

University of Pittsburgh ILL



ILLiad TN: 142462

Borrower: PMC

Lending String: *PIT,VWM,VWM,VYF,TAI

Patron:

Journal Title: Research on child language disorders : a decade of progress /

Volume: Issue:

Month/Year: 1990**Pages:** I don't know, but I just want

Article Author: Brian MacWhinney

Article Title: Connectionism as a framework for language acquisition theory

Imprint: Austin, Tex. : Pro-Ed, ©1991.

ILL Number: 184413159



Call #: RJ496 L35R47; 1990

Location: Engineering Library

ODYSSEY ENABLED

Charge

Maxcost: 35.00IFM

Shipping Address:

Carnegie Mellon University-Hunt Library-ILL
Office
4909 Frew St.
Hunt Library Room 305
Pittsburgh, Pennsylvania 15213
United States

Fax: (412)268-6944

Email: es82@andrew.cmu.edu

Research on Child Language Disorders:

A Decade of Progress

*A volume marking the Tenth Anniversary
of the Wisconsin Symposium for
Research on Child Language Disorders*

Edited by

Jon F. Miller, Ph.D.

Professor of Communicative Disorder

University of Wisconsin, Madison

and

Associate Director for Development

Waisman Mental Retardation Research Center

pro·ed

8700 Shoal Creek Boulevard
Austin, Texas 78758

© 1991 by Jon F. Miller

All rights, including that of translation, reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, recording, or otherwise, without the prior written permission of the publisher.

Library of Congress Cataloging in Publication Data

Main entry under title:

Research on child language disorders: a decade of progress / Jon F. Miller, editor.
p. cm.

"A volume marking the 10th anniversary of the Wisconsin Symposium for Research on Child Language Disorders.

Includes bibliographical references.

ISBN 0-89079-408-1

1. Language disorders in children. 2. Language disorders in children—Research. I. Miller, Jon F. II. Wisconsin Symposium on Research in Child Language Disorders.

[DNLM: 1. Language Development. 2. Language Disorders—in infancy & childhood. 3. Research. WL 340 R432]

RJ496.L35R47 1990

618.92'855—dc20

DNLM/DLC

for Library of Congress

89-25120

CIP

Printed in the United States of America

pro·ed

8700 Shoal Creek Boulevard
Austin, Texas 78758

CHAPTER

4

Connectionism as a Framework for Language Acquisition Theory

BRIAN MACWHINNEY

Connectionism is "in." Not since the Dark Ages of the pre-Chomskyan era have we seen so much interest in associationist models of human thinking. Streaming forth from their banishment in the Skinnerian dungeons are dozens of detailed computational models based on the new language of networks, nodes, and connections. At the crest of this wave are the Parallel Distributed Processing (PDP) models of Rumelhart, McClelland, and their colleagues (1986). Proponents of these new models present them as a major challenge to the *ancien régime*—a definitive revolution in the way in which we understand the human mind. Yet revolutions in academia are seldom bloodless. Inevitably, the proponents of the new paradigm tend to overstate their case and, inevitably, the Old Guard tries to form a unified front to challenge the contributions of the newcomers. The ensuing confusion infuriates some, galvanizes others, and perplexes everyone. All of these things are happening now in the Great Debate that is taking place between the New Connectionism and the Classical Model.

From the viewpoint of the onlooker to this controversy, what most characterizes this debate is the amount of heat being generated and the uniform avoidance of compromise positions or mixed models. From the general viewpoint of the human sciences, both the connectionists and their opponents seem to agree on many crucial issues. They agree on the importance of generating precise models of complex phenomena that can

then be matched to empirical data. They agree on the importance of areas such as phonology, morphology, and syntax for the construction of such models. Both groups realize that the problems of one approach are often the strengths of the other. Given this, it is surprising not to see researchers considering ways of integrating the classical and connectionist models. Presumably, the participants in this debate have decided that they could present their points more clearly by assuming a position of stark opposition. From the viewpoint of the history of science, this is useful. However, the consumer of both connectionist and classical views of language processing and acquisition should be warned that a reconciliation of connectionist and classical views may not be as impossible as many authors suggest.

This chapter will present a series of completed connectionist simulations of the acquisitions of morphology in German, English, and Hungarian and will show how these simulations successfully address a variety of problems unsolved in earlier work. Then we will examine connectionist accounts for the learning of word meanings and the processing of syntax. Before looking at the connectionist simulations, it is necessary to review earlier work in this area to understand its successes and its failures.

EARLIER MODELS OF MORPHOLOGICAL LEARNING

Modern investigations of the learning of morphology began with Berko's (1958) famous "wugs" experiment. Berko, Ervin-Tripp, Braine, Brown and others looked in detail at overregularizations such as "feets" or "bringed" as ways of understanding language learning more generally. This early work showed the extent to which language use is based on the productive application of patterns and not mere rote memorization. Although morphology is only a very small part of the general picture of language acquisition, it has some important properties that make it an ideal topic for detailed investigation. Above all, it is extremely easy to collect and quantify data on morphological productions, including both correct forms such as "jumped" and overregularizations such as "falled."

The first serious theoretical account of morphological learning across languages was provided by Slobin (1973). In a masterful overview of both experimental and diary data on the acquisition of dozens of languages, Slobin was able to propose a set of general operating principles that accounted for the most well-documented aspects of word formation by children. Some of the most important data in Slobin's account came from studies of plural and past tense formation in English and from nominal declension in Russian. Among the most crucial of Slobin's principles were those that led the child to:

1. pay attention to the ends of words,
2. realize that the phonological forms of words can be systematically modified,
3. pay attention to the order of words and morphemes,
4. avoid interruptions,
5. mark underlying semantic notions overtly and clearly,
6. avoid exceptions, and
7. try to make semantic sense out of grammatical markers.

From each of these general operating principles, Slobin derived a further set of universals that appeared to be true of the 40 languages for which data was then available. Slobin grounded these principles on fundamental psychological facts and an insightful view of language function.

Despite its widespread appeal, Slobin's synthesis failed to specify in more exact terms the ways in which the principles should interact. Without such specifications, it was difficult to make exact predictions regarding the course of morphological learning. Providing this type of specification was a top priority and the earliest account of this sort was the one worked out by MacWhinney (1978) on the basis of earlier proposals by Braine (1971). Because learning in this mode was based on a process of error correction much like the traditional Hegelian dialectic, the model was called the *dialectic model*. The six types of processing in the model were:

1. *Rote use*. Early production of forms such as "feet" and "fell" was considered to be by rote.
2. *Simple combination*. After analyzing out basic forms of affixes, the child could apply them to form patterns such as "foots" or "falled."
3. *General morphophonemic rules*. The control of vowel harmony patterns, voicing assimilations and other general alternations was based on the applications of fairly surface-oriented transformational rules.
4. *Morpheme-specific rules*. Minor rules were acquired by encoding archi-morphemic alternations on both stems and suffixes. Such rules can be used to form irregulars such as "knives" and "sang."
5. *Paradigms*. When all else failed, the system began to hypothesize general word-formation paradigms.
6. *Analogy*. Paralleling the development of combination in levels 2-5, the child also was able to produce analogies such as "brang" as the past tense of "bring" or "rew" as the past tense of "row."

