The Dinosaurs and the Ring

Brian MacWhinney
Carnegie Mellon University

In R. Corrigan, G. Iverson, & S. Lima (Eds.) (1994). The reality of linguistic rules. Amsterdam: Benjamins.

Abstract

Cognitive models based on rules and symbols typically require a high degree of hand-wiring. One way of avoiding the hand-wiring of rules is to rely on connectionist networks for modeling some of the core processes in language acquisition. However, connectionist models have problems with scalability and generativity. Ideally, we want a model that will keep the strengths of both symbolic and connectionist models, while overcoming their respective shortcomings. One model that may be able to do this is a connectionist extension of the Competition Model which relies on the lexical item and lexical categories as ways of coordinating processing and learning on the levels of syntax, audition, articulation, and semantics.

1. The Dinosaurs

Rules have a long and venerable history within both linguistics and cognitive science. Within linguistics, the dynamic duo of rules and categories has formed the backbone of some of the best work in both theory and description. However, recent work in connectionist modeling (Rumelhart & McClelland, 1986) has challenged the ascendancy of rules and categories, focusing attention instead on models based on simple, observable cues and connections between these cues.

If rules were so successful, why did researchers feel the need to explore an alternative paradigm? At least part of the answer can be framed in the context of the events that occurred during a period which I will call the era of the Big Mean Rules. During this Golden Age of Transformational Grammar, linguistic categories and complex rules based upon these categories flourished in great diversity. Like the dinosaurs of the Late Cretaceous, these rules became huge and terrifying. An example of a piece of one of the scariest of the rules of this period is the Laxing rule from the Chomksy and Halle (Chomsky & Halle, 1968).

LAXING

(I) AUXILIARY REDUCTION — I

$$V \rightarrow \begin{bmatrix} \text{- stress} \\ \text{- tense} \end{bmatrix} \Big/ \left\{ \begin{array}{l} \langle VC_0 \rangle \begin{bmatrix} \overline{\quad\quad} \\ \alpha\text{stress} \\ \langle +\text{tense} \rangle \end{bmatrix} C_0^1(=C_0) \begin{bmatrix} \beta\text{stress} \\ V \end{bmatrix} \\ \left\{ \begin{array}{l} \begin{bmatrix} \overline{\quad\quad} \\ \gamma\text{stress} \end{bmatrix} \\ [1\ \text{stress}]\ C_0 \underline{\quad\quad} C_0[\text{-cons}] \end{array} \right\} [\text{-stress}]_0\ \# \end{array} \right\} \qquad \text{III (118)}$$

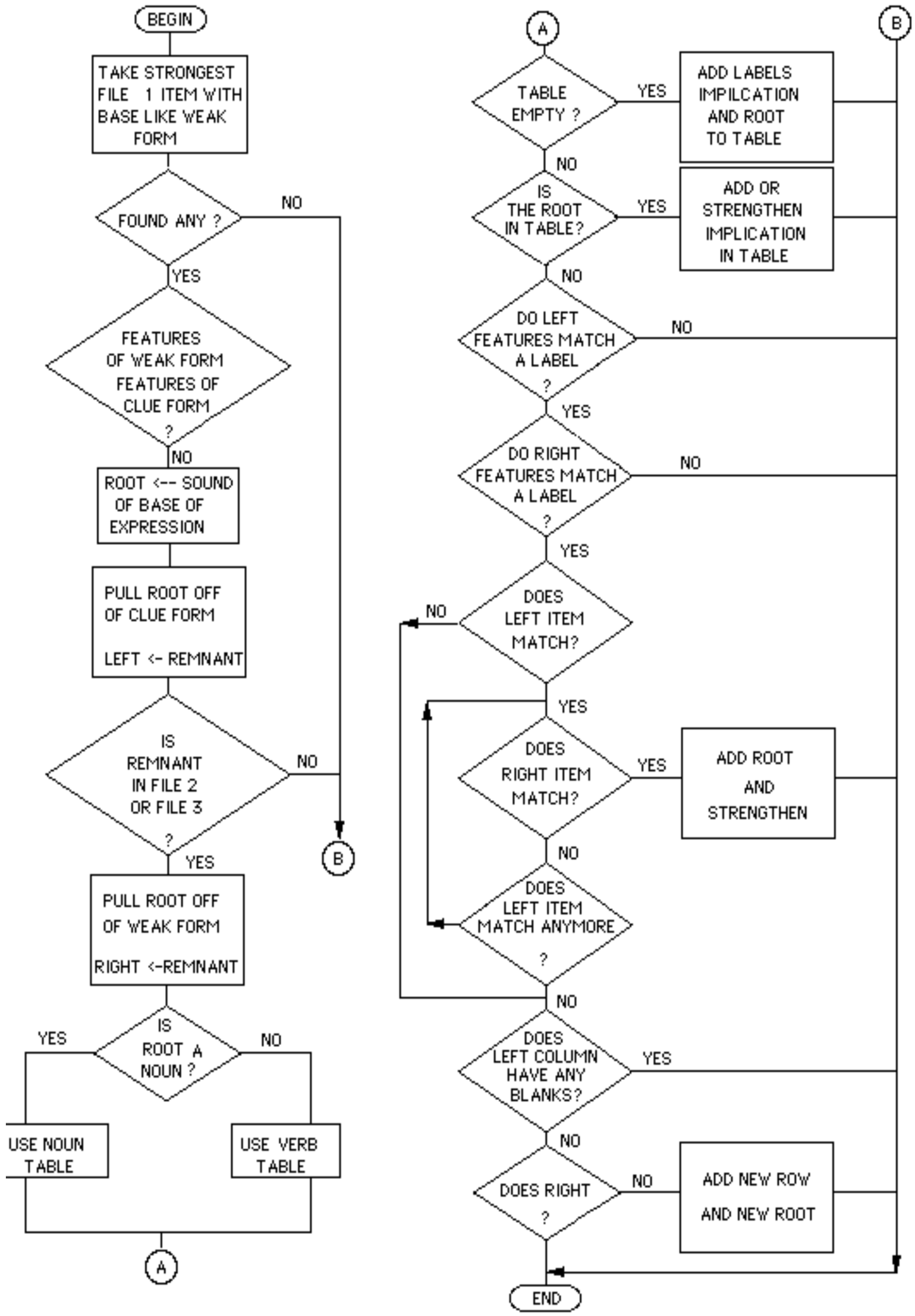
Conditions: β=1,2,3
α is weaker than β
γ is weaker than 2

(II) $V \rightarrow [\text{- tense}] \ / + \underline{\quad\quad}\ \text{r+i[-seg]}$ III (142); IV p.202

(III) $V \rightarrow [\text{- tense}] \ / \ \underline{\quad\quad}\ [\text{+cons}] \begin{bmatrix} \text{+cons} \\ \text{-voc} \end{bmatrix}$ IV (8)

(IV) $\begin{bmatrix} V \\ \alpha\text{round} \\ \alpha\text{back} \end{bmatrix} \rightarrow [\text{-tense}] \ \Big/ \ \underline{\quad\quad}\ C \left\{ \begin{array}{l} C_0 + \mathit{ic}, + \mathit{id}, + \mathit{ish} \\ (C_1+) \begin{bmatrix} \text{-stress} \\ V \end{bmatrix} C_0[\text{-cons}] \end{array} \right\}$ IV (19), note 16, and p. 192

Just as the Big Mean Rules of transformational grammar dominated the Linguistic Plain, so the Big Mean Flow Charts of information-processing psychology lorded over the Psycholinguistic Plain. An representative example of this other class of symbolic dinosaurs is this flow chart for German paradigm formation from MacWhinney (MacWhinney, 1978):



The success of the great rule systems was closely linked to the belief in the rule-governed nature of linguistic behavior. If rules are real, it should be possible to observe their operation through standard psychological techniques. However, attempts to demonstrate the psychological reality of these rules (Fodor, Bever, & Garrett, 1974; Jaeger, 1980; Linell, 1979; Ohala, 1974a; Ohala, 1974b; Trammell, 1978) yielded disappointing results. Without strong evidence for psychological reality, the Big Mean Rules and the Big Mean Flowcharts began to collapse under the weight of their enormous mass of ad hoc assumptions and nativist theoretical baggage. There were proofs of learnability and non-learnability, critiques of the role of curly brackets, restatements of the competence-performance distinction, and attempts to modularize assumptions. But none of these last-ditch attempts to save this dying breed could rescue the great rule-based carnivores from the mass extinctions occurring in the pyrrhic victories at the end of the Linguistic Wars (Newmeyer, 1980).

2. The Ring

The great rule systems were destroyed not so much by their size and complexity as by their excessive reliance on formal power. They derived this power from the insights embodied in the computational architecture of the Von Neumann serial computer and the corresponding application of this architecture to human cognition by Simon, Newell, and their followers (Klahr & Kotovsky, 1991; Newell & 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. The ability of production systems to model any structure or process directly and completely was the Magic Ring of the symbolic approach. As in Tolkien's Lord of the Rings (Tolkien, 1965), the builders of the great rules had access to a special gift whose exercise demanded modesty and caution, lest excessive use of its power lead to the corruption and downfall of its owner. A modeler could take a few symbols, concatenate them into rules and, magically, the computer could conjure up a working model of mental processing. But this power was deceptive and could turn on its creator. Virtually any human or machine behavior could be formalized through a potentially limitless number of alternative, equally persuasive production systems. Moreover, attempts to identify the correct model without adding further constraints were shown to be flawed in principle (Anderson, 1978). The most extreme impact of this indeterminacy was felt in linguistics, where academic wars between generative semantics and generative syntax were waged under the banners of competing production system architectures, both of which allowed themselves access to limitless power.

The Age of the Big Mean Rules was followed by the Period of the Constraints. Awed by the power of the Ring, researchers became increasingly concerned with the need to rein in the power of rule systems. Within the symbolic tradition, there were attempts to trim the power of rules by moving down the Chomsky hierarchy (Hopcroft & Ullman, 1979) with some systems even flirting with the once ridiculed Markov Model (Chomsky, 1959; Hausser, 1989). Constraint models, such as G-B theory (Chomsky, 1981), allow patterns to multiply profusely, but then attempt to bring them back under control by specifying constraints or filters on their output. However, the hand-wired complexity of these constraint systems often rivals that of the great rule systems. The more modern weight-watcher approach to rule construction takes a somewhat different tack. It focuses attention on a few, clearly productive rules, specifying only the core of the rule, leaving the exceptions to exhaustive listings or analogistic processing. The new race of models based on these kinder, gentler rules echoes the linguistic descriptivism of the beginning of the century (Bloomfield, 1914). A frequently discussed rule of this type is Pinker's "add -ed" rule (Pinker, 1991), which states, quite unobtrusively, that one can form the past tense of an English verb by simply adding -ed. There is no need to add any further complexity to this rule, since all of the complexities are bequeathed to either rote memorization or connectionist nets.

Symbolic models based on constraints and kinder, gentler rules are now competing with a totally new breed of models -- the neural nets (Grossberg, 1987; Hopfield, 1982; Kohonen, 1982). These models take a determined approach to limiting the power of the Ring, by imposing two stringent limitations on cognitive models. First, these models require that the computations involved in the models echo the connectionist architecture of the brain. Neuroscience has shown that the brain cannot use memory addresses to bind variables; there is no plausible neural mechanism that can assign an absolute "address" to a particular neuron (Squire, 1987). There is no general scheme for register assignment, data pathing, or memory addressing in the brain. 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 that provides the power underlying the von Neumann architecture. 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 control the power of the Ring. One can 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, & Palmer, 1991). A early example of a fully hand-wired connectionist architecture was Lamb's stratificational grammar (Lamb, 1966). More recently, we have seen hand-wired models in areas such as speech errors (Dell, 1986; MacWhinney & Anderson, 1986; Stemberger, 1985), ambiguity resolution (Cottrell, 1987), and lexical activation (Marslen-Wilson, 1987). The only apparent limitation to the power of these hand-wired systems is the energy and creativity of the modeler. Yet another approach to hand-wiring spares the modeler the tedium of hand-wiring by running the wiring procedure off symbolic templates. For example, Touretzky (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 the power of the symbolic Ring.

In order truly to constrain the power of the Ring, modelers must match the constraint against symbol passing with the constraint that networks be self-organizing on the basis of simple input data. In some cases, this self-organization can emerge without external instruction, although it would be unwise to assume that some form of negative information is not available to the language learner (Bohannon, MacWhinney, & Snow, 1990). When one applies the twin constraints of local computation and self-organization on input data, the class of potential models of language learning becomes extremely limited. In fact, there is currently no reasonably detailed model of language acquisition that can rigorously 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. Moreover, 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 & Kawamoto, 1986), graceful degradation (Harley & MacAndrew, 1992; Hinton & Shallice, 1991; Marchman, 1992), and property emergence (MacWhinney, Leinbach, Taraban, & McDonald, 1989).

Building models without the power of the Ring is a tough job. Some modelers try to make headway by providing detailed accounts of small pieces of the language acquisition puzzle. For example, there are modular networks which have constrained themselves to well-defined topics such as the acquisition of the English past tense (Cottrell & 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 & Elman, 1986). These toy models use only a few sentences or a few words. When one attempts to add additional words or sentences to these models, their performance typically begins to degenerate. Both the network modules and the toy models suffer from these problems with scalability. In order to achieve scalability, connectionist models will need to develop self-organizing ways of solidifying structures that can support intercommunication between the various forms of language processing.

3. Two Babies and the Report Card

Currently, neither connectionist models nor kinder, gentler symbolism provides a complete and satisfactory framework for understanding the acquisition of language. Both approaches have their strengths and their weaknesses. On balance, one could register the following report card from the two approaches, with pluses indicating strengths and minuses indicating weaknesses:

	Symbolic	Connectionist
Generativity	+	-
Scalability	+	-
Crispness	+	-
Minimize hand-wiring	-	+
Gradation, Leakage	-	+
Learning from input	-	+
Constrain Symbol Passing	-	+

On this report card, we see the symbolic approach excelling in terms of its ability to generate crisp, scalable models of a wide variety of human behaviors. Moreover, it is precisely in the areas of generativity, scalability, and crispness that connectionist models do most poorly. But, when we look at the relative ability of the models to minimize hand-wiring, to account for gradation and leakage, to place a full focus on learning from input, and to subscribe to brain-like prohibitions on symbol passing, it is the connectionist models that clearly excel.

We want to avoid throwing out either the symbolic baby with the symbolic bath water or the connectionist baby with its connectionist bath water. We want to rid ourselves of the ad hoc hand-wiring of the symbolic approach. But we don't want to sacrifice the generativity and scalability achieved by the symbolic approach. We want to rid ourselves of the problems of toy models and systemic opacity so often found in connectionist models, but we want to retain the principles of connected computation and self-organization. We cannot achieve a better report card by simply patching up a few holes in our AI programs or by adding a few hidden units to our connectionist nets. Attempts to address all of the key criteria measured in this report card will require a fundamentally new level of conceptualization.

4. Core Issues in Language Acquisition

Like the symbolic models, the connectionist models seek to provide 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. To get a sense of the scope of connectionist concerns, let us quickly list some of the issues that have been addressed. In each case I will only cite one or two articles, just to give the reader a pointer to further reading.

1. **Cues vs. rules.** Do children learn linguistic rules or do they acquire systems for managing cue interactions? (MacWhinney et al., 1989)
2. **Competition vs. hypothesis testing.** Do children use hypothesis testing and innate ideas to form categories or does language development depend on the formation of coalitions and competitions between cues? (MacWhinney, 1989)

3. **Recovery from overgeneralization.** How does a child learn to stop saying "costed" and say "cost" instead? How does a child learn not to say "unsqueeze?" (Li, 1992; Plunkett & Marchman, 1991)
4. **Structure building**. How does the mind achieve a representation of syntactic structure sufficient to control the building of conceptual structure? When we hear that the "big dog and the little dog chased the cat" how do we know that it wasn't a medium sized dog that chased a cat? (Barnden & Srinivas, 1991)
5. **Segmentation**. How does the learner pull new words out of the ongoing stream of speech? (Grossberg & Stone, 1986)
6. **Analysis**. How does the child analyze complex words and phrases? (Elman, 1993).
7. **Semantic mapping**. When acquiring a new word, how does the child solve Quine's famous "gavagai" problem? (MacWhinney, 1989)
8. **Fast mapping - slow learning**. How is it that the child can so quickly form an association between sound and meaning, but requires years for learning the full range of meanings of that word? (Harris, 1990)
9. **Underlying form**. What is the role of abstract or underlying phonological form in lexical representations? Is this form primarily auditory, primarily articulatory, or both? (Hare, 1990)
10. **Phonological processes**. How does the child learn to overcome the articulatory processes that so severely limit the shape of early articulatory output? (Menn, Markey, Mozer, & Lewis, 1992)
11. **Storage in processing**. What is the role of the articulatory loop or other memories in imitation, segmentation, and syntactic processing? (Burgess & Hitch, 1992; Houghton, 1990)
12. **Catastrophic interference**. How can models of second language acquisition avoid the problem of catastrophic interference? (Kruschke, 1992)
13. **Transfer**. How can models capture the ways in which transfer from L1 to L2 is facilitated by structural similarities? (Chrisman, 1991; Gasser, 1990)
14. **Fossilization and critical periods.** Are there inherent limits on the capacity for the acquisition of a second language in adulthood? (Elman, 1993; Johnson & Newport, 1989).
15. **Plasticity**. Why do children who suffer damage to large areas of the brain often end up with better control of language than other children with no apparent damage? (Feldman, Holland, Kemp, & Janosky, 1992; Marchman, 1992; Tallal & Stark, 1980)
16. **Gradation of Loss**. How much does the pattern of variability in processing by normals account for processing by subjects with disorders? (Kilborn, 1989; Marchman, 1992)
17. **Dissociations and modularity**. Are there clear examples of dissociations between component language skills and between language and general cognition? (Farah & McClelland, 1991)

Although connectionists have begun to formulate approaches to each of these core issues, much of this work is still extremely preliminary and fragmentary, with no overall picture of language acquisition yet emerging from the various pieces.

5. The English Past Tense

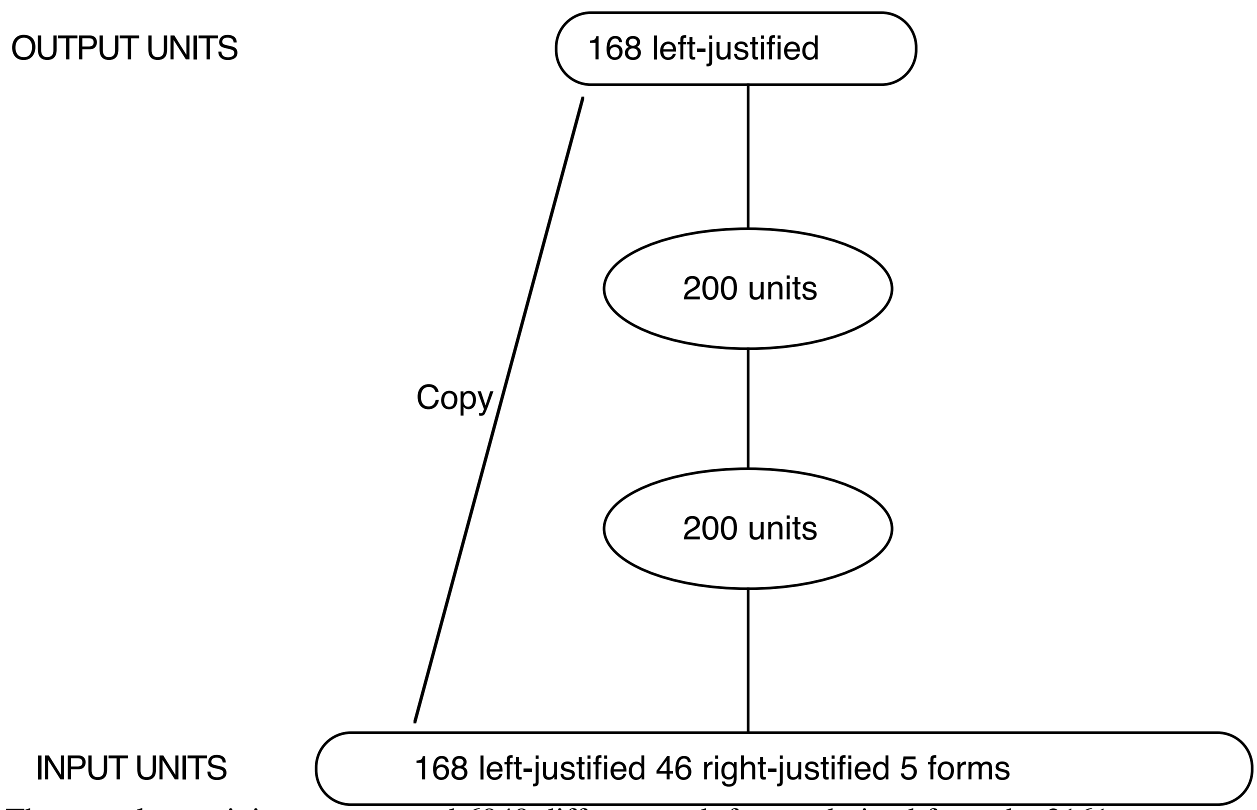
The remainder of this paper sketches out the achievements of connectionist models in one specific well-researched area and then examines how we can move from these preliminary achievements to a fuller, more explanatory, unified approach to all of the core issues facing language acquisition theory.

Let me begin by reviewing some recent connectionist models of the learning of inflectional morphology. To begin, let us examine the simulation of the acquisition of

English verb marking developed by MacWhinney and Leinbach (MacWhinney & Leinbach, 1991). The task of the MacWhinney-Leinbach model was to convert the stem of an English verb into another inflected form. For example, given a stem such as “eat”, the model could produce “eats”, “eating”, “ate”, or “eaten.” The model took as its input a rich, slot-coded, phonological representation. The basic phonological representation used 12 consonantal slots and 6 vowel slots. Together these 18 slots constituted the autosegmental grid for the model. The actual segments of the stem were filled into this grid in either a left-justified or a right justified fashion. For example, the word “bet” would fill out the grid in this way:

```
left-justified:     bCC EV tCC VV CCC VV CCC
right justified:    CCC VV CCC VV CCb VE CCt
```

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 consonantal slot in the left-justified representation for “bet” would have 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. The architecture of the network had this shape:



The complete training corpus used 6949 different verb forms, derived from the 2161 highest frequency verbs in English (Francis & Kucera, 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 Kucera (Francis & Kucera, 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, & 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. The model correctly addressed the various complaints leveled by Pinker and Prince (Pinker & Prince, 1988) against the earlier verb-learning model of Rumelhart and McClelland (Rumelhart & McClelland, 1987).

Despite its basic successes, there were four weaknesses in the MacWhinney-Leinbach model, and these failures are actually more instructive than the successes.

1. **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.
2. **Homophony**. Several English verbs have present tense forms that sound the same along with past tense forms that are clearly different. For example, we say "the maid wrung out the clothes", "the soldiers ringed the city", and "the choirboy rang the bell." These three different past tense forms all have the same sound /rIN/ in the present. Because the MacWhinney-Leinbach model uses only the phonology of the stem as its input, it could not control this distinction.
3. **Compounds.** The 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". If 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 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 "undergo-underwent".
4. **Derivational status.** Finally, the model was not capable of utilizing information regarding the derivational status of lexical items. As Kim, Pinker, Prince, and Prasada (Kim, Pinker, Prince, & Prasada, 1990) have noted, the past tense forms of denominal verbs are uniformly regular. For example, the verb "fly" as used in baseball derives from the noun "a fly". We say that a batter "flied out to center field" rather than "flew out to center field". It is clear that inflectional processes must be able to access information regarding the derivational status of words.

These four weaknesses can be linked to a single core problem: an overreliance on phonological cues. The model was limited to only performing a sound-to-sound conversion without any meaning-to-sound conversion. Sound-to-sound conversion models make the assumption that, in producing inflected words, we first access the base of a stem and then activate its derived forms. This boils down to the claim that the only information available to the inflectional processor is phonological form. This strict assumption fits in well with a modular processor of the type proposed by Fodor (Fodor, 1983). However, it does not fit in well with the interactive ethos of connectionist modeling, nor does it fit in very well with what we know about lexical production (Stemberger, 1985; Stemberger & MacWhinney, 1988).

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 (this volume) has developed a fully connected auto-associative network that uses both phonological regularities and specific meaning-sound associations in the same simple network. The model does this not through overt construction of two separate routes, but through activation of two separate processes with different time courses. For example, when the state of the network is read out quickly, the orthographic form "pint" is pronounced as if it rhymed with "hint." Kawamoto argues that this shows the operation of a quick, pattern-based mode of activation. When more time is allowed, the model pronounces "pint" correctly. When this happens, the model appears to be retrieving specific facts about the semantics, orthography, and pronunciation of individual lexical item "pint." Because the model can act in these two modes, it can mimic the operation of a "dual-route" model of lexical access (Besner, Twilley, McCann, & Seergobin, 1990), despite the fact that only a single route is implemented in the network. What gives rise to these dual-route effects is the fact that general phonological patterns have a quick rise time and item-specific patterns are slower.

Other models that perform meaning-to-sound conversion have been developed by Hoeffner (Hoeffner, 1992), Cottrell and Plunkett (Cottrell & Plunkett, 1991), and MacWhinney and Leinbach (MacWhinney & Leinbach, 1991). For example, MacWhinney and Leinbach supplemented their basic model with a second meaning-based model which could distinguish between "ringed", "rang", and "wrung", because the inputs in the three cases were very different. However, the model began to bog down when it was asked to deal with hundreds of forms. In an attempt to improve on this performance, Hoeffner (Hoeffner, 1992) used a simplified coding scheme and a more powerful architecture with recurrent connections. He found that it was possible to encode up to 500 past tense forms directly from meaning. However, at this point his model also began to bog down. Problems with scalability also seem to crop up with the other meaning-based models. All of these models represent the meanings of lexical items through a distributed pattern of nodes. As long as the model is only encoding a few items, this pattern is sufficient to encode the needed distinctions. However, as the number of words rises, the network starts to become clogged with the numbers of distinctions that have to be stored.

Attempts to address the issue of access from meaning appear to be running up against a serious scalability problem. Somehow these systems must be able to add additional lexical items to their access path without having the introduction of new items choke down the computational power of the network. A similar conclusion emerges from work with the learning of German morphology. In the next section, we will see that the learning of German morphology depends crucially on the processing of lexical structures and lexical categories.

6. Learning Arbitrary Form Classes

The inflectional systems of languages such as German, French, and Latin have often been used as illustrations of the extreme arbitrariness and unpredictability of grammatical systems. Mark Twain (Twain, 1935) 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 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." Maratsos (1979) echoed this theme when he wrote that, "... the gender categories of most Indo-European languages ... do not agree with anything in the practical world ...

there seems to be no practical criterion by which the gender of a noun in German, French, or Latin could be determined (pp. 271, 280)." Twain and Maratsos are certainly not alone in this opinion. Anyone who has studied one of these languages, 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 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 following paradigm for the definite article:

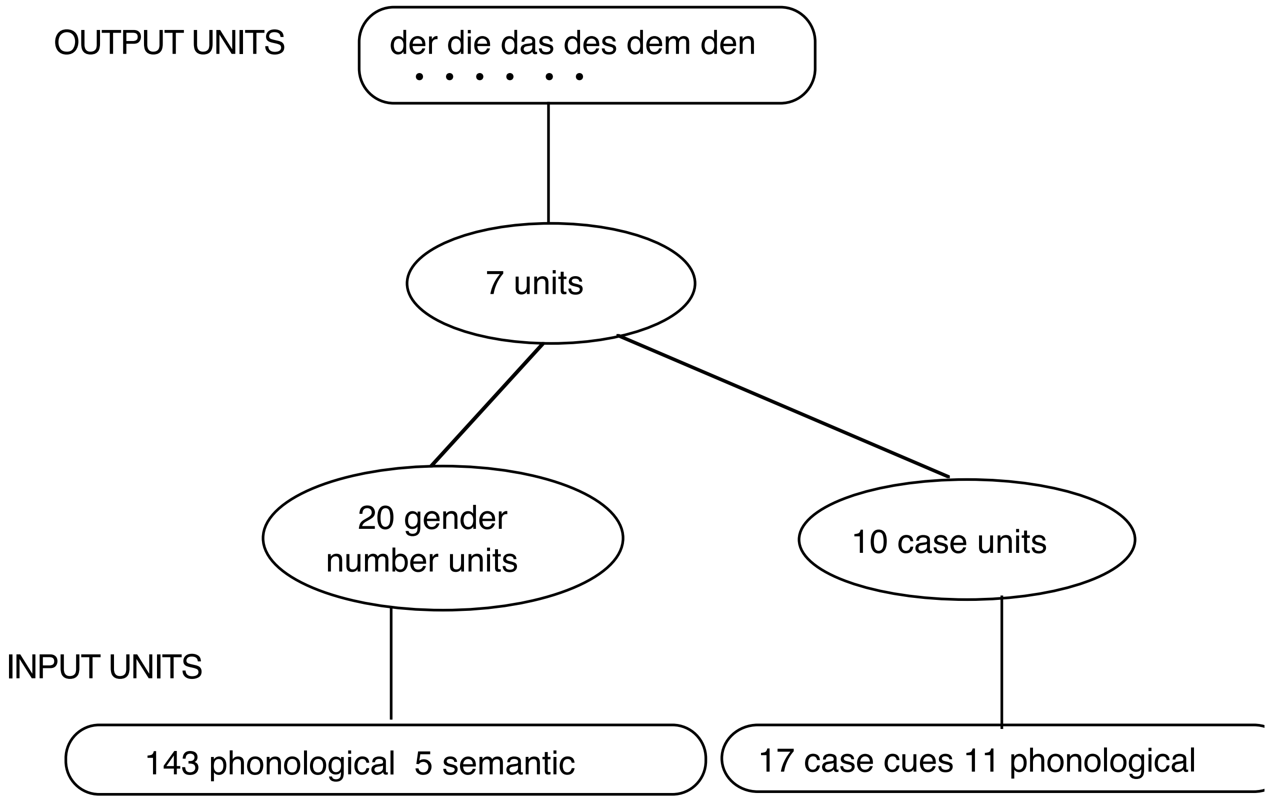
	Masc	**Fem**	**Neut**	**Plural**
Nom	**der**	**die**	**das**	**die**
Gen	**des**	**der**	**des**	**der**
Dat	**dem**	**der**	**dem**	**der**
Acc	**den**	**die**	**das**	**die**

This paradigm is rife with homonymy. The six forms of the definite article (der, die, das, dem, des, den) must cover the 16 cells in the paradigm. This is done by having a single form 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 bears an obvious relation to real-world categories. Case is somewhat more abstract, but can generally be figured out through a combination of cues from the verb, prepositions, and some word order patterns. There is little in the objective situation or the other words in the sentence that can help the child figure out the gender of a noun. It is possible that the noun's gender could be simply memorized or even inferred on the basis of its use within the paradigm. However, recent work by Köpcke and Zubin (Köpcke, 1982; Köpcke, 1988; Köpcke & Zubin, 1983; Köpcke & Zubin, 1984; Zubin & Köpcke, 1981; Zubin & Köpcke, 1986) has shown that the Twain-Maratsos 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 specify a particular gender.

MacWhinney, Leinbach, Taraban, and McDonald (MacWhinney et al., 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 characterizations of linguistic systems. The architecture of the more successful non-handcrafted German simulation had this form:



The input to the network was a pattern across 143 phonological units to represent the noun stem and 11 phonological units to represent suffixes attached to the noun. In addition, there were 5 semantic units representing inherent gender and 17 cues that provided a distributed pattern of information regarding the target case for the noun. This network was trained with 2000 German nouns from all cells in the paradigm. It learned the training set perfectly. 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.

The model also succeeded in capturing a variety of important developmental phenomena. Like the children studied by MacWhinney (MacWhinney, 1978) and Mills (Mills, 1986), the model showed early acquisiiton 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.

Although the MacWhinney et al. model was successful in many ways, it displayed many of the same weaknesses that plagued the MacWhinney and Leinbach model for English. The model did not have a problem with failure to acquire early irregulars, since real learners display no consistent regular pattern and no u-shaped learning curve for the

German article. However, the problems of homophony and derivational status were similar, and there were some new problems:

1. **Homophony**. Just as the English model suffered from an inability to control the alternation between homophones, the German model could not control the fact that phonologically identical nouns can take different genders. For example "der Band" (volume, book) contrasts with "das Band" (rubber band), just as "der Bund" (alliance) contrasts with "das Bund" (bundle). These problems also affect the formation of the plural. For example, the singular form "das Wort" has two plural forms "die Wörter" (several words) and "die Worte" (connectioned words or speech). The MacWhinney et al. model cannot control these alternations because it has no capacity to encode lexical semantics.
2. **Derivational status.** German provides some clear examples of the importance of derivational status. Given the assumptions of the model, there was no way to encode the fact that nouns formed through zero-derivation from verbs are typically masculine. For example, the noun "der Schlag" (blow, cream) derives from the verb "schlagen" (to hit), but including this fact in the simulation would have required an ad hoc "zero-derivation" node.
3. **False derivationals**. The model includes no independent way of representing derivational suffixes and prefixes. 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 clear 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.
4. **Compounds**. Just as the English model was unable to treat "undergo" as a combination of a prefix and the stem "go", so the German model had trouble treating "Grossmutter" (grandmother) as a combination of a modifier and the stem "Mutter" (mother). Because the frequency of "Grossmutter" is much less than that of "Mutter" and because the ending -er tends to cue masculine gender, the model had trouble learning that "Grossmutter" should be feminine.

In the English simulations, these weaknesses had no obvious impact outside the area of the past tense. In German, however, problems in learning noun gender lead to further problems in other parts of the grammar. Pronouns and relativizers depend crucially on the gender of the governing noun and can be misinterpreted when gender marking is incorrect. Plural formation relies on gender cues in concert with phonological and morphological cues. Inflection of the adjective both in noun phrases and as a predicate adjective has a complex set of relations to both noun gender and the presence of absence of other adjectives and determiners.

The basic weaknesses of these models should appear in only slightly different forms for other languages. For example, the problems with "Grossmutter" in German or "overthrown" in English should crop up in a slightly different form in French. In both English and German, the principle that requires the final noun to determine the gender of the compound is fairly solid. In French, however (Alain Desrochers, personal communication), deverbal compounds such as "le portefeuille" (carry-paper, wallet) go against this principle and are uniformly masculine, even when they contain a final feminine noun such as "la feuille" (the paper). Without an ability to deal with properties specific to lexical items, networks will have an extremely hard time dealing with these patterns and counterpatterns. French also shows how phonological patterns can lead modifiers to shift from their otherwise clear gender marking. For example, "son echelle" (his/her ladder) has a masculine modifier which switches into the feminine before an adjective beginning with a consonant as in "sa grande echelle" (his/her long ladder).

7. Lexical items and lexical categories

Given the seriousness of these problems and the extent to which they have limited the full effectiveness of these network models for English and German, we decided to explore an alternative conceptualization. The core assumption in this new model is that the lexical item serves a central controlling and stabilizing role in language learning and processing. Localist connectionist models of the type developed by Dell (Dell, 1986) and Stemberger (Stemberger, 1985) provide just such a central role to the lexical item. However, because of their localist node-based architecture, these models have been forced to rely on hand-wiring. The Competition Model (MacWhinney, 1987a; MacWhinney, 1988; MacWhinney, 1989; MacWhinney & Anderson, 1986) also relies on localist, hand-wired connections. These models provide good descriptions of particular experimental results, but at the price of relying on symbolic hand-wiring.

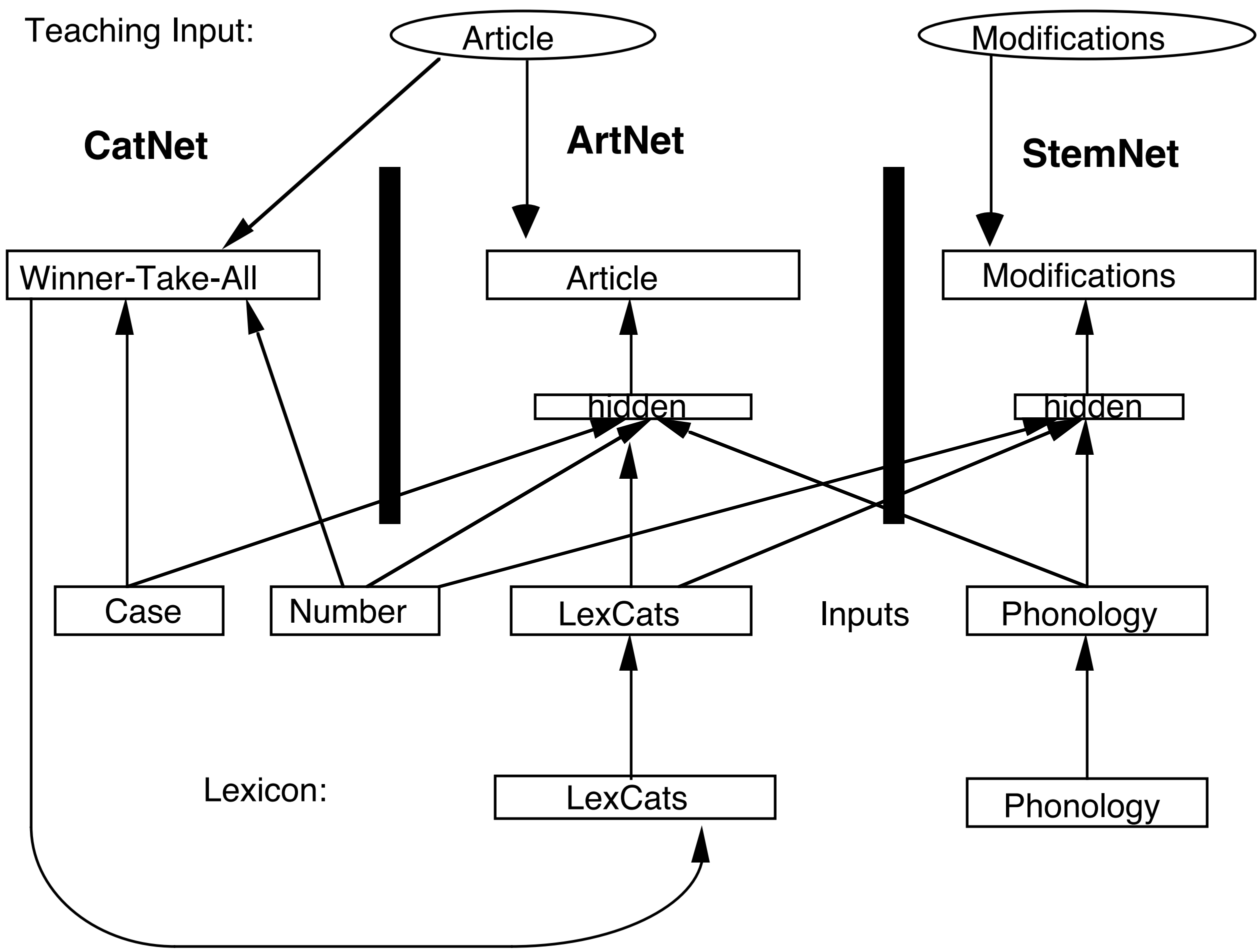
A model designed to address these problems is the Connectionist Competition Model or ConComp model. There is no sharp discontinuity between earlier versions of the Competition Model and the ConComp model. For example, Taraban, McDonald, and MacWhinney (Taraban, McDonald, & MacWhinney, 1989) showed that the Competition Model constructs of cue availability, reliability, and conflict reliability correlate well with successive phases of learning in the the connectionist model for German (MacWhinney et al., 1989). McDonald and MacWhinney also explored relations between the Competition Model and the Fuzzy Logical Model of Perception of Massaro (Massaro, 1987). Similarities between a variety of statistically-oriented models of this type have often been noted (Anderson, 1990; Cohen & Massaro, 1992).

What most clearly distinguishes the newer ConComp model from other connectionist models (Daugherty & Seidenberg, 1994; MacWhinney et al., 1989; Plunkett & Marchman, 1991; Rumelhart & McClelland, 1987) is the central role assigned within the ConComp model to the lexical item. The first version of the ConComp model was developed by Gupta and MacWhinney (Gupta & MacWhinney, 1992) to coordinate article marking with learning of the case and number markings on the noun. As an example of a typical set of case/number markings for a German noun, consider this paradigm for the noun “Wort” (word):

	Singular	**Plural**
Nom.	**Wort**	**Wörter**
Gen.	**Wortes**	**Wörter**
Dat.	**Wort**	**Wörtern**
Acc.	**Wort**	**Wörter**

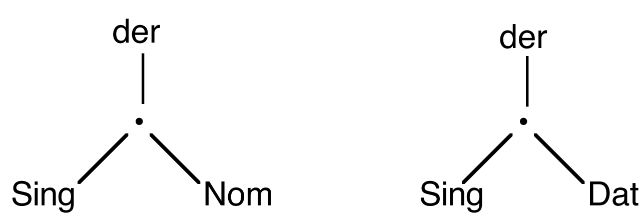
There are nine different ways in which German plurals can be formed and the exact shape of the endings in the genitive and the plural dative varies in complex ways with the phonology of the stem. Altogether, one could distinguish about 20 major inflectional paradigms for German nouns (Engel, 1991).

There are two new structures built into the ConComp model: a system for extracting lexical categories and a method for storing lexical categories as properties of individual lexical items. The overall network for this system is given in the following diagram. Computation in this network occurs in parallel interactively. The separation of the network into three segments is done largely for explanatory purposes.



Lexical category learning is done in the CatNet which is responsible for acquiring a set of lexical categories (lexcats) to express the association of particular articles with particular cases and numbers. These lexical category units have one-to-one connections to matching units on each lexical item, as indicated at the bottom of the chart. The basic article selection performed by the MacWhinney-Leinbach model is done in the ArtNet, and the new task of inflecting the noun stem is done in StemNet. The construction of gender categories is done by the operation of hidden units in the ArtNet on the basis of the information in both the phonology of the stem and its lexical categories. The ArtNet and the StemNet share the same set of inputs, but differ in having a separate set of hidden units and different outputs.

The network works in the following way. When a noun is presented to the system, there are four types of inputs to the hidden units in the ArtNet and the StemNet. The case and number units code the four cases and two numbers of German. The phonology of the noun is coded in the left-justified form discussed earlier. Finally, the input also includes a set of activations for the lexical category units of the current noun. When a noun is activated, its lexical categories are turned on and these then transmit activation to the hidden units. The training signal provides input on the correct form of the definite article and the correct modifications for the noun. The information regarding the correct form of the definite article goes both to the ArtNet back-propagation network and to the winner-take-all competitive learning system in the CatNet. The CatNet begins learning with a set of 40 units that are available for category extraction. The choice of a particular number of units for this learning is arbitrary. We have also tried learning with a larger set of potential units and the results are similar. Each of the units in the winner-takes-all net competes for activation whenever there is a particular combination of an article, a case, and a number, as in these examples:



Through the first five epochs of learning, these 40 units come to settle on a encoding system in which 13 units respond distinctively and the other 27 units fall into disuse. This learning is based on an unsupervised association between the article and the cooccurring case and number units.

As particular units begin to emerge in the CatNet, their activation is copied across through one-to-one identity connections to lexical category entries in the lexicon. For example, if we look at three units that code for different meanings of the article "der" we see how the feminine noun "Mutter" picks up weightings on two lexcat units, whereas the masculine noun "Mann" comes to pick up weightings for the other lexical category unit related to the article "der".

```
                •   •   •   •   •   •   •   •     •
                der                  der              der
                sg                   sg               sg
                nom                  dat              gen

Mutter          -    -    -    -    •    -    -    •    -

Mann            •    -    -    -    -    -    -    -    -
```

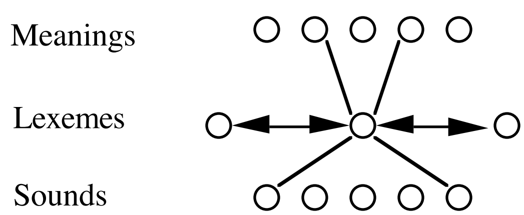
Although this representation has some noise in the first epochs, non-repeated activations decay and the pattern cleans itself up over time.

It is useful to compare the operation of the network with and without lexical category units. With these units present, the model was able to learn all the articles for a 2094 word corpus in 65 epochs and all the noun modifications in 70 epochs. Without lexical categories, there were still 34 article errors at 65 epochs and there were still 53 noun inflection errors at 70 epochs. With a larger input corpus, errors will increase and networks without lexical categories will be unable to fully learn their training set. It is clear that the learning of lexical categories is crucial for the successful learning of the full system.

7.1. Lexical associations as nodes

This new scheme extends the capacity of the network to acquire the full system of inflectional markings for German. It also casts a spotlight on the pivotal role of lexical items as controllers of the learning process. If we can assume that lexical items are available, then we can address all of the residual problems with our earlier models for English and German, including the problems of homophony, early irregulars, derivational status, the influence of the head of compounds, syntactic influences, and potential catastrophic interference effects (Kruschke, 1992; McCloskey & Cohen, 1989). But how can lexical items be learned in a connectionist framework?

One approach looks at lexical items as single nodes (Morton, 1970; Morton, 1980) competing in a winner-take-all system, as in this sketch:



Our simulations of lexical learning in this scheme have succeeded in associating patterns of sound to patterns of meaning, thereby showing how the lexical items assumed in the ConComp model can plausibly be learned. However, like the other network architectures we have discussed, these associative nets soon reach an upper limit when asked to encode several hundred lexical associations.

7.2. Lexical associations as networks

In order to get beyond this upper limit, it may be necessary to dig deeper into the structure of the lexical item. One interesting approach, for which we are just now developing simulations, views each lexical item as a complex network of “cells” or “nodes”. At the beginning of language learning the cells within each lexical network are richly interconnected. Each cell in the network maintains at least one of these five types of connections:

1. **Phonological receptors**. Cells with connections to auditory cortex are able to activate the lexical item on the basis of phonological input.
2. **Conceptual receptors**. Cells with diffuse connections to other parietal regions are able to activate visual and conceptual interpretations of lexical items.
3. **Motor activators**. Cells with connections to Broca's area and motor cortex will eventually be able to activate the articulatory form for the lexical item.
4. **Syntactic connections**. Another class of cells within the lexical net maintains connections to cells that can specialize in the extraction of lexical categories.
5. **Cohort inhibitors**. Yet another set of cells within the lexical net functions to inhibit competitors within the phonological cohort or the semantic field. The system does not require that all networks inhibit all other networks, although initial inhibitory connections are assumed to be fairly dense.

Associations between these five types of information should be sufficient to control language learning and processing within the ConComp model.

The pattern of interconnectivity in the overall lexical net must be such that no single cell needs to maintain connections to all other cells. Instead, it is the overall pattern of connectivity that allows lexical nets to integrate diverse types of information. It is possible that this pattern of connectivity is a neurological given, although it is also possible that it is at least partially an emergent structure. In any case, both this connectivity and the shape of the units of phonological and conceptual input are already fairly well established by the beginning of language learning.

Acquiring the auditory image. What happens during language learning is that the thousands of lexical networks available to the learner all compete in a highly parallel fashion. All of the networks have connections to auditory units, but the strengths of connections to particular auditory patterns vary randomly across nets. When the child attends to an auditory form, many lexical nets will receive some activation. By chance, a few networks will emerge as more responsive than others. The cohort inhibitor connections between networks will tend to allow one network to emerge as the winner for a particular pattern of auditory input. The winning network will start to adapt itself as a recognizer for this sound pattern. When the same word is heard again, this net will have an even clearer advantage and will continue to tune itself to recognize the word in the winner-take-all competition with the other networks.

Coding order within the auditory image. Within the individual lexical nets, phonological receptor units come to respond to input auditory cues in the correct serial order. For example, when acquiring the word "bat", there will be several lexical nets with phonological receptor units that respond to the individual sounds /b/, /a/, and /t/. Initially, the order between units responding to these sounds will not be fixed. However, units will tend to fire soonest for those sounds that come first. The network that does this most effectively will have a processing advantage the next time the word is heard, since it will fire more smoothly and will have a head start on inhibiting its competitors. The more smoothly this process works within a network, the more strongly it will compete with other lexical networks for the correct recognition of the word. For example, a network that has become tuned to recognize "bat" in correct linear sequence will have a quicker activation than one that is tuned to simply recognize the sounds /b/, /a/, and /t/ without a particular linear sequence. The linear ordering of sounds through an activation sequence is similar to the scheme for serial order suggested by Rumelhart and Norman (Rumelhart & Norman, 1982) and Houghton (Houghton, 1990).

Linking to meaning. Just as some of the cells in the network are well connected to auditory units, others are well connected to visual image and conceptual units. When a word is used in a meaningful semantic context, these units fire in association with the auditory units and, through Hebbian learning, become associated. Those networks that make this connection most successfully become most highly activated and are able to inhibit their competitors.

Cohort formation. Initially, the inhibitory connections between networks are randomly distributed. However, as the identity of particular word networks begins to emerge, these connections can be pruned and strengthened. For example, inhibitory connections between the words “cat” and “pull” are probably never needed. These words share no common phonemes and have minimal semantic overlap. Although there are some abstract commonalities between “cat” and “pull”, such as the fact that both are monosyllables and both begin with a voiceless obstruent, these commonalities are not enough to place them in direct competition for a match to actual input. On the other hand, there must be inhibitory connections between phonological neighbors such as “cat” and “cap” or between semantic neighbors such as “axe” and “hatchet.” As these connections become pruned and strengthened, auditory cohorts (Marslen-Wilson, 1987; Walley, 1988) and semantic fields (Warrington & McCarthy, 1987) arise. The training of these inhibitory connections can proceed on the Hebbian model. As in Kawamoto (this volume) and McClelland (McClelland, 1987) top-down preactivation of some of the semantics of a word can further faciliate activation.

Articulatory learning. The learning of the articulatory form of a word is scaffolded by the prior learning of its auditory form. When the word is activated, the auditory units become activated in their correct order. Each auditory unit then stimulates some closely matching articulatory unit. In this way, activation of the articulation for a new word being learned is more or less on track initially. However, full control of the articulation of a word requires direct connections from the lexical network to motor cortex without the mediation of the auditory scaffold. As these direct connections form, the importance of the auditory scaffold diminishes. Of course, there is no direct correspondence between auditory forms and articulatory forms and there is a great deal of learning of coarticulation that occurs without reliance on auditory patterns. However, the auditory skeleton can provide a major scaffold for the acquisition of articulatory form.

7.3. Segmentation

The model of lexical learning given above assumes that lexical items and words come nicely packaged with just the right amount of phonology and meaning. Of course, nothing could be farther from the truth. Consider this example from Garrett (Garrett, 1990):

Remember	a spoken	sentence	often	includes
Ream ember	us poke can	cent tense	off in	ink lewds

many	words	not	intended	to be	heard.
men knee	words	knot	in ten did	tube bee	herd.

In early child language, segmentation involves building up a contrast between lexical items that are already known and material that is new and unfamiliar. For example, if the child already knows the words “Mommy” and “Daddy”, but does not know the word “like” the sentence “Daddy likes Mommy” would be represented in this way:

d a d i	l aI k s	m a m i
Daddy	| unknown |	Mommy

For the known stretches, there are two lexical items that successfully compete. The unknown stretch stimulates lexical learning of the new word “likes”. In this way, the use of the lexicon leads to the correct segmentation of the input and further lexical learning.

The process of segmentation by lexical activation also facilitates the learning of inflections. If the child knows the word “cat”, but not the plural form, the word “cats” can be represented in this way:

k ae t s
cat | unknown

The acquisition of the plural is a subcase of the general case of lexical learning. The fact that productive use of the plural is relatively late, despite its high frequency, indicates that the segmentation process tends to work first on items that have a relatively clearer prosodic and phonological separation and that the extrasyllabic (Stemberger & MacWhinney, 1984) quality of the plural suffix may tend to delay the segmentation process.

7.4. Allomorphy

Our discussion of lexical learning has assumed that each word has a single phonological shape and a single meaning. But this is seldom the case. Words take on different phonological forms for a variety of reasons, including dialect variation, fast speech simplifications, and general phonological processes. For example, the English word "that" is sometimes pronounced as "dat" or even as "dah". Dialect and regional variations of this type affect many words. When listening to one of these variant forms, the system has to treat the new form as a variant of an old form. This is done on the basis of an overall matching procedure. Both "that" and "dat" activate the same semantic and syntactic units. They only differ in terms of one phonological feature. When the learner hears "dat" for the first time, the activation of "that" will be somewhat weaker than usual. However, any attempt to establish a new lexical item in this case will be unsuccessful, because the old item is a good enough match to the input. Unless the mismatch exceeds a certain threshold value, no new lexical network will be recruited. If the forms differed more radically, as in "on" and "onto" the the attempt to set up separate items would be more successful.

The incorporation of multiple phonological forms within a single lexical item also allows the system to pick up allomorphic variation. Consider how the system learns the five allomorphs of the Hungarian plural: -<u>ok</u>, -<u>ak</u>, -<u>ek</u>, -<u>ök</u>, and -<u>k</u>. As in the learning of the English plural suffix, one of these five forms will be segmented out into a new lexical network. If the first analyzed plural is <u>asztalok</u> "tables", the initial form of the suffix will be -<u>ok</u>. If the next analyzed plural is <u>székek</u> "chairs", the resulting form of the suffix will involve the phonological union of -<u>ok</u> and -<u>ek</u>. This union preserves the common phonology of the two suffix forms and their common plural meaning. However, it adds competing values for the features of rounding, height, and fronting of the vowel. When the speaker comes to activating the suffix for new words, the competition between the alternative values of these features must be resolved through a process much like that described by MacWhinney and Leinbach for the English past tense. In the new framework of the ConComp Model, we can refer to this as the process of allomorphic resolution.

7.5. Allomorphic resolution

This scheme for allomorphic extraction brings us full circle back to the back-propagation models for inflection. However, now the input to these models is not a single surface form, but the archisegmental representation proposed by Lamb (Lamb, 1966), Hudson (Hudson, 1980), MacWhinney (1975; 1978), and others. The choice between competing values of the vowel in the Hungarian suffix is now resolved by the same type of back propagation network used in the StemNet branch of the model for German discussed earlier.

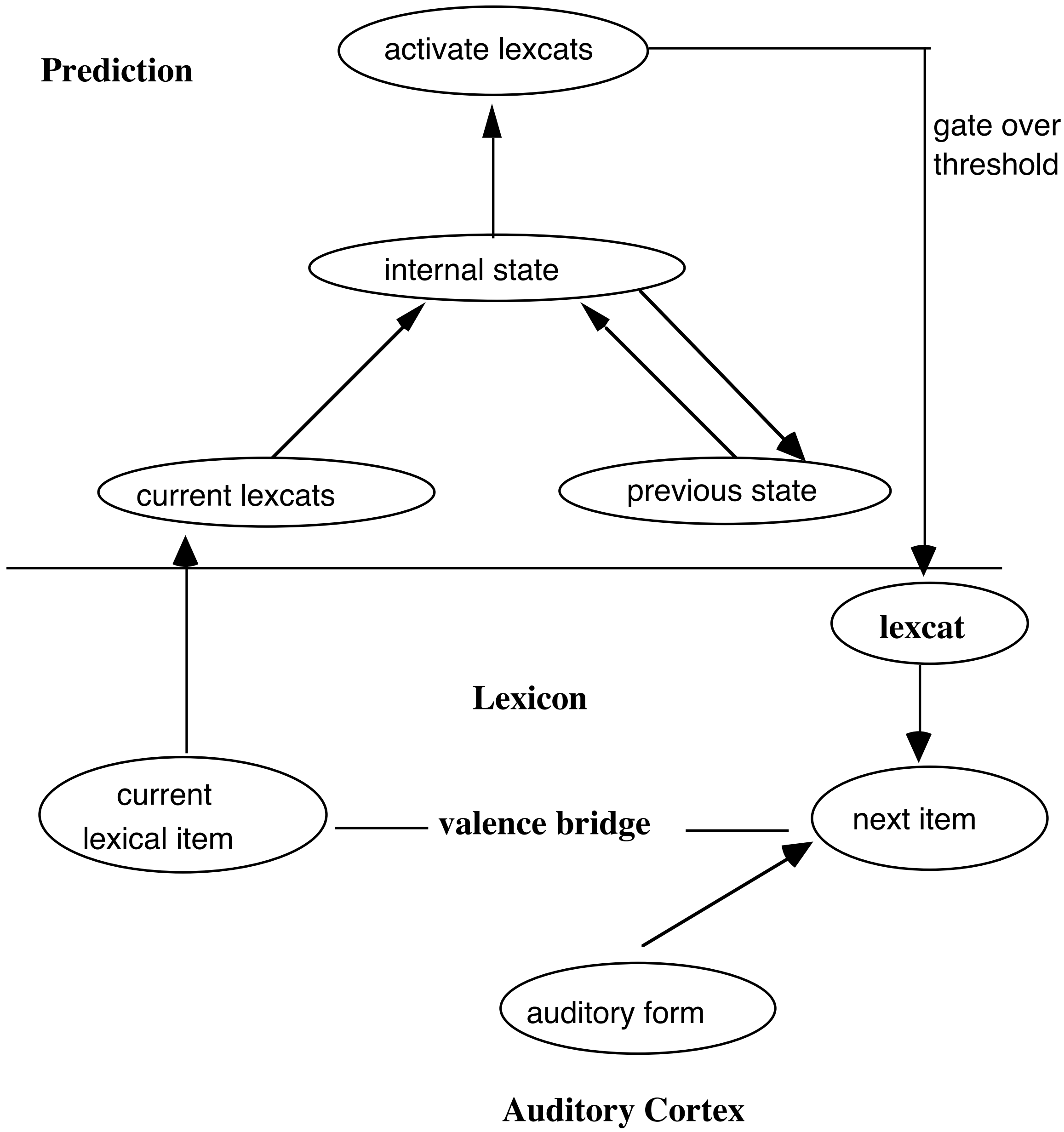
The ConComp model actually provides two sets of cues for allomorphic resolution. In addition to the cues based on the phonological shape of the lexical items being combined, there is the lexical category information. In German, this mechanism is important for extracting gender. In other languages, it can also be used to extract irregular morphological classes. For example, the choice between the plural allomorphs -ok and -ek in Hungarian can be determined almost exclusively on the basis of phonological cues. However, the choice between the two back vowel variants -ok and -ak is far more idiosyncratic. For example, the noun híd "bridge" takes forms like hidak, hidas, and hidam. When applied to Hungarian, the LexCat system will record the cooccurrence of plural -ak with híd, and this can serve as a good cue for producing forms like hidas and hidam. A similar system for controlling irregular paradigms can help in the learning of weak masculine nouns in German such as der Herr - des Herren - die Herren (Klaus-Michael Köpcke, personal communication).

7.5. Polysemy

The system for the learning and resolution of polysemy has the same form as the system for the learning and resolution of allomorphy (MacWhinney, 1985). For example, the contrast between the use of "truck" to refer to a toy and the use of "truck" to refer to a real working vehicle can be coded in terms of competing features and then resolved through a back-propagation network using polysemic cues from related words. The complete framework for this type of processing is discussed in MacWhinney (MacWhinney, 1989).

8. Lexical Control of Processing

We can extend this construction of a connectionist basis for the Competition Model by using lexical networks to control syntactic processing in both comprehension and production. The key to this elaboration is an architecture that uses lexical categories to build "valence bridges" (MacWhinney, 1988; MacWhinney, 1989). Valence bridges are semantic-syntactic connections between slots activated by lexical items and the arguments that fill these slots. For example, in the phrase "another coffee", the word "another" generates a slot for a count noun which is then filled by the noun "coffee". The filling of the slot creates a valence bridge between the two words and a new conceptual unit. For comprehension, the model would look like this:

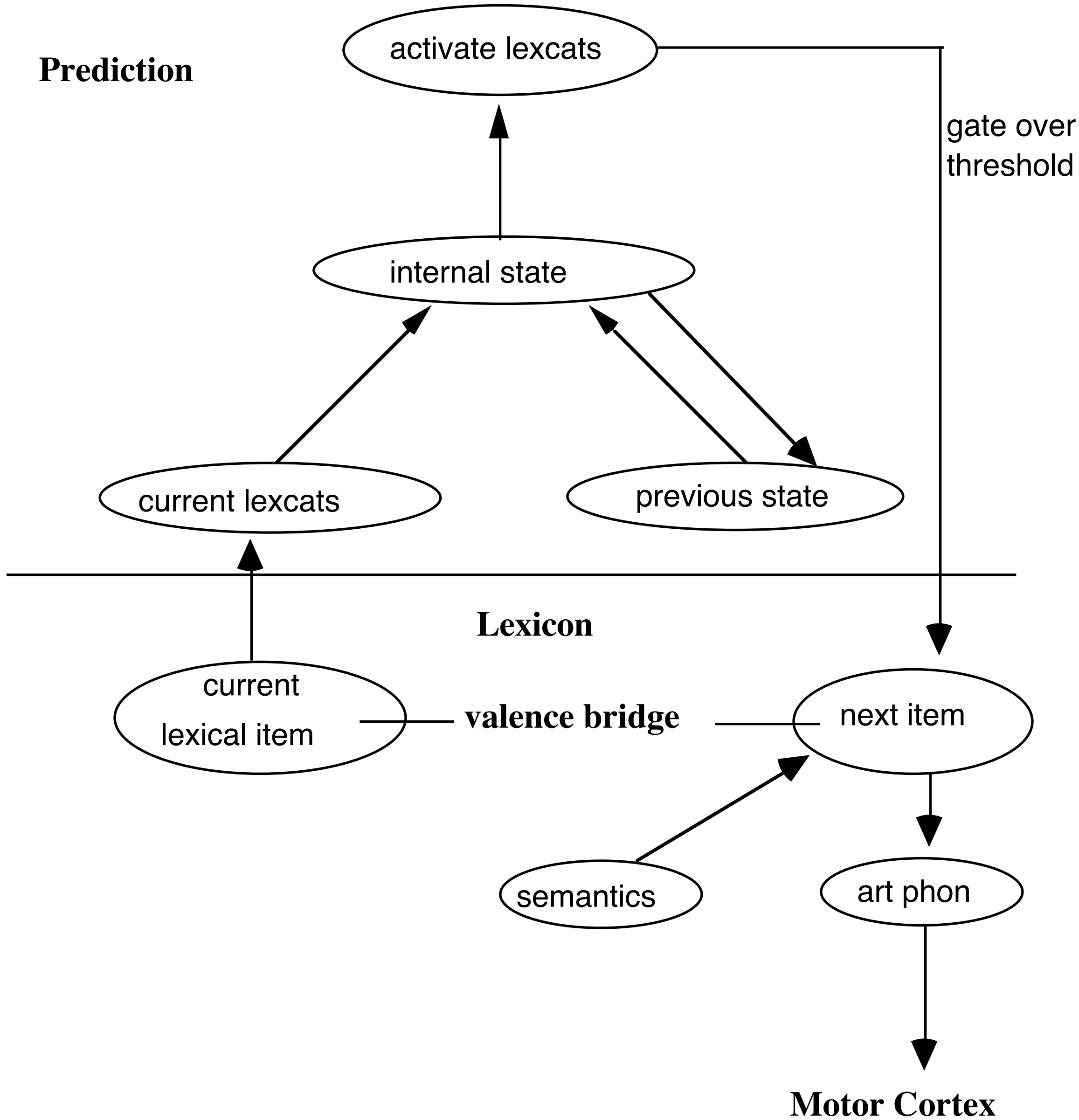


It may be convenient to think of this network as divided into two segments. The segment that is labelled "prediction" is based upon a recurrent network of the type proposed by Elman (Elman, 1990; Elman, 1993). This segment takes the lexical category of the word currently being processed and generates a set of expectations for the lexical categories of the next item. For example, if the current word is the adjective "big", the anticipations for the next word would be for either a noun or an adjective.

Lexical processing uses the syntactic expectations generated by predictive processing to activate large cohorts of words. Then the actual auditory form of incoming material would help to narrow down the competition between lexical items. The output of this process is the formation of a conceptual relation between the previous lexical item and the word which is the new current item. For many sequences of words, no solid valence relation can be established (MacWhinney & Pléh, 1988) and the system needs to store the items in something like a working memory (Baddeley, 1990; Just & Carpenter, 1992). One approach to the working memory problem is suggested by Hausser's (Hausser, 1990) left-associative grammar. In Hausser's system, intermediate category states are complex category strings which are carried along until valence cancelling can be achieved. Hausser has shown that this system processes in linear time using a Markov process that operates on categories, rather than terminal strings. Such a system is highly compatible with Elman's proposals for the construction of internal category states in a recurrent network.

The second aspect of sentence comprehension that must be captured in a network is the parallel computation of alternative interpretations of ambiguous sentences. For example, in a sentence such as "Bill bought her pancakes", the pronoun "her" can be either the indirect object or a possessive pronoun. The proposal advanced by Ford, Bresnan, and Kaplan (Ford, Bresnan, & Kaplan, 1982), MacWhinney (MacWhinney, 1987b), and Hausser (Hausser, 1990) is that the competition between these two readings is controlled through the lexical ambiguity itself. Specifically, the lexical item "her" has two lexical categories that are in competition. Each of these categories can be used to build a set of valence bridges. In this particular case, two full interpretations are built in parallel (Kurtzman, 1985; Taraban & McClelland, 1988). In a disambiguating sentence such as "Bill bought her best pancakes" only one of the chain of valence relations survives. In a garden-path sentence such as "The communist farmers hated died" the weakest reading is suppressed too early and must be resurrected through reprocessing.

Lexical categories can also be used to control sentence production. The network architecture is the same as that involved in comprehension. However, the role of the valence bridge in production is to support the activation of the next item, rather than to construct a relation between the current item and the previous item. Here, again, lexical categories work together with lexical semantics to pick out a candidate word for articulation.



In both comprehension and production the role of the "predictive" recurrent network is to gate the activation of lexical items. The view of Broca's aphasia as leading to a deficit in activation (Milberg, Blumstein, & Dworetzky, 1988) and of Wernicke's aphasia as an excess of activation fits in well with this view of lexical categories as devices for sending the activation of particular lexical items over threshold. Either posterior damage or damage to the connections between anterior and posterior (Damasio & Damasio, 1983) can impede lexical activation in production (Bates, Friederici, & Wulfeck, 1987; MacWhinney & Osman-Sági, 1991) and comprehension (MacWhinney, Osman-Sági, & Slobin, 1991).

9. Networks and the Ring

We began with the history of the demise of the Big, Mean Rules and the dangers of the power of the symbolic Ring. This paper has attempted to chart out a path that will escape from the dangers of hand-wired complexity. But is it possible that this work will simply replace Big, Mean Rules and Big, Mean Flowcharts with Big, Mushy Networks? Is it possible that these Big, Mushy Networks will also fall prey to power of the symbolic Ring? Although we have to always worry about hidden assumptions and hand-crafting in networks, it seems to me that the danger of excessive power in connectionist models is less than the danger of insufficient power and insufficient clarity. Many of the mechanisms proposed here have been tested and are reliable for a small set of problems. However, many are untested and may require significant modification when they are fully implemented. What is important at this point is to show that there are at least some candidate proposals for how a comprehensive connectionist model of language acquisition could conceivably function. The account that I have offered is one such proposal. Will this model or others like it be able to attain the full generativity and scalability of the great symbolic systems? Only time will tell.

References

Anderson, J. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249-277.

Anderson, J. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Needham Heights, MA: Allyn & Bacon.

Barnden, J., & Srinivas, K. (1991). Encoding techniques for complex information structures in connectionist systems. *Connection Science, 3,* 269-313.

Bates, E., Friederici, A., & Wulfeck, B. (1987). Grammatical morphology in aphasia: Evidence from three languages. *Cortex, 23,* 545-574.

Besner, D., Twilley, L., McCann, R., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review, 97,* 432-446.

Bloomfield, L. (1914). *An introduction to the study of language*. New York: Holt, Rinehart & Winston.

Bohannon, N., MacWhinney, B., & Snow, C. (1990). No negative evidence revisited: Beyond learnability or who has to prove what to whom. *Developmental Psychology, 26,* 221-226.

Burgess, N., & Hitch, G. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31,* 429-460.

Chomsky, N. (1959). Review of Skinner's verbal behavior. *Language, 35,* 26-58.

Chomsky, N. (1981). *Lectures on government and binding*. Cinnaminson, NJ: Foris.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Chrisman, L. (1991). Learning recursive distributed representations for holistic computation. *Connection Science, 3,* 345-366.

Cohen, M., & Massaro, D. (1992). On the similarity of categorization models. In F. Ashby (Ed.), *Multidimensional models of perception and cognition* . Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

Cottrell, G. W. (1987). Toward connectionist semantics. *TINLAP, III,* 1-6.

Damasio, A. R., & Damasio, H. (1983). The anatomic basis of pure alexia. *Neurology, 33*(12), 1573-1583.

Daugherty, K., & Seidenberg, M. (1994). Beyond rules and exceptions: A connectionist approach to inflectional morphology. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 353-388). Amsterdam: John Benjamins.

Dell, G. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283-321.

Elman, J. (1990). Finding structure in time. *Cognitive Science, 14,* 179-212.

Elman, J. (1993). Incremental learning, or the importance of starting small. *Cognition, 48,* 71-99.

Engel, U. (1991). *Deutsche Grammatik*. Heidelberg: Julius Groos.

Farah, M., & McClelland, J. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120,* 339-357.

Feldman, H., Holland, A., Kemp, S., & Janosky, J. (1992). Language development after unilateral brain injury. *Brain and Language, 42,* 89-102.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, Mass.: M. I. T. Press.
Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
Ford, M., Bresnan, J., & Kaplan, D. (1982). A competence-based theory of syntactic closure. In J. Bresnan & R. Kaplan (Eds.), *The mental representation of grammatical relations* . Cambridge, MA: MIT Press.
Francis, W., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
Garrett, M. (1990). Sentence processing. In D. Osherson & H. Lasnik (Eds.), *An invitation to cognitive science: Language* . Cambridge, MA: MIT Press.
Gasser, M. (1990). Connectionism and universals of second language acquisition. *Second Language Acquisition, 12,* 179-199.
Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science, 11,* 23-63.
Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93*(46-74).
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.
Hare, M. (1990). The role of similarity in Hungarian vowel harmony: A connectionist account. *Connection Science, 2,* 123-150.
Harley, T., & MacAndrew, S. (1992). Modelling paraphasias in normal and aphasic speech, *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 378-383). Hillsdale, NJ: Lawrence Erlbaum Associates.
Harris, C. (1990). Connectionism and cognitive linguistics. *Connection Science, 2,* 7-33.
Hausser, R. (1989). *Computation of language: An essay on syntax, semantics, and pragmatics in natural man-machine communication*. Berlin: Springer.
Hausser, R. (1990). Principles of computational morphology. *Computational Linguistics,* 47.
Hertz, J., Krogh, A., & Palmer, R. (1991). *Introduction to the theory of neural computation*. New York: Addison-Wesley.
Hinton, G., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*(1), 74-95.
Hoeffner, J. (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* . Hillsdale, NJ: Lawrence Erlbaum.
Hopcroft, J., & Ullman, J. (1979). *Introduction to automata theory, languages, and computation*. Reading, Mass.: Addison-Wesley.
Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79,* 2554-2558.
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.
Hudson, G. (1980). Automatic alternations in non-transformational phonology. *Language, 56,* 94-125.
Jaeger, J. (1980). Testing the psychological reality of phonemes. *Language and Speech, 23,* 233-253.
Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60-99.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.
Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech, 32,* 1-23.
Kim, J., Pinker, S., Prince, A., & Prasada, S. (1990). Why no mere mortal has ever flown out to center field. *Cognitive Science, 15,* 173-218.
Klahr, D., & Kotovsky, K. (1991). *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43,* 59-69.
Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer.
Köpcke, K.-M. (1988). Schemas in German plural formation. *Lingua, 74,* 303-335.
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.
Kruschke, J. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22-44.
Kurtzman, H. (1985). *Studies in syntactic ambiguity resolution.* , Massachusetts Institute of Technology.
Lamb, S. (1966). *Outline of stratificational grammar*. Washington: Georgetown University Press.
Li, P. (1992). *Overgeneralization and recovery: Learning the negative prefixes of English verbs* : Center for Research on Language, UCSD.
Linell, P. (1979). *Psychological reality in phonology: A theoretical study*. Cambridge: Cambridge University Press.
MacWhinney, B. (1975). Rules, rote, and analogy in morphological formations by Hungarian children. *Journal of Child Language, 2,* 65-77.
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. (1985). Hungarian language acquisition as an exemplification of a general model of grammatical development. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition* (pp. 1069-1156). Hillsdale, NJ: Lawrence Erlbaum.
MacWhinney, B. (1987a). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249-308). Hillsdale, NJ: Lawrence Erlbaum.
MacWhinney, B. (1987b). Toward a psycholinguistically plausible parser. In S. Thomason (Ed.), *Proceedings of the Eastern States Conference on Linguistics* . Columbus, Ohio: Ohio State University.
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., & Anderson, J. (1986). The acquisition of grammar. In I. Gopnik & M. Gopnik (Eds.), *From models to modules* (pp. 3-25). Norwood, N.J.: Ablex.
MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 29,* 121-157.
MacWhinney, B., & Osman-Sági, J. (1991). Inflectional marking in Hungarian aphasics. *Brain and Language, 41,* 165-183.

MacWhinney, B., Osman-Sági, J., & Slobin, D. (1991). Sentence comprehension in aphasia in two clear case-marking languages. *Brain and Language, 41,* 234-249.

MacWhinney, B., & Pléh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition, 29,* 95-141.

MacWhinney, B. J., Leinbach, J., Taraban, R., & McDonald, J. L. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28,* 255-277.

Marchman, V. (1992). Constraint on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 4,* 215-234.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71-102.

Massaro, D. (1987). *Speech perception by ear and eye*. Hillsdale, NJ: Lawrence Erlbaum.

McClelland, J., & Elman, J. (1986). Interactive processes in speech perception. In J. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing* (pp. 58-121). Cambridge, MA: MIT Press.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Performance XII* (pp. 41).

McClelland, J. L., & Kawamoto, A. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel Distributed Processing* (pp. 272-376). Cambridge, MA: MIT Press.

McCloskey, M., & Cohen, N. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. Bower (Ed.), *The psychology of learning and motivation: Volume 23* . New York: Academic Press.

Menn, L., Markey, K., Mozer, M., & Lewis, C. (1992). Connectionist modeling and the microstructure of phonological development: A progress report. In B. Boysson-Bardies, P. Jusczyk, J. Morton, & S. de Schonen (Eds.), *Changes in speech and face processing in infancy: A glimpse at developmental mechanisms of cognition* .

Milberg, W., Blumstein, S., & Dworetzky, B. (1988). Phonological processing and lexical access in aphasia. *Brain and Language, 34:2,* 279-293.

Mills, A. E. (1986). *The acquisition of gender: A study of English and German.* Berlin: Springer-Verlag.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models of Human Memory* (pp. 203-248). New York: Academic Press.

Morton, J. (1980). The logogen model and orthographic structure. In U. Frith (Ed.), *Cognitive processes in spelling* . London: Academic Press.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Newmeyer, F. (1980). *Linguistic theory in America*. New York: Academic Press.

Ohala, J. J. (1974a). Phonetic explanation in phonology. In A. Bruck, R. Fox, & M. La Galy (Eds.), *Papers from the parassession on natural phonology* (pp. 251-274). Chicago: Chicago Linguistic Society.

Ohala, M. (1974b). The abstractness controversy: Experimental input from Hindi. *Language, 50,* 225-235.

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

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a Parallel Distributed Processing Model of language acquisition. *Cognition, 29,* 73-193.

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

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by back propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition* . Cambridge, MA: M.I.T. Press.

Rumelhart, D., & Norman, D. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6,* 1-36.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing*. Cambridge, Mass.: 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.

Squire, L. (1987). *Memory and brain*. New York: Oxford University Press.

St. John, M. (1992). The story gestalt: a model of knowledge-intensive processes in text comprehension. *Cognitive Science, 16,* 271-306.

Stemberger, J. (1985). *The lexicon in a model of language production*. New York: Garland.

Stemberger, J., & MacWhinney, B. (1984). Extrasyllabic consonants in CV Phonology: An experimental test. *Journal of Phonetics, 12,* 355-366.

Stemberger, J., & MacWhinney, B. (1988). Are inflected forms stored in the lexicon? In M. Hammond & M. Noonan (Eds.), *Theoretical Morphology* . New York: Academic Press.

Tallal, P., & Stark, R. (1980). Speech perception of language delayed children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology: Perception* (Vol. 2, pp. 155-171). New York: Academic Press.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language, 27,* 597-632.

Taraban, R., McDonald, J. L., & MacWhinney, B. M. (1989). Category learning in a connectionist model: Learning to decline the German definite article. In R. Corrigan, F. Eckman, & M. Noonan (Eds.), *Linguistic categorization* . New York: Benjamins.

Tolkien, J. R. R. (1965). *The lord of the rings*. Boston: Houghton Mifflin.

Touretzky, D. (1990). BoltzCONS: Dynamic symbol structures in a connectionist network. *Artificial Intelligence, 46,* 5-46.

Trammell, R. (1978). The psychological reality of underlying forms and rules for stress. *Journal of Psycholinguistic Research, 7,* 79-94.

Twain, M. (1935). The aweful German language, *The family Mark Twain* . New York: Harper & Brothers.

von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. Shannon & J. McCarthy (Eds.), *Automata studies* . Princeton: Princeton University Press.

Walley, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development, 3,* 137-165.

Warrington, E., & McCarthy, R. (1987). Categories of knowledge: Further fractionations and an attempted integration. *Brain, 110,* 1273-1296.

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.