In general, the model includes no

independent way of representing morphological relationships between words. Thus, no distinction is made between true phonological cues such as final /e/ or initial /kn/ and derivational markers such as *-chen* or *-ett*. This leads to some very obvious confusions. For example, masculines such as *der Nacken* ('neck') and *der Hafen* ('harbor') end in phonological /en/, whereas neuters such as *das Wissen* ('knowledge') and *das Lernen* ('learning') end in the derivational suffix *-en*. Confusion of these two suffixes leads to inability to correctly predict gender for new nouns ending in /en/.

3.4 Problem 4: Early Irregulars

A well-known child language phenomenon is the u-shaped learning curve for irregular verbs in English. For a verb such as *go*, children may begin with *went*, then show some occasional usage of *goed*, and finally settle in on correct usage with *went*. During the period of oscillation between *goed* and *went*, it is usually *went* that predominates. However, not all irregular verbs show this pattern and not all overregularizations enter at the same time. The MacWhinney-Leinbach model showed the oscillation between *goed* and *went* terminating in correct usage, but it did not show early use of *went*. The reason for the failure of the model to produce early *went* is that the network is configured to construct the past tense as a variation on the phonological form of the present tense. A more accurate model would allow direct learning of *went* as a rote form. But the capacity to learn rote associations between sound and meaning involves the capacity to learn lexical items and this means that we will need a connectionist architecture specifically designed for this type of learning.

3.5 The Core Problem

These four weaknesses we have discussed can be linked to a single core problem: the absence of any way of representing lexical items. Because these models have no lexical items, they are forced to rely on sound features as the only way to determine inflectional morphology. Of course, it would be a mistake to imagine that the sound form of words has no impact on inflection and derivation. In fact, it seems that what really happens during both production and comprehension is that both the sound and meaning of stems and affixes are available in parallel, although the time course of their activation may vary (Kawamoto 1993).

One way of addressing this problem is to mix both sound features and meaning features without providing any explicit representation of lexical items. Attempts to achieve lexical access without lexical representations have been partially effective in models of reading (Kawamoto 1993; Plaut and McClelland 1993) and spelling (Seidenberg and McClelland 1989). Models of reading and spelling can avoid lexical representations, because orthographic-phonological correspondences typically make little reference to lexical items. However, these models run into more serious problems (Cottrell and Plunkett 1991; Hoeffner 1992), when dealing with language learning and word production. Models of the Hoeffner type display this problem most clearly. They learn to associate sound to meaning and store these associations in a distributed pattern in the hidden units. This approach works well enough until the model is given more than

about 700 forms. At this point, the large pool of hidden units is so fully invested in distinguishing phonological and semantic subtypes and their associations that there is simply no room for new words. Adding more hidden units doesn't solve this problem, since all the interconnections must be computed and eventually the learning algorithm will bog down. It would appear that what we are seeing here is the soft underbelly of connectionism—its inability to represent Islands of Stability in the middle of a Sea of Chaos. Perhaps the problem of learning to represent lexical items is the Achilles' heel of connectionism.

4. A Solution to the Lexical Learning Problem

4.1 Lexicalist Connectionism

Given the seriousness of these problems and the extent to which they have limited the full effectiveness of connectionist models for English and German, we decided to explore alternatives to fully distributed representations. The core assumption in our new approach is that the lexical item serves a central controlling and stabilizing role in language learning and processing. We can refer to this revised approach as **lexicalist connectionism**. Predecessors to lexicalist connectionist models can be found in localist connectionist models of the type developed by Dell (1986) and Stemberger (1985), where a central role is given to the lexical item. However, because of their localist node-based architecture, these models were forced to rely on hand-wiring.

In order to model lexical learning without hand-wiring, we turned to the self-organizing feature map (SOFM) framework of Kohonen (1982) and Miikkulainen (1990; Mikkulainen and Dyer 1991). In this framework, word learning is viewed as the association of a large number of phonological features to a large number of semantic features. These many features constitute a high-dimensional space. However, the association of these many dimensions can be compressed onto a 2-D feature map in which nearby vectors in the input space are mapped onto nearby units in the 2-D map. The two dimensions of the visible representation do not have any direct relation to features in the input data set; rather they preserve the topological relations inherent in the high-dimensional space.

Schematically, one can think of the map as a 2-D compression that associates sound and meaning in the way suggested by Figure 3.

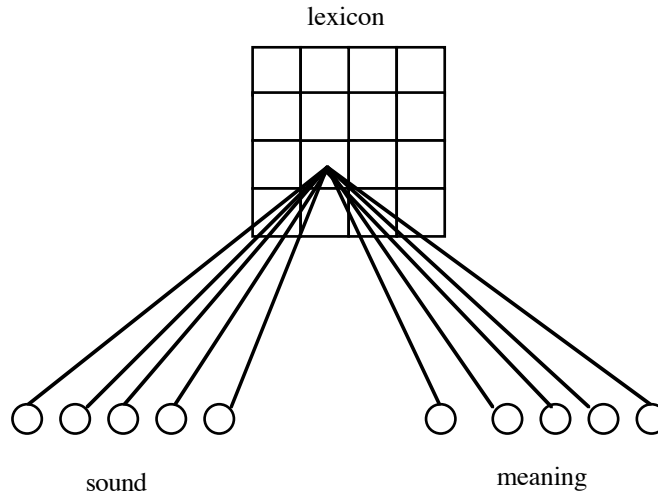


Figure 3. Mapping Lexical Items as Sound-Meaning Associations

Learning involves the strengthening of weights between particular inputs and units on the map. This can be done in strict accord with established biological principles of lateral inhibition and the redistribution of syntactic resources (Kohonen 1982) using a computationally efficient algorithm that is faithful to these biological principles (Miikkulainen 1990).

Using this algorithm, we found that a network with 10,000 nodes can learn up to 6000 lexical associations with an error rate of less than 1%. In this implementation, we used four floating-point numbers to represent sound and four additional floating-point numbers to represent meaning. The shape of these eight numbers for each item was generated randomly. At the beginning of learning, the first input vector of eight numbers would lead by chance to somewhat stronger activation on one of the 10,000 cells. This one slightly more active cell would then inhibit the activation of its competitors. Once it has won this particular competition, its activation would be negatively damped to prevent it from winning for all of the items. Then, on the next trial, another cell would win in the competition for the next lexical item. This process would repeat until all 6000 items had developed some “specialist” cell in the feature map. During this process, the dynamics of self-organization would ensure that items that shared features would end up in similar regions of the feature map.

One way of following the development of the feature map is to track the average radius of the individual items. After learning the first 700 words, the average radius of each word was 70 cells; after 3000 words, the radius was 8; after 5000 words the radius was 3; and after 6000 words the radius was only 1.5 cells. Clearly, there is not much room for new lexical items in a feature map with 10,000 cells that has already learned 6000 items. However, there is good reason to think that the enormous number of cells in the human brain means that the size of the initial feature map is not an important limiting constraint on the learning of the lexicon by real children.

4.2 Using Maps for Retrieval

In order to permit full use of the lexicon in processing, we next proceed to supplement the basic SOFM architecture with additional connections. We have done this by simultaneously building up three identical parallel feature maps: the sound map, the meaning map, and the association map. Figure 4 illustrates the relations of these three maps. In this figure, a particular location on the association map relates a location on the sound map to a location on the meaning map.

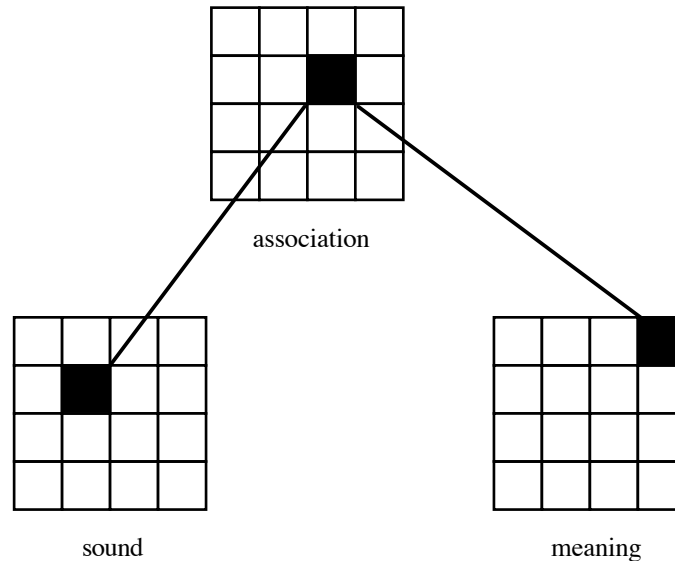


Figure 4. The Development of Bidirectional Sound-Meaning Associations

During training, the sound and meaning maps are given full input patterns. Each then comes to take on its own pattern of internal organization. Although these patterns emerge as quite similar, they are not constrained externally to be identical. Once patterns are established on the two basic maps, we start a period of Hebbian learning. When a particular sound and a particular meaning are active on the two lower maps, we then reinforce the most highly active unit on the association map. This unit then becomes the effective center of the lexical item. Once this initial association is formed, we can use this device as an auto-associator. We can then turn on a sound and retrieve the corresponding meaning by activating the correct associator unit. This models lexical comprehension. We can also turn on a meaning and retrieve the corresponding sound. This models lexical production.

5. Using Lexical Representations

This implementation allows us to put aside our earlier worries regarding lexical learning as an Achilles' Heel for connectionist models of language learning. We now have structures that function like lexical items and which developed in a fully self-organizing way without external intervention or corrective training. We cannot manipulate these items with standard symbol passing techniques, but additional input can develop their connections to other processes. In this section, I will discuss some of ways in which we can use these representations to make further progress in connectionist models of language acquisition.

5.1 Learning Inflectional Morphology

The first crucial application of this new modeling effort has been to the acquisition of inflectional morphology. Our initial goal was to see if the model could learn to inflect verbs for the English past tense. The network is first given the set of input forms used in the MacWhinney-Leinbach simulations for English. Each phonological input is paired with a randomly generated, but consistent, semantic representation. Included in the semantic representation are features that consistently represent the meanings signaled by English verb inflections: present, past, perfect, progressive, and third singular. During this lexical training, the network learns inflected items such as *went* and *gone*, as well as regularly inflected items such as *goes*, *going*, *jumped*, and *runs*. The network also acquires a large number of bare-stem verbs such as *go*, *run*, and *jump*.

Before this training is completed, we introduce a few generalization trials. These trials take verbs that have been learned as bare-stems, but not yet as inflected forms. For example, the network already knows *jump*, but has not yet learned *jumped*. At this point, the network engages in the process of **masking** (Burgess 1995; Burgess and Hitch 1992; Carpenter, Grossberg and Reynolds 1991; Cohen and Grossberg 1987; Grossberg 1987) which leads to the temporary suppression of *jump* and the isolation of *-ed* as the not-yet-recognized form. Masking is a general process that is used to suppress the status of any phonological string that has been recognized. In this particular case, masking leads to the isolation of the part that the child already knows from the residue. In other words, isolation is a result of the operation of the process of masking. The suffix is then learned as a new lexical item that associates *-ed* with the meaning of the past tense. In the terms of MacWhinney (1978a), learning involves associating the "unexpressed" meaning of the past tense with the "uncomprehended" sound string *-ed*. Learning of the other inflectional suffixes proceeds in a similar way. When the system is asked to produce the past tense of a new verb, masking is used in reverse. For example, the past tense of *jump* is produced by (1) activating *jump*, (2) masking the semantics and phonology of *jump*, (3) reactivating the network with 'past,' and (4) retrieving *-ed*. The output phonology of *jump* and the *-ed* suffix is then available for read-out from the articulatory buffer and the child can produce *jumped*.

This approach to the problem of learning the English past tense solves a number of problems faced by earlier non-lexical models. First, the model succeeds in capturing both rote lexicalization and combinatorial lexicalization within a single connectionist model. Rote forms are picked up directly on the feature map. Combinatorial forms are created by the isolation of the suffix through masking and the use of masking in production. Second, the model no longer faces the earlier problems that stemmed from a lack of lexical items.

Homophony is not a problem, because the meanings of the homophones of /rêN/ are now representationally distinct. The model can locate *go* inside *undergo* and *Mutter* inside *Grossmutter* because it now has lexical items and can use masking. In German, derivational suffixes like *-chen* can be used as cues to gender because these suffixes now have their own representational status. Finally, the model no longer has any problem with the early acquisition of irregulars such as *went*. Since the learning is grounded now on lexical items, these high-frequency forms are some of the first forms learned.

This model relies on three crucial processing mechanisms. The first mechanism is the self-organizing competitive learning incorporated in the feature map. The second mechanism uses Hebbian learning to develop a central associative map for lexical comprehension and production. The third mechanism is the masking process which works to extract inflections.

5.2 The Logical Problem

The ability to produce *went* by rote and **goed* by combination within the same lexical feature map also allows us to begin work on a general solution to the so-called “logical problem of language acquisition” (Baker and McCarthy 1981; Gleitman 1990; Gleitman, Newport and Gleitman 1984; Morgan and Travis 1989; Pinker 1984; Pinker 1989; Wexler and Culicover 1980). 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 combinational *-ed* form. Figure 5 illustrates this competition:

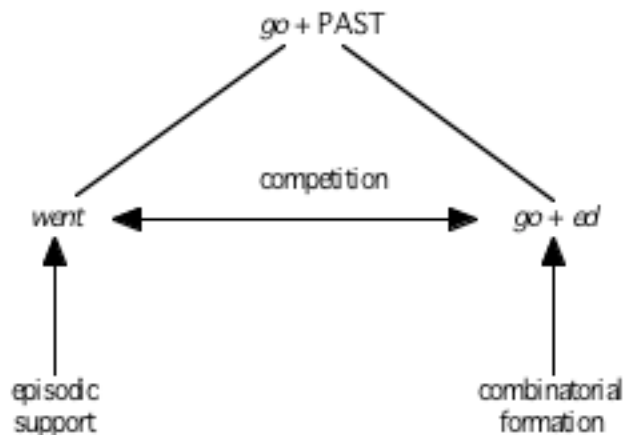


Figure 5. Competition between Episodic and Combinatorial Knowledge

This particular competition is an example of what Baker and McCarthy (1981) call a “benign exception to the logical problem.” It is benign because everyone will agree that there is only one way to say ‘goed’ and that is *went*. However, in other cases, people will argue that there is no obvious way to “correct” a given error by using an obviously

Daddy | unknown | Mommy

For the first and last phonological stretches, there are lexical items that match. These strings and the semantics they represent are masked. The unknown stretch is not masked and therefore stimulates lexical learning of the new word *likes*. The core of the learning for *likes* is the association of the sound /laɪk/ to the meaning ‘like.’ In addition to this association, the central association feature map now constructs links not only to sound and meaning, but also to argument relations. The initial argument frame for the word *likes* is [[arg1, preposed, *Daddy*][arg2, postposed, *Mommy*]]. Further exposures to sentences such as *Daddy likes pancakes* or *Billy likes turtles* will soon generalize the dependency frame for *likes* to [[arg1, preposed, HUMAN][arg2, postposed, OBJECT]].

The implementation of the acquisition of argument frames follows the logic developed by Gupta and MacWhinney (1992) for the acquisition of German declensional marking. That model used a SOFM for the extraction of cooccurrence patterns between articles and nouns. Using these patterns, the full shape of the German declensional pattern emerged inside the SOFM. Nodes in the map took on the role of associating particular constellations of case and number marking on the article with one of the three grammatical genders of German. This system was then linked to a back-propagation system that responded to additional phonological cues to gender. The general shape of this type of model is given in Figure 6 which includes the basic components of Figure 4 along with additional maps for argument structures and articulatory control.

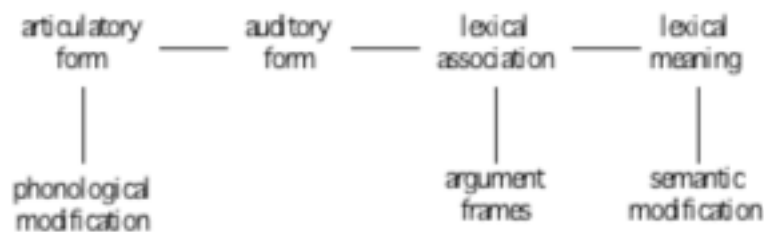


Figure 6. A Connectionist Account of the Shape of Linguistic Representations

The argument frame feature map is intended to capture a few basic grammatical categories which will then be related to particular lexical items. However, these argument frames will also activate patterns in the meaning and auditory form feature maps. The feature map is designed to include two basic effects. One is the activation of the correct argument frame for a specific lexical item. The other is the activation of argument frames for semantically related groups of words or lexical “gangs.” Words that have similar meanings will tend to activate similar argument structures.

Lexical gang effects help us address some remaining aspects of the “logical problem of language acquisition.” Bowerman (1988) cites overgeneralization errors such as **I poured the tub with water*, **I unsqueezed the ball*, and **I recommended him a soup* as potential evidence for the logical problem. However, recovery from these errors can be viewed as similar to recovery from errors such as **goed*. Figure 7 illustrates the situation.

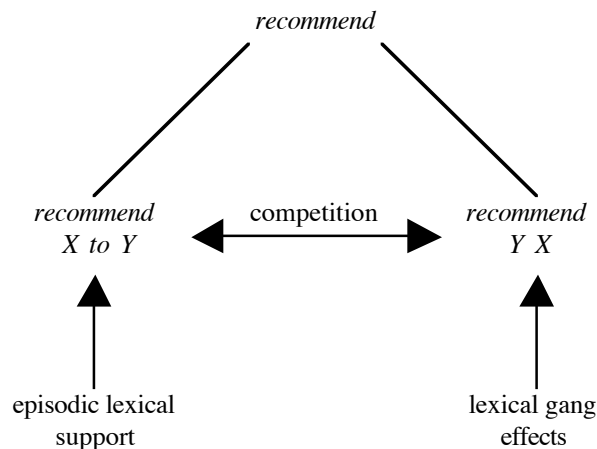


Figure 7. Competition between Lexical Argument Frames and Lexical Gang Effects

The Competition Model (MacWhinney 1988) treated the argument structures for inflectional morphemes as similar to the structures for verbs and prepositions. For the English past tense, the masking and extraction of the phonological form /t/ from the verb *jumped* produces a frame specifically linked to the sound and meaning of *jump*. Over time, the meaning of [arg1] of the past tense morpheme becomes generalized to any verb and the sound becomes generalized so that the /t/ and /d/ forms of the morpheme require stem-final non-dentals (voiceless and voiced, respectively) and vowels, whereas the /Id/ form requires a final dental consonant.

5.5 Modification Systems

In addition to the basic maps for lexical associations and argument frames, Figure 6 has systems for phonological and semantic modification. Phonological modification operates to enforce general phonological patterns when words are linked together. Semantic modification works to adapt meanings when words are linked together. Often this linkage is predictable and obvious. However, sometimes words are placed together even at the expense of standard argument frames. Because connectionist systems are constraint-satisfaction systems, rather than rule systems, they can deal with partial violations in the combinations of words. Consider a combination like *another sand*. Typically, the word *another* requires a count noun and *sand* is a mass noun. However, when the listener is confronted with this particular combination, it is still possible to retrieve an interpretation by treating *sand* as a count noun. This can be done by thinking of bags of sand, types of sand, alternative meanings of the word *sand*, or even the act of applying sandpaper to something. MacWhinney (1989) talks about these semantic extension effects in terms of a process of “pushy polysemy.”

5.6 Sentence Compilation

Currently, our actual implementation of argument frame effects has focused on the use of a simple recurrent network, rather than feature maps. The task we have selected is the acquisition of English and Dutch as both first and second languages. Figure 8 illustrates the shape of the model we have used:

Figure 8. *Network Architecture for Acquisition of English and Dutch*

The input to the model is a corpus of sentences from the language illustrating the major syntactic frames, including relative clauses, imperatives, and various word order types. For each word, the input cues include animacy, number, part-of-speech, and case-marking for pronouns.

This model learns the input corpus in a few runs. We then test it using the experimental agrammatical sentences used in Competition Model studies by McDonald (1987; 1989) and Kilborn and Cooreman (1987) for Dutch, and Bates, McNew, MacWhinney, Devescovi, and Smith (1982) for English. Overall, the match of the model to the experimental data was excellent.

6. Limitations and Prospects

This modeling work with connectionist nets has advanced to the point where it can compete on an equal footing with the more powerful rule-based symbolic models. These networks minimize hand-wiring and maximize self-organization. Because of this, they are attractive to developmental psychologists. They also avoid reliance on hardware address and symbol passing. And it is this that makes them attractive to cognitive neuroscientists. But they are still incomplete in many ways. In our modeling of language learning, we have made good progress in the areas of lexical learning, inflectional morphology, and role assignment. But there is still much work to be done. We need to link up the high-level modeling of role assignment to detailed aspects of lexical processing and local role assignment and attachment. We need to model the learning of more different types of inflectional structures. And we need to deal in greater detail with lexical effects on syntax. There may still be an Achilles' Heel that will doom this whole process to failure. But, for now, we can look at the progress we have made as grounds for cautious optimism.

References

- Anderson, John. 1978. Arguments concerning representations for mental imagery. *Psychological Review* 85, 249-277.
- Baddeley, Alan. 1986. *Working Memory*. Oxford: Oxford University Press.
- Baker, Charles L. and John J. McCarthy (eds.). 1981. *The Logical Problem of Language Acquisition*. Cambridge, MA: MIT Press.
- Bates, Elizabeth, Sandra McNew, Brian MacWhinney, Antonella Devescovi, and Stan Smith. 1982. Functional constraints on sentence processing: A cross-linguistic study. *Cognition* 11, 245-299.
- Bowerman, Melissa. 1988. The "no negative evidence" problem. In John Hawkins (ed.), *Explaining Language Universals*, 73-101. London: Blackwell.
- Bullinaria, John A. 1997. Modelling reading, spelling and past tense learning with artificial neural networks. *Brain and Language* 59, 236-66.
- Burgess, Neil. 1995. A solvable connectionist model of immediate recall of ordered lists. In Gerald Tesauro, David Touretzky, and Joshua Alsppector (eds.), *Neural Information Processing Systems 7*. San Mateo, CA: Morgan Kaufmann.
- Burgess, Neil and Graham Hitch. 1992. Toward a network model of the articulatory loop. *Journal of Memory and Language* 31, 429-460.
- Carpenter, Gail, Stephen Grossberg, and John Reynolds. 1991. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks* 4, 565-588.
- Cohen, Michael and Stephen Grossberg. 1987. Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics* 26, 1866-1891.
- Cottrell, Garrison. 1985. *A Connectionist Approach to Word Sense Disambiguation*. Rochester, NY: University of Rochester.
- Cottrell, Garrison and Kim Plunkett. 1991. Learning the past tense in a recurrent network: Acquiring the mapping from meaning to sounds. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, 328-33. Hillsdale, NJ: Lawrence Erlbaum.
- Daelemanns, Walter, Steven Gillis, and Gert Durieux. 1994. The acquisition of stress: A data-oriented approach. *Computational Linguistics* 20, 421-51.
- Dell, Gary. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283-321.
- Fodor, Jerry A., Thomas G. Bever, and Merrill F. Garrett. 1974. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.
- Francis, W. Nelson and Henry Kucera. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Gleitman, Lila. 1990. The structural sources of verb meanings. *Language Acquisition* 1(1), 3-55.
- Gleitman, Lila R., Elissa L. Newport, and Henry Gleitman. 1984. The current status of the motherese hypothesis. *Journal of Child Language* 11, 43-79.
- Goldsmith, John. 1976. An overview of autosegmental phonology. *Linguistic Analysis* 2, 23-68.

- Grossberg, Stephen. 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, Stephen. 1987. Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science* 11, 23-63.
- Gupta, Prahlad and Brian MacWhinney. 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*, 253-8. Hillsdale, NJ: Lawrence Erlbaum.
- Gupta, Prahlad and Brian MacWhinney. 1994. Is the articulatory loop articulatory or auditory? Re-examining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language* 33, 63-88.
- Gupta, Prahlad and Brian MacWhinney. 1997. Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language* 59, 267-333.
- Harley, Trevor, and Siobhian MacAndrew. 1992. Modelling paraphasias in normal and aphasic speech. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 378-83. Hillsdale, NJ: Lawrence Erlbaum.
- Hertz, John, Anders Krogh, and Richard Palmer. 1991. *Introduction to the Theory of Neural Computation*. New York: Addison-Wesley.
- Hinton, Geoffrey and Tim Shallice. 1991. Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review* 98(1), 74-95.
- Hoeffner, James. 1992. Are rules a thing of the past? The acquisition of verbal morphology by an attractor network. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 861-6. Hillsdale, NJ: Lawrence Erlbaum.
- Hopfield, John J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79, 2554-2558.
- Houghton, George. 1990. The problem of serial order: A neural network model of sequence learning and recall. In Robert Dale, Christopher Mellish, and Michael Zock (eds.), *Current Research in Natural Language Generation*, 287-319. London: Academic Press.
- Jaeger, Jeri J. 1984. Assessing the psychological status of the vowel shift rule. *Journal of Psycholinguistic Research* 13(1), 13-36.
- Kawamoto, Alan. 1993. Non-linear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language* 32, 474-516.
- Kilborn, Kerry and Ann Cooreman. 1987. Sentence interpretation strategies in adult Dutch-English bilinguals. *Applied Psycholinguistics* 8, 415-431.
- Kim, John, Steven Pinker, Alan Prince, and Sandeep Prasada. 1990. Why no mere mortal has ever flown out to center field. *Cognitive Science* 15, 173-218.
- Klahr, David and Kenneth Kotovsky. 1991. *Complex Information Processing: The Impact of Herbert A. Simon*. Hillsdale, NJ: Lawrence Erlbaum.
- Kohonen, Teuvo. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43, 59-69.
- Köpcke, Klaus-Michael. 1982. *Untersuchungen zum Genusssystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer.

- Köpcke, Klaus-Michael. 1988. Schemas in German plural formation. *Lingua* 74, 303-335.
- Köpcke, Klaus-Michael and David Zubin. 1983. Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für germanistische Linguistik* 11, 166-182.
- Köpcke, Klaus-Michael and David Zubin. 1996. Prinzipien für die Genuszuweisung im Deutschen. In Ewald Lang and Gisela Zifonun (eds.), *Deutsch - typologisch*, 473-91. Berlin: Walter de Gruyter.
- Köpcke, Klaus-Michael and David Zubin. 1984. Sechs Prinzipien für die Genuszuweisung im Deutschen: ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte* 93, 26-50.
- Lamb, Sydney. 1966. *Outline of Stratificational Grammar*. Washington: Georgetown University Press.
- Linell, Per. 1979. *Psychological Reality in Phonology: A Theoretical Study*. Cambridge: Cambridge University Press.
- Ling, Charles. 1994. Learning the past tense of English verbs: The symbolic pattern associator vs. connectionist models. *Journal of Artificial Intelligence Research* 1, 209-229.
- Ling, Charles and Marin Marinov. 1993. Answering the connectionist challenge. *Cognition* 49, 267-290.
- MacWhinney, Brian. 1978a. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43(1), 1-122.
- MacWhinney, Brian. 1978b. Conditions on acquisitional models. *Proceedings of the 1978 Annual Conference*, 421-427. New York: Association for Computing Machinery.
- MacWhinney, Brian. 1988. Competition and teachability. In Richard Schiefelbusch and Mabel Rice (eds.), *The Teachability of Language*, 63-104. New York: Cambridge University Press.
- MacWhinney, Brian. 1989. Competition and lexical categorization. In Roberta Corrigan, Fred Eckman, and Michael Noonan (eds.), *Linguistic Categorization*, 195-241. New York: Benjamins.
- MacWhinney, Brian. 1993. Connections and symbols: Closing the gap. *Cognition* 49, 291-296.
- MacWhinney, Brian. 1994. The dinosaurs and the ring. In Roberta Corrigan, Susan Lima, and Gregory Iverson (eds.), *The Reality of Linguistic Rules*, 283-320. Amsterdam: John Benjamins.
- MacWhinney, Brian and John Anderson. 1986. The acquisition of grammar. In Irwin Gopnik and Myrna Gopnik (eds.), *From Models to Modules*, 3-25. Norwood, N.J.: Ablex.
- MacWhinney, Brian and Jared Leinbach. 1991. Implementations are not conceptualizations: Revising the verb learning model. *Cognition* 29, 121-157.
- MacWhinney, Brian, Jared Leinbach, Roman Taraban, and Janet L. McDonald. 1989. Language learning: Cues or rules? *Journal of Memory and Language* 28, 255-277.
- Maratsos, Michael and Mary Chalkley. 1980. The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In Keith Nelson (ed.), *Children's Language*, Vol. 2, 127-214. New York: Gardner.

- Marchman, Virginia. 1992. Constraint on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience* 5, 215-234.
- Marslen-Wilson, William D. 1987. Functional parallelism in spoken word-recognition. *Cognition* 25, 71-102.
- McClelland, James and Jeffrey Elman. 1986. Interactive processes in speech perception: The TRACE model. In James McClelland and David Rumelhart (eds.), *Parallel Distributed Processing*, Vol. 2, 58-121. Cambridge, MA: MIT Press.
- McClelland, James and Alan Kawamoto. 1986. Mechanisms of sentence processing: Assigning roles to constituents. In James McClelland and David Rumelhart (eds.), *Parallel Distributed Processing*, Vol. 2, 272-326. Cambridge, MA: MIT Press.
- McDonald, Janet L. 1987. Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics* 8, 379-414.
- McDonald, Janet L. 1989. The acquisition of cue-category mappings. In Brian MacWhinney and Elizabeth Bates (eds.), *The Crosslinguistic Study of Language Processing*, 375-396. New York: Cambridge University Press.
- Miikkulainen, Risto. 1990. A distributed feature map model of the lexicon. *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 447-54. Hillsdale, NJ: Lawrence Erlbaum.
- Miikkulainen, Risto and Michael Dyer. 1991. Natural language processing with modular neural networks and distributed lexicon. *Cognitive Science* 15, 343-399.
- Mills, Anne E. 1986. *The Acquisition of Gender: A Study of English and German*. Berlin: Springer-Verlag.
- Morgan, James and Lisa Travis. 1989. Limits on negative information in language input. *Journal of Child Language* 16, 531-552.
- Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Newell, Allen and Herbert Simon. 1972. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Ohala, John. 1974a. Phonetic explanation in phonology. *Papers from the Parasession on Natural Phonology*, 257-274. Chicago: CLS.
- Ohala, John J. 1974b. Experimental historical phonology. In John M. Anderson and Charles Jones (eds.), *Historical Linguistics*, Vol. 2, 353-89. New York: American Elsevier.
- Ohala, Manjari. 1974c. The abstractness controversy: Experimental input from Hindi. *Language* 50, 225-235.
- Pinker, Steven. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- Pinker, Steven. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.
- Plaut, David and James McClelland. 1993. Generalization with componential attractors: Word and nonword reading in an attractor network. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, 824-9. Hillsdale, NJ: Lawrence Erlbaum.
- Rumelhart, David, Geoffrey Hinton, and Ronald Williams. 1986. Learning internal representations by error propagation. In David Rumelhart and James McClelland (eds.), *Parallel Distributed Processing*, Vol. 1, 318-62. Cambridge, MA: MIT Press.

- Rumelhart, David and James McClelland. 1986a. *Parallel Distributed Processing*, Vol. 1. Cambridge, MA: MIT Press.
- Rumelhart, David E. and James L. McClelland. 1986b. On learning the past tense of English verbs. In James L. McClelland and David E. Rumelhart (eds.), *Parallel Distributed Processing*, Vol. 2, 216-71. Cambridge, MA: MIT Press.
- Rumelhart, David and James McClelland. 1987. Learning the past tenses of English verbs: Implicit rules or parallel distributed processes? In Brian MacWhinney (ed.), *Mechanisms of Language Acquisition*, 195-248. Hillsdale, N.J.: Lawrence Erlbaum.
- Seidenberg, Mark and James McClelland. 1989. A distributed, developmental model of word recognition and naming. *Psychological Review* 96, 523-568.
- Squire, Larry. 1987. *Memory and Brain*. New York: Oxford University Press.
- St. John, Mark. 1992. The story gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Science* 16, 271-306.
- Stemberger, Joseph. 1985. *The Lexicon in a Model of Language Production*. New York: Garland.
- Touretzky, David. 1990. BoltzCONS: Dynamic symbol structures in a connectionist network. *Artificial Intelligence* 46, 5-46.
- Trammell, Robert. 1978. The psychological reality of underlying forms and rules for stress. *Journal of Psycholinguistic Research* 7, 79-94.
- Twain, Mark. 1935. The awful German language. *The Family Mark Twain*. New York: Harper and Brothers. [From *A Tramp Abroad*, originally published 1880.]
- von Neumann, John. 1956. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In Claude Shannon and John McCarthy (eds.), *Automata Studies*, 43-98. Princeton, NJ: Princeton University Press.
- Wexler, Kenneth and Peter Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.
- Zubin, David and Klaus-Michael Köpcke. 1981. Gender: A less than arbitrary grammatical category. *Papers from the Seventeenth Regional Meeting of the Chicago Linguistic Society*, 439-449. Chicago: CLS.
- Zubin, David A. and Klaus-Michael Köpcke. 1986. Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. In Colette Craig (ed.), *Noun Classes and Categorization*, 139-180. Amsterdam: John Benjamins.