The dialectic model was successful in accounting for many detailed aspects of the acquisition of morphology in English, German, Hungarian, and Finnish. Indeed, it is still the most complete account available for the actual empirical findings in the learning of morphology. However, the account suffered from a fundamental inability to deal with analogic processing. We will discuss that issue in more detail shortly.

Several years later, Pinker (1984) constructed a revised account of morphological learning, which kept certain aspects of the dialectic model while rejecting others. Pinker's model is interesting for three reasons. First, by basing all learning on paradigm formation and analysis, it achieves a cleaner theoretical structure. Second, Pinker portrays the model as an elaboration of learnability theory—a framework that requires that all models of language learning be demonstrated to converge in finite time on target grammars. Third, Pinker and Prince (1988) refer to the model of Pinker (1984) as a prime example of a well-established account of language learning developed within the classical framework and one that is clearly superior to connectionist alternatives. Given these claims for the value of the model, it is important for us to consider exactly how much of an advance it represents over the dialectic model or over connectionist alternatives. Pinker's model can be summarized in terms of a series of paradigm building operations.

1. *Add columns.* Every time the child encounters a new form of a word with a new meaning, he can add a column to a word-specific paradigm.

		PERSON		
		1	2	3
TENSE	Present	go	go	goes

2. *Add rows.* When the child finds a new form of an old word that expresses a new dimension, he must add a new dimension to the paradigm. For example, a one-way paradigm can be made into a two-way paradigm by adding rows. This can also be done for dimensions such as phonological patterns. In general any relevant cue can be used as the basis for setting up new columns or rows.

		PERSON		
		1	2	3
TENSE	Present	go	go	goes
	Past	went	went	went

3. *Delete rows and columns.* If all of the cells in a given row or column have the same entries, eliminate the distinction.
4. *Find the stem.* The child is supposed to examine all the forms in the paradigm and extract the common phonetic material as the stem. This means that the stem will be the "least common denominator" rather than the "greatest common denominator." In this paradigm for the Latin noun *puella* "girl," the stem is *puell-* in the Pinker account.

	<i>Singular</i>	<i>Plural</i>
Nom	puella	puellāe
Acc	puellam	puellas
Gen	puellae	puellarum
Dat	puellae	puellis
Abl	puellā	puellis

5. *Create general paradigms from the affixes.* After extracting the stem from a word-specific paradigm, the remaining affixes constitute the general paradigm. For Latin first declension, the paradigm is

	<i>Singular</i>	<i>Plural</i>
Nom	-a	-ae
Acc	-am	-ās
Gen	-ae	-ārum
Dat	-ae	-īs
Abl	-a	-īs

For agglutinative forms such as Hungarian *ablak-od-nál* "window-yours-near," Pinker provides a recursive extraction procedure that would enter the suffix *-od* into the paradigm for the person suffixes and *-nál* into the paradigm for the case suffixes, while also constructing a word structure template of the form: *stem + person + case*.

6. *Use new forms to split old paradigms.* If there are any irregular forms or minor patterns in the language, repeated application of 4 and 5 can be used to split the general paradigm along "arbitrary" dimensions such as noun gender or verb conjugation in Indo-European. Alternatively, repeated application of 4 and 5 can be used to focus the child's attention on some otherwise ignored syntactic or discourse dimension not encoded in word-specific paradigms.
7. *Use the general paradigm to fill any remaining gaps in the word-specific paradigms.* If a particular word-specific paradigm is missing an entry, the child can use the general paradigm to fill that gap. For example, the child may not know the past tense of "fall." The general paradigm would tell him or her that it is "falled." Like MacWhinney, Pinker allows for a competition between the word-specific form and the general form.

Let us examine some of the technical problems that arise in Pinker's framework. Some of these problems also occur in MacWhinney's system, whereas others only arise in the Pinker account.

1. *New word or new cell?* When a child hears a new word, how does he or she decide whether it should be added to an existing paradigm or simply

entered as a new word into the lexicon? On the one hand, a child might decide not to related *went* to *go*, assuming they are too discrepant phonologically. On the other hand, the child might attempt to form a paradigm for words such as *thump*, *dump*, and *bump*. There are no sure principles in Pinker's system to guide the child through the Scylla and Charybdis of correct suppletion and erroneous paradigm formation. Since, in Pinker's system, word analysis cannot begin until general paradigms are formed, there is no way for the child to find stems in inflected words. Of course, a child could use meaning communalities to guide paradigm construction, but this is not a part of Pinker's proposal. The dialectic model has no problem with forms such as *go* and *went*, since it never puts them into paradigms in the first place. Later on, if it constructs unnecessary rules for alternations that do not really occur, those rules will simply be so weak that they will not survive.

2. *How does the child know which values belong to which dimensions?* In order to add a new value to a row or column, the child must know that the value belongs to a particular dimension. In order to add the value of dual to the dimension of number, the child has to understand that a given noun cannot be both plural and dual at the same time. In other words, the child must have already constructed the conceptual framework underlying the paradigm. Some contrasts, such as that between singular and plural, are obvious. Others, such as that between conditional and past or plural and dual are not so clearly nonoverlapping. Both the Pinker proposal and the MacWhinney proposal share the problem of basing paradigm construction on precisely the knowledge that is being constructed. However, the problem is worse for Pinker since his entire model relies on paradigm formation and MacWhinney only uses paradigm formation when all other forms of learning have failed.

3. *Should stems be reduced to the least common material?* As Braine (1987) has shown, Pinker's procedure runs into serious problems with strong morphophonological alternations. For example, for third declension nouns like *miles* "soldier" the procedure would analyze out *mīl*- instead of the correct form *mīl(it)*-. In regard to the suffix, the procedure would yield a large array of inconsistent suffixes such as *-item*, *-inem*, and the like, all with stray pieces of the stem. In addition, there would be different ending forms for each of the different declensional types. In general, the Pinker solution adds a great deal of complexity and nonuniformity that is avoided in the proposals of MacWhinney (1978) and Braine (1987).

4. *How many words are needed to support a general paradigm?* Pinker tells us that new paradigms will not be set up when only a "small" number of forms are involved. Doing this leads to serious problems for irregular words in agglutinative languages. For example, Hungarian has several irregular noun groups with fewer than six members. According to the Pinker procedure, none of these groups could produce general paradigms. Within

the three large gender classes in German, there are about 20 further classes that show minor irregularities in the genitive, the dative plural, and elsewhere. What leads the child to treat these alternations as secondary and the major gender divisions as primary? Indeed, what evidence is there that minor rules and minor classes are treated in a fundamentally different way from major rules and major classes? Indeed, data from language history, language acquisition, and adult speech errors show that even the smallest paradigmatic groups have some productivity. Indeed, Malkiel (1968) showed that the single strong verb *dicere* was able to play a major role in the shaping of the development of verbal conjugations in Romance.

5. *Should paradigms be used as the basis for extracting affixes?* The problem here is that Hungarian children pull out affixes from simple forms before they construct full paradigms. For example, they extract the first person singular definite *-om* from verbs such as *tudom* long before they construct the full six dimensional verb paradigm out of forms such as *tud-hat-gat-ná-tok* "know-potential frequentative-conditional-2PL". This is only a problem for Pinker's system, since the MacWhinney system orders analysis before paradigm formation.

6. *Learnability.* Despite his interest in learnability, Pinker provides neither learnability proofs for the model nor computational implementations to test the model's operation.

By attempting to provide a uniform treatment based on paradigms, Pinker has expanded on the most questionable part of the dialectic model and abandoned the part that was most fully motivated empirically.

Both MacWhinney and Pinker tried to emphasize the ways in which the child uses real semantic and phonological cues to predict morphological patterns. A rather different emphasis can be found in the work of Maratsos (1982). Maratsos analogized the learning of language to the learning of arbitrary patterns of etiquette. Why is the knife placed on the right of the plate and the fork on the left? According to Maratsos, there is no inherent reason. Why in German is *das Messer* "the knife" neuter, *die Gabel* "the fork" feminine, and *der Löffel* "the spoon" masculine? According to Maratsos, there is no good reason. Rather, language is arbitrary and language learning involves picking up arbitrary co-occurrence patterns between unmotivated form classes. The mechanism that Maratsos proposed to acquire these co-occurrences was one that would simply form classes out of things that go with other things. For example, in German, masculine nouns would be defined as ones which co-occur with *der*, *den*, and *dem*. Neuter nouns are defined as those that co-occur with *die* and *der*, and so on. Other categories are defined in similar ways. The nominative is the category of nouns occurring before verbs, with either *der*, *die*, or *das* and with no nominal endings and no preceding prepositions, and so on. Maratsos never stated exactly how these co-occurrences would be detected in a model.

These earlier models all had their strengths and their limitations. The Slobin account was empirically accurate, but inadequate mechanistically. The MacWhinney account was empirically accurate, but its mechanistic components were rather diverse. The Pinker account had greater symmetry, but paid a price in empirical adequacy. Maratsos makes many good points in emphasizing the arbitrariness of many aspects of language learning. However, his view fails to give enough attention to the other side of the coin—the predictable aspects of language patterns. Even for the seemingly clearcut example of German gender, it turns out that there is a great deal more predictability than Maratsos recognized (Köpcke & Zubin, 1983, 1984).

A CONNECTIONIST MODEL OF MORPHOLOGICAL COMPETITION

All of the accounts examined here suffer from an inability to express both arbitrariness and predictability within a single comprehensive framework. The model examined next goes beyond these limitations. It is a connectionist network that is able to learn the basic properties of the German declensional paradigm. This model is a direct outgrowth of a variety of work within the competition model of MacWhinney and Bates (1989). Earlier reports on this line of research can be found in Taraban, McDonald, and MacWhinney (in press) and MacWhinney, Leinbach, Taraban, and McDonald (1989).

Before we look at the simulation itself, we need to review the way in which declensional facts are marked in German. The declensional paradigm is configured around the dimensions or number, case, and gender. Number is either singular or plural. Case is either nominative, accusative, genitive, or dative. Gender is either masculine, feminine, or neuter. The bulk of the work of marking gender, number, and case is done by the article or adjective that precedes the noun. A complete cross of the categories of gender, number, and case would yield 24 possible cells for the full declensional paradigm for the definite article. Fortunately for the German child, gender distinctions for the definite article disappear in the plural, reducing the paradigm to the 16 cells shown below.

	<i>Singular</i>			<i>Plural</i>
	M	F	N	
Nom.	der	die	das	die
Acc.	den	die	das	die
Gen.	des	der	des	der
Dat.	dem	der	dem	den

THE ACQUISITIONAL DATA

The two most comprehensive experimental studies of the learning of German declension are those done by MacWhinney (1978) and Mills (1986). The findings of these studies match well with nonexperimental observations from Park (1981) and the various other sources cited in MacWhinney (1978) and Mills (1986). Some of the most important findings of this literature are:

1. *Early acquisition of the nominative.* Children first achieve correct mastery of the use of the nominative case, often overgeneralizing it for the accusative (MacWhinney, 1978).
2. *Delayed acquisition of the genitive.* Of the four cases, it is the genitive that continues to cause problems for article marking. The dative plural is also a late difficult form, but this difficulty involves nominal marking rather than article selection.
3. *Children often omit the article.* Many of the cues to gender assignment are hard to detect and many are only imperfectly reliable. This forces the child to turn his or her attention to other ways of controlling gender categorization. One simple way of solving the problem is to omit the article. In fact, early on, omission of the article is very common and even later on, the article may be omitted when the child is in doubt about the correct gender assignment.
4. *Children often overgeneralize one gender.* Mills (1986) observed a tendency to overgeneralize the use of the feminine gender.
5. *Children make early use of the highly frequent -e cue.* Mills (1986) examined the role of some of the Köpcke-Zubin cues in the acquisition of German gender and found evidence for their use. MacWhinney (1978) conducted his work before the Köpcke-Zubin cues were available, but his experiment still included some of the cues. Both Mills (1986) and MacWhinney (1978) found early acquisition of the most highly available and reliable of the cues—the presence of final -e as a cue to feminine gender.
6. *Children make early use of highly reliable cues.* MacWhinney (1978) also found that children between the ages of 4 and 6 were able to make correct use of the morphological marking -ei as a cue to feminine gender and -chen as a cue to neuter gender. Schneuwly (1978) reports similar findings. These data indicate that children are indeed sensitive to the various phonological and morphological cues to gender and that the stronger these cues are, the earlier they are used consistently by children. Tucker, Lambert, and Rigault (1977) report on a set of careful and detailed studies of cue use in predicting French gender, which make it entirely clear that the higher the reliability of a cue, the stronger its use by adult subjects.

7. *Children can use paradigmatic marking cues to infer word classes.* MacWhinney (1978) showed that 4-year-old children were able to make reliable use of the pronoun as a cue to the gender of nonce words. The experiment involved using the masculine form of the accusative personal pronoun "him" *ihn* to refer to a nonce word represented by a small toy. When the experimenter said, "I am picking him (*ihn*) up in my hand," children were able to successfully infer that the thing being picked up was masculine even though it was an object they had never seen before with a name they had never heard before.

THE SIMULATION

The behavior that the model is designed to simulate is the selection of one of the six forms of the definite article given a noun and its case context. This task is clearly a production task and not a comprehension task. We will discuss comprehension-production relations later. The simulation relies on the "back-propagation" architecture elaborated by Rumelhart, Hinton, and Williams (1986). Like other connectionist models, the model consists of a large number of densely interconnected "units" or "nodes" operating in parallel. The model has three layers of units: input units, output units, and intervening units, as can be seen in Figure 4-1. The network's "knowledge" is contained in the strength of the connections between the units in the network. Nodes in the network can receive or send activation or both. Activation is sent across connections. Receiving nodes update their activation as a function of the sum of their inputs. Each input is the product of the activation of the sending node times the strength of the connection. The input layer encodes the presence or absence of cues associated with a particular noun and its sentential context. Each node on the input layer represents a single cue feature. If the cue is present for a particular noun, the input node is fully activated, and if it is not present the node remains off. The words are represented as sets of cues. The activation of the input layer produces activation on the internal layer(s), which in turn produces activation on the output layer. Each of the six German definite articles is represented by a unit on the output layer.

Input Units

The model uses a uniform coding of the phonological shape of input words. As indicated in Figure 4-1 above, there are three types of input units: 143 phonological units, 5 semantic units, and 20 case context units. The 143 phonological units represent the full form of the noun in actual phonological features. These units are distributed over 13 slots with 11 features in each slot. The 11 features are standard phonological distinctive features such as [+labial], [+coronal], [+voice], [+high], etc. Diphthongs and affri-

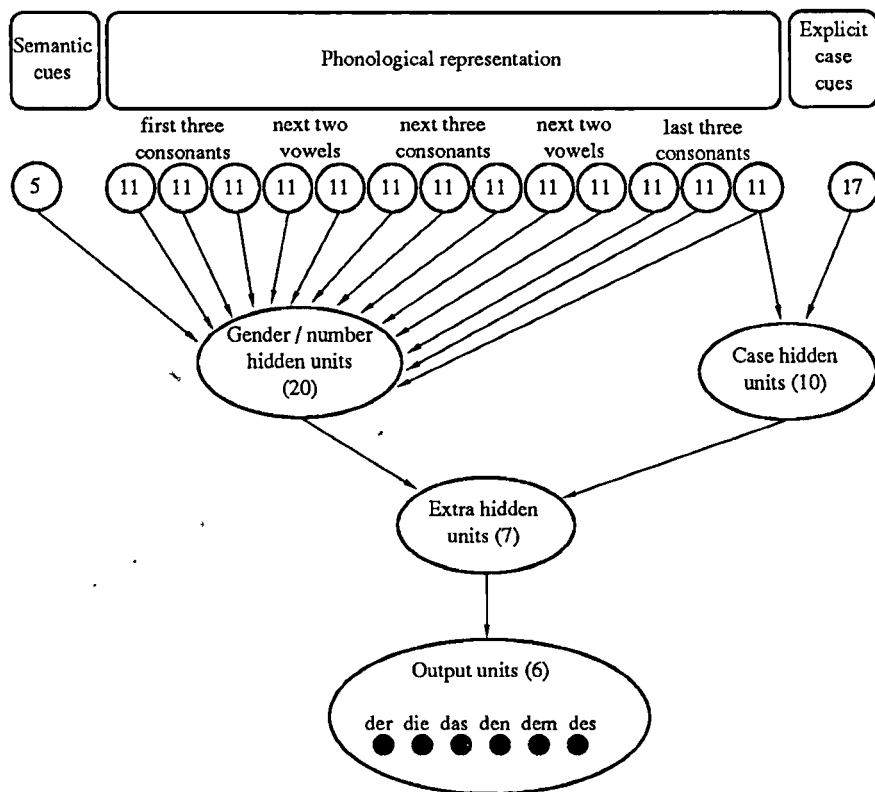


Fig. 4-1. The architecture for the German simulation.

cates are coded as pairs of phonemes. These features provide a unique 11-unit feature code for each German phoneme. The 13 slots are divided across various syllable positions. The 20 case context units code for seven prepositions, seven word order configurations (NNV, NVN, VNN, NN, first noun, second noun, third noun), and three verb types (verb of motion, copular verb, and plural verb). In addition, the hidden units that connect to the case context units also connect to the final phoneme of the word, allowing them to detect the presence of the case endings *-s* for masculine and neuter nouns in the genitive and *-n* for plural nouns in the dative.

The phonological and semantic units all project to a set of gender/number hidden units. The explicit case units and the phonological units for the final consonant of the stem all project to a set of case hidden units. The phonological information for the final consonant is used to code the presence of noun-final markers for the genitive and dative plural. Both sets

of hidden units project to a third set of hidden units, which then activates the six output units.

Running the Model

Here is how the network was taught to perform the task. Initially, all the weights on the connections were assigned small random weights. A training set for the simulation was then presented to the network. The training set consisted of sets of cues for each word in the list and the correct article for that set of cues. During the training phase, the cues for each word were presented on the input layer and activated an output pattern. The activated output pattern was compared to the correct pattern and the difference between the two was used to compute an error measure. After a complete pass through all the words in the training set (an epoch), each weight in the network was individually strengthened or weakened so that during the next pass through the training set the activated patterns would be closer to the correct patterns (i.e., there would be less error). Each weight was changed according to the back propagation algorithm. The learning method was consistently applied to all the connections in the network, and there was no *ad hoc* intervention into the learning process.

The Training Set

The training set consisted of 102 different nouns selected from a frequency count of a spoken German corpus of over 80,000 words (Wängler, 1963). The higher frequency nouns were included several times for a total of 305 tokens in the training set. The network learned the training set completely by the end of 100 epochs of training.

GENERALIZATION ACROSS THE PARADIGM

Although the speed and accuracy of this learning is impressive, one could argue that the network simply developed a complicated rote-like representation of the data presented to it without really acquiring anything that corresponds to rule-like behavior. In order to see if the network had learned something beyond the specific associations between combinations of cues and definite articles present in the training set, two different tests of generalization were made. The first test checked how well the network was able to apply the declensional paradigm. The test set consisted of the same 102 nouns used in the training set, but each noun was paired with the subset of case contexts it had not been paired with in the training set. For example, a word that had occurred only in the dative singular in training was then tested in the nominative plural and the genitive singular.

This test was given to the network after it had achieved 100% performance on the training set. The results of this test were excellent. On the five generalization runs, the model had an average success rate of 94%. The chance level here would be 16%. This high level of performance provides strong evidence that the network was remarkably good at generalizing the overall paradigm to noun-case pairings that it was seeing for the first time. Many of the errors that the network made were caused by ambiguities in the paradigm. These are the kinds of errors children are expected to make. For example, if a noun occurred in the training set in the nominative *die*, the accusative *die*, and the genitive *der*, but did not occur in the dative, neither a child nor the network could know whether the noun stem was a feminine singular or a plural, since the plural takes the same articles as the feminine singular in these three cases. The ambiguity is even more confounded, since one of the most frequent cues to feminine, final *-e* on the noun, is also a plural marker. When a noun with final *-e* was presented in the dative case in the test, the network most often assigned it the article *der*—the marker of a feminine singular noun in the dative. Thus, plural nouns, which should take the article *den* in the dative, were sometimes assigned the incorrect article in the dative case in the generalization test. Another case of ambiguity in the training set occurred when a noun appeared in the dative case with *dem* and the genitive case with *des*, but did not occur in the nominative or accusative cases. In this situation it would be impossible to discriminate masculine singular from neuter singular nouns. Because of this overlap in the paradigm between masculine and neuter singular, the network often confused or conflated the two.

PREDICTION OF THE GENDER OF NEW NOUNS

In order to test the ability of the network to predict the gender of new nouns, the next simulation used a much larger input set. This set included all the 2095 high frequency German nouns in the Wängler corpus. From these, 199 nouns were picked at random and reserved for generalization testing. The remaining 1896 nouns constituted the training set. For this test the architecture of the system was simplified in various ways. The five semantic cues were eliminated, as well as the dimensions of case and number from the simulation, using only the articles *der*, *die*, and *das* of the nominative. This simplified architecture allowed the ability of the network to acquire cues for predicting gender to be seen more clearly. The results were exactly as expected. The simulation was able to predict the gender of new nouns with over 70% accuracy. Since there are only three articles used in this simulation, chance is 33%. The nouns for which the simulation chose the wrong gender were generally ones that resembled patterns of another gender. The model was not expected to achieve perfect or even near perfect perform-

ance in this task, since even native German speakers cannot achieve perfect accuracy in predicting the gender of new words. However, the strong performance of the model on this very large data set in this simplified architecture indicates that there are indeed many powerful cues to the prediction of German gender and that a connectionist network is a good tool for picking up these cues.

COMPARISON TO THE DEVELOPMENTAL LITERATURE

Finally, the model's errors at various points were examined during its learning of the training set. The results from these further analyses uniformly matched the first six major phenomena noted in the developmental literature. Early acquisition of the nominative, delayed acquisition of the genitive, omissions of the article, overgeneralizations of the feminine gender, early use of the strong *-e* cue, early use of the reliable *-chen* cues and others like it were found. The earlier simulation had already shown the model's ability to use the "paradigm" to infer word classes. This last finding is perhaps the most remarkable, given the fact that there was no direct representation of a paradigm anywhere in the model. Although the model used no formal inferential logic, it was able to behave as if it were making this inference.

The model provides an interesting alternative to the information-processing account of morphological learning presented first in MacWhinney (1978) and later in Pinker (1984). Within a single network, the processes of rote, combination, analogy, and paradigm application are all expressed in terms of patterns of associations between cues. The ad hoc nature of the processes proposed in the earlier accounts is entirely eliminated. Whereas earlier research on morphological systems such as that of Tucker, Lambert, and Rigault (1977) or MacWhinney (1978) was forced to think of generalization in terms of rule use, generalization can now be thought about in terms of cue acquisition. The model also allows merging of the insights of the co-occurrence model of Maratsos and Chalkley (1980) with the cue-based learning emphasized in the competition model (MacWhinney, 1987). Within a single network can be found prediction of form class on the basis of both co-occurrences and cues. The network was able to deal successfully with both arbitrary relations and cue-based predictable relations.

Finally—and this is no small matter—this is the first real operational simulation of morphological learning that has ever been fully completed.

GENERALIZING THE MODEL

The architecture used in the simulation discussed so far was designed to model a very specific aspect of German language production. Given a noun

and its case context, the simulation could activate the correct form of the definite article. Using a slightly different type of architecture, it is possible to simulate the processes of production and comprehension within a single network. To explore this possibility, the next study focused on the learning of nominal case marking in Hungarian. A set of 92 Hungarian nouns, each having ten different declined forms, was chosen. These ten different cases are what Hungarian grammarians call *ragok*. They include the accusative, the plural, the inessive, the dative, the benefactive, various possessive forms, and so on. Each inflected form is produced by combining a nominal stem with an affix. During the process of suffixation, both stem and suffix can undergo a variety of transformations that are described in detail in MacWhinney (1978, 1985). For example, when the stem *bokor* ("bush") combines with the suffix *-ok*, the resultant form is *bokrok*. The actual shape of the suffix varies depending on the phonological shape of the stem. For example, the various possible forms of the plural suffix include *-k*, *-ok*, *-ak*, *-ek* and *-ök*. Here is a very small piece of the Hungarian nominal paradigm.

<i>Nom-Sg</i>	<i>Nom-Pl</i>	<i>1PSposs-Acc</i>	<i>Acc</i>	<i>Allative</i>	<i>Super</i>
<i>ló</i>	<i>lovak</i>	<i>lovamat</i>	<i>lovat</i>	<i>lóhoz</i>	<i>lónon</i>
<i>ablak</i>	<i>ablakok</i>	<i>ablakomat</i>	<i>ablakot</i>	<i>ablakhoz</i>	<i>ablakon</i>
<i>madár</i>	<i>madarak</i>	<i>madaramat</i>	<i>madarat</i>	<i>madárhoz</i>	<i>madáron</i>
<i>epér</i>	<i>eprek</i>	<i>epremet</i>	<i>epret</i>	<i>epérhez</i>	<i>epéren</i>
<i>ötös</i>	<i>ötösök</i>	<i>ötösömet</i>	<i>ötöst</i>	<i>ötöshöz</i>	<i>ötösön</i>

The complete paradigm has about 30 rows and at least 96 columns. Our current simulation sampled only 10 of the columns and 8 of the rows.

The goals of the simulation were (1) to model production by producing the phonetics of a declined form, given the semantics of the noun and the semantics of the desired declination, (2) to model comprehension by generating the semantics of a noun and the semantics of its declination type, given the phonetics of its declined form, and (3) to be able to improve performance on both of these tasks through training on the *other*. All of these goals were achieved with surprising success.

The structure of this network, as with the German simulation, was quite simple. Three layers of units were used—an input layer, an intermediate "processing" layer, and an output layer. Training was again carried out using the back propagation algorithm described above. In this simulation, however, both the input and output layers were used to represent the same thing—a semantics/phonetics pair that fully described the declined form of a noun, as indicated in Figure 4-2. The semantics of a declined form were represented by a random unique pattern specific to the noun, followed by a pattern across a set of 14 meaningful units which together expressed the meaning of a declination type. The phonetics of a declined form were simply represented as a string of the phonemes of that form, each phoneme

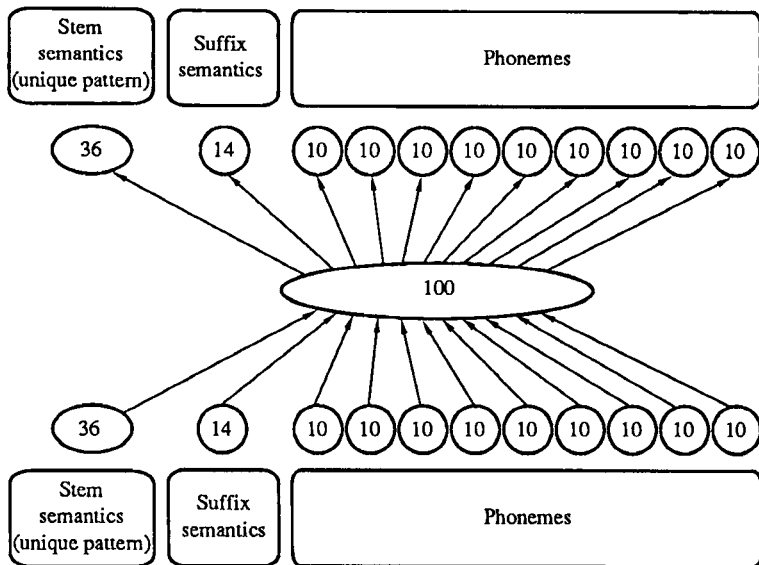


Fig. 4-2. An architecture that relates comprehension and production.

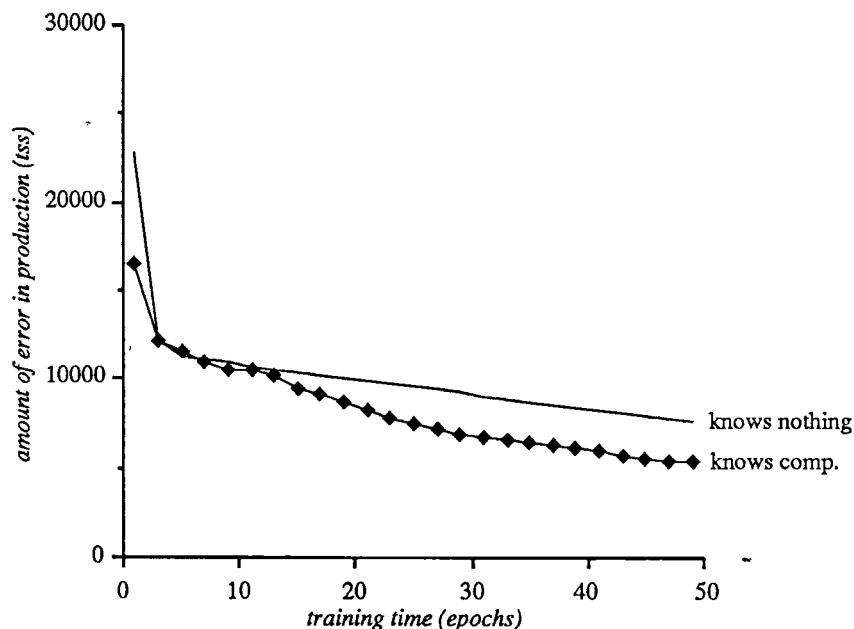
being represented by a pattern across 10 feature units that together specified the desired phoneme.

Training was conducted in two modes. During comprehension training, the network was taught to generate a complete semantic and phonological description, given just the phonological form. The idea here is that children can often infer the meanings of words from context. They hear the word and can then check to see if what they thought the word might mean actually corresponds to what the context indicates. During production training, the network was taught to generate a complete semantic and phonological description of a declined form, given just the semantics of that form. One third of the nouns were randomly selected to have one of their declined forms (randomly chosen) excluded from training (both comprehension and production). The ability to produce and comprehend these unseen forms was then tested, once the network had learned correctly to produce and comprehend all of the forms it was trained with. The purpose of this test was to ensure that the network was actually using appropriate rules of comprehension and production, and not simply memorizing which output patterns went with each input patterns. The results indicated that the network had in fact learned some excellent rules for these tasks. When asked to comprehend the untaught forms, the resulting semantic descriptions of the forms were better than 99% correct

in terms of the stem semantics and 100% correct in terms of the declination-type semantics. When asked to produce the unseen forms, the resulting phoneme strings were 98.3% correct.

The network also demonstrated an interesting interplay between comprehension and production. The observation that the language learner is better prepared to produce a word if he or she has already learned to comprehend it was also captured by this network, as hoped. This was made possible by the "full-description" word representations that were used as output targets. These representations forced the network to learn not only how to produce semantics from phonology and phonology from semantics, but also to reproduce semantics from semantics and phonology from phonology. In doing so, each task generated middle-level representations that were useful to the other task. The comprehension task generated middle-level representations that were useful for producing phonological output (generally useful for performing production), and the production task generated middle-level representations useful for producing semantic output (generally useful for performing comprehension). This meant that, if the network had already learned one of these tasks, when attempting to learn the other, it could exploit this previously learned ability. This is apparently what the network did. Figure 4-3 shows the network's ability to

Fig. 4-3. Savings from prior comprehension training.



remove error from its production performance over time, in each of two conditions. Clearly it is much more adept at learning the production task if it has already learned to comprehend the words it is trying to produce.

This architecture could also be applied to the German declination system, providing a much more robust consideration of the many processes involved there. Such a project and a project involving the declination of English verbs into various tenses are currently underway. These projects will all use a uniform phonological architecture, a uniform network topology, and a uniform learning rule to model morphological learning in these very different languages.

MODELS IN OTHER DOMAINS

The next three sections examine ways of constructing connectionist models for phonological, lexical, and syntactic competition. These proposals are still largely speculative, although some simple simulations in these areas have already been constructed. The reader needs to remember that models of morphological processing are inherently easier to develop than models of semantic processing. Morphology is a small, tightly defined domain for which empirical data are relatively easy to obtain. The development and processing of word meaning, on the other hand, is a much larger area without the same sharp data and tight definitions. Despite these practical differences, there are reasons to believe that the same connectionist concepts developed for the study of morphological development will also be useful in studying learning in these other domains.

MODELING PHONOLOGICAL COMPETITION

McClelland and Elman (1986) have constructed a connectionist account of phonological processing called the trace model. However, that model did not have a learning component and is not well adapted to use in language acquisition studies. However, the architecture being used in our current simulations for morphology can also be adopted to the study of phonological development, if certain modifications are made. The first modification is the replacement of semantic features in the network with phonological features. Thus, instead of trying to go from phonological features to semantic features, the network will try to go from phonetic features to words as characterized by phonological features and then back to phonetic features. In this way, the network produces a unique phonological code that could be used for further connectionist processing without having to compute at the same time the largely arbitrary mapping from sound to meaning. Figure 4-4 indicates the possible shape of such a model.

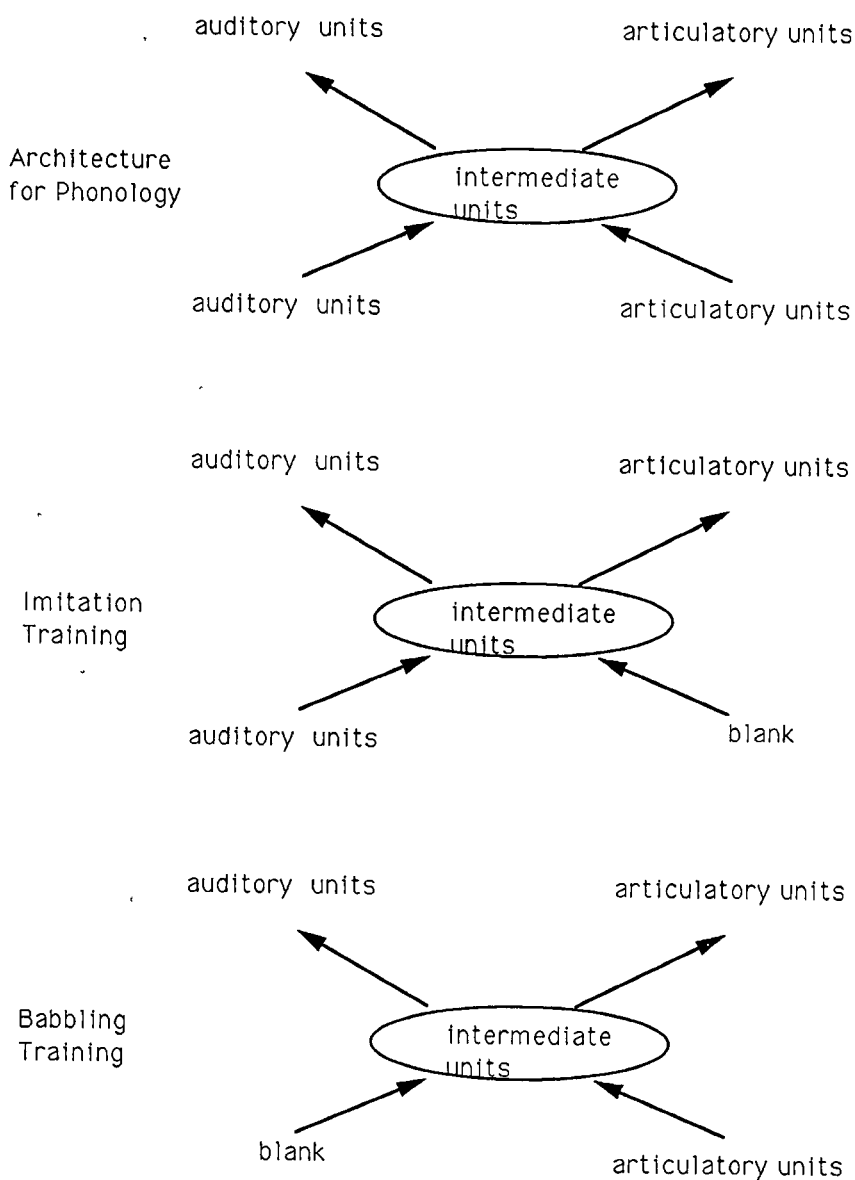
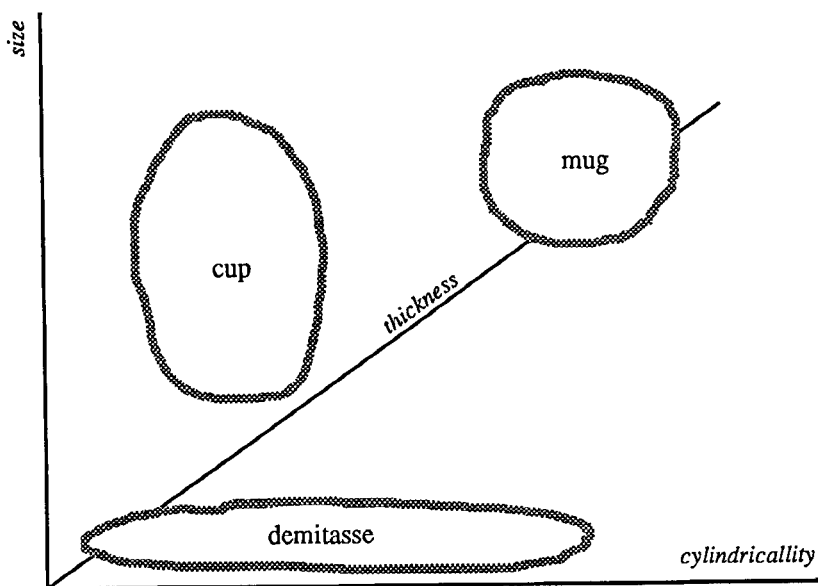


Fig. 4-4. A model for phonological learning.

MODELING LEXICAL COMPETITION

The most basic type of semantic competition is lexical competition. This type of competition arises during production when one is trying to decide what word to use to refer to a particular object or activity. Consider a set of competing words like "cup," "mug," and "demitasse." In the competition model account, these three forms are seen as occupying neighboring parts of a multidimensional semantic topography. For simplicity, imagine that the crucial attributes distinguishing these three forms are size, thickness, and cylindricity. Figure 4-5 illustrates the core semantic territory for each of the three words on these three dimensions. For the adult, objects that fall within the core territories are clear cases of cups, mugs, and demitasses. Objects that fall outside the core will be attracted to one of the three neighboring semantic clusters depending on a feature-weighting algorithm. The closer they are to strong cues of a particular neighbor, the more likely they are to be pulled into the semantic influence of that neighbor. Thus the cue of cylindricity can be in competition with the cue of size for a smallish cylindrical object. Although most demitasses are not cylindrical and most mugs are, the size cue would probably win over the shape cue for most adults. As a result, a very small cup would be called a demitasse, even if it is cylindrical.

Fig. 4-5. Semantic space for "cup," "mug," and "demitasse."



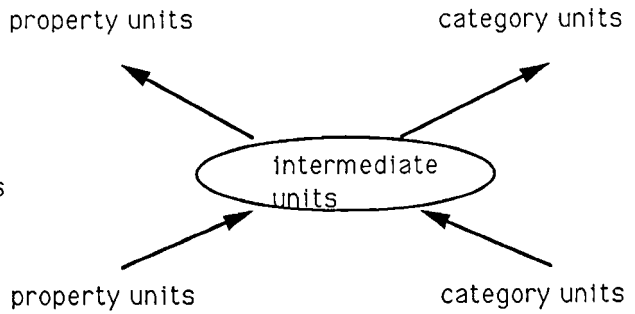
Cues are sometimes not available when we need them to make distinctions between competing forms. For example, the attribute of heat-resistance might often be used to distinguish mugs from cups. However, it may not be possible to judge whether a given drinking utensil is capable of holding hot liquids until it is actually used. Even if this cue is available, it still may not be entirely reliable, since many porcelain cups are as capable of holding hot liquids as are mugs.

The three-level connectionist networks used for work with morphology can also be used to study interactive effects in lexical competition. Figure 4-6 shows a network of this type for the domain of eating utensils. The input to the network would include perceptual and functional properties of the type discussed above—properties such as cylindricality, size, thickness, heat-resistance, and so on. The category inputs would be words like “cup” or “mug.” The intervening or hidden units would detect nonlinear interactions between the properties or categories. The network could be used to predict categories from properties, as in production, or to predict properties from categories, as in comprehension. This type of network is easy to build. However, detailed coding of the features themselves is a laborious process and can only be handled reasonably for small domains. Lexical fields where connectionist simulations could make interesting contacts with current semantic theory include locative prepositions, transfer verbs, verbs of covering, reversible actions, and quantifiers.

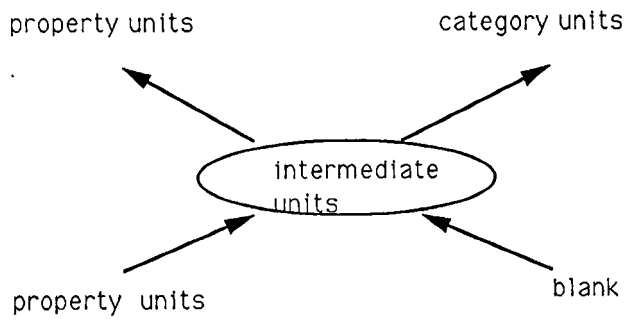
POLYSEMIC COMPETITION

The previous section discussed the competition that occurs during production when we attempt to choose among totally different words. However, there is a level of semantic competition that occurs below the level of the word. This is the competition that occurs during comprehension when we have to select among alternative meanings or polysemes. When we hear a word such as “palm,” we must decide whether to think of it as a tree or as a part of a hand. This competition between alternative meanings of the same word is an extremely pervasive aspect of human language. When a lexical item is detected, it automatically activates each of its polysemes (Swinney, 1979). These polysemes are then placed into competition (Small, Cottrell, & Tanenhaus, 1988). The polyseme supported by the strongest cues wins. Like the competition between words, the competition between polysemes is determined by a process of cue strength summation. The notion of a multidimensional semantic topography is a useful way of understanding the way in which alternative meanings compete. This topography makes distinctions not just between words, but also within words.

Architecture
for Semantics



Labeling
Training



Inference
Training

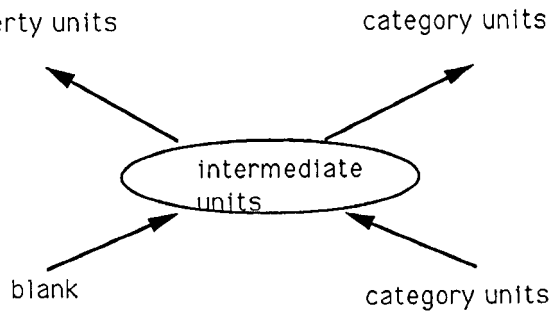


Fig. 4-6. A model for semantic learning.

Consider a word like "ball." Webster's Third contains three major entries for "ball." The first major entry is for a noun that describes round physical objects; the second is for a verb that involves forming things into balls; the third is for a noun describing a formal gathering for social dancing. Within each of the major entries is a series of minor readings or polysemes. For example, the third entry for "ball" has one polyseme for "a formal gathering" and a second for "a good time." Or the first entry has polysemes for things like "odd ball," "a ball game," "a fast ball," "testes," or "keep the ball rolling," along with the basic meaning of "a round object."

Within a given minor polyseme, further polysemy can be found. For example, within the basic polyseme for the first entry of "ball," there are 15 minor polysemes. Various types of round objects called "balls" include the ball of the foot, a baseball, the ball of the earth, an eyeball, a fall of fudge, and so on. Even within the minor polyseme for objects that bounce or roll, there is a long list of types, including baseballs, footballs, golf balls, and so on. Nor does polysemy really end at this level. Within the interpretation of "ball" as "football," specific objects such as "nerf football," "mini football," and "regulation football" can be further distinguished.

The word "ball" may refer to any one of these many different polysemes at the various levels. It is the listener's job to try to decide which of the many competing options is the one which is currently intended. If the listener wants to understand the message at all, it is almost always important to pick out the correct major polyseme. If the listener heard that Cinderella "went to the ball," he would have a very strange idea of what happened if he thought of her as approaching a round object. If he had heard "the baby threw the ball," he would need to avoid thinking of a baby throwing a wad or hot candy or the bone in someone's foot. He would certainly want to focus in on the reading of "round object for throwing or bouncing." However, within this general minor polyseme, it might not yet be possible to distinguish between a beach ball, a volley ball, or a nerf ball. There may be no further information in the discourse or in the discourse context that could tell him which of these particular objects is being thrown by the baby. If there is further disambiguating information, he will attempt to use it. A series of studies by Anderson and Ortony (1975), Anderson et al. (1976), and Anderson, Reynolds, Schallert, and Goetz (1977) demonstrated the degree to which discourse context influences the final interpretation of lexical items. For example, in a sentence such as "The Coca-Cola poured all over the table, and then the container was empty," subjects tend to interpret "container" as a "bottle." However, in a sentence such as "The apples rolled all over the table, and then the container was empty," subjects tend to interpret "container" as "basket." There are three major types of cues for resolving the competition between polysemes.

Connectionist networks can be applied to the problem of processing polysemy. The simplest type of network would have all of the possible

polysemes of a single word as output and all of the cues that are likely to be important in the polysemic processing as possible inputs. For example one could build a network for the processing of the polysemes of the word "ball." In order to activate the reading of "ball" as a dance, the input could include features such as "music" or "costumes." These features would be turned only if there are words containing these features in the actual input. The inputs could activate intervening units that might be generally useful for collecting evidence regarding the concept "dance." Other intervening units would be more activated by meanings of "ball" that have to do with candymaking, and so on. There are four major problems with networks of this type.

1. Word-specific nets of this type would fail to capture the interrelatedness of the lexicon. Given the word "ball," the easiest way to activate the "golf ball" interpretation over the "basketball" interpretation and other competitors, is to activate the "golf" concept. However, if our networks only resolve competitions for single words, this interrelatedness is missed.
2. Isolated single-word networks also fail to capture general polysemic competitions such as metonymies, personifications, and the like. The same extensional logic that allows us to talk about "having a ball" can be used to talk about "having a spin." If the polysemes of "ball" and "spin" are processed in total separation, these effects cannot be captured.
3. Isolated single-word networks cannot, by themselves, resolve lexical competition of the type discussed earlier for "cup," "mug," and "demitasse."
4. In reality, some words are more important than others in determining the polysemic competition. In particular, as seen below, words tend to exert the greatest pressure on other words with which they have syntactic relations.

One straightforward way of dealing with these limitations is to build bigger nets. Nets can be constructed that model both the major lexical competitions for a given semantic domain and all of the polysemic competitions within that domain. Such nets could relate comprehension and production in the way outlined for the simulations of Hungarian morphology above. These richer nets would need to have more input features. For example, in the domain of "cup" and "mug," one would want to add to the cues of shape and function various cues to distinguish between "mug" as a drinking vessel and "mug" as a derogatory term for a person's visage. Of course, the cues for the latter polyseme of "mug" are really ones that are more relevant to the competition between "face," "mug," "visage," and "puss." Each of the first three problems mentioned above points out the

need for constructing really large nets to capture any of the interesting aspects of lexical processing. The fourth problem noted above points in yet another direction to be explored in the next section.

MODELING SYNTACTIC PROCESSING

Languages differ markedly in the way they use grammatical cues to govern attachment competitions. As Bates and MacWhinney (in press) have shown, the cue of preverbal positioning is the strongest cue in English to identification of the subject role. Given a sentence like "The eraser are chasing the boys," English-speaking subjects show a strong tendency to choose "the eraser" as the subject and, hence, the actor. This occurs despite the fact that the noun "boys" has the cues of verb agreement, animacy, and humanness all on its side. These three weak cues are just not enough to counterbalance the strength of the preverbal position cue in English. In Italian, however, the corresponding sentence is *la gomma cacciano i ragazzi* in which *la gomma* "the eraser" has support from the cue of preverbal positioning and *i ragazzi* "the boys" has support from the cues of agreement, animacy, and humanness. As Bates et al. (1984) have shown, agreement is a much stronger cue in Italian than it is in English. In Italian, the strongest cue is verb agreement and the second strongest cue is preverbal positioning. Thus Italians interpret this sentence as meaning "The boys are chasing the eraser."

How can a connectionist model account for this type of competition for syntactic attachment? It would be easy enough to construct a net for a particular sentence. The inputs to the net would be cues such as preverbal positioning, verb agreement, case marking, animacy, and so on. The output units would be the competing nouns. There must also be a way of identifying or keeping track of each competing noun. For example, in the sentence "the dogs are chasing the cat" the first noun phrase would be "the dogs" and the second would be "the cat." The net would be designed specifically to resolve the competition between the *first noun* and the *second noun* for the role of subject of the verb. The cues that would be used include noun animacy, agreement, stress, and certain semantic features of the verb. Having set up the network in this way, it provides a faithful connectionist rendition of the basic syntactic claims of the competition model in regard to the competition between nouns for grammatical roles.

In a sense, the main function of the syntactic net is not to determine syntactic relations, but to resolve ambiguity in the meanings of the words being related. Consider the word "run." The phrase "another run," as in "let's take another run," forces the verb "run" to behave like a noun. If one talks about wanting "a deeper blue," the adjective "blue" is forced to behave like a noun. This use of syntactic combinations to push words into

other part-of-speech categories can be called *pushy polysemy*. What is interesting about pushy polysemy is that it only works between words that are syntactically related through the type of valence relation discussed earlier. In this example, it is the word "another" that is forcing the word "run" to behave like a noun.

Pushy polysemy is strong enough to overcome most of the standard categorizations of words into parts of speech and subclasses of the parts of speech. It can easily force a mass noun to assume a reading as a common noun. For example often it is said that "sugar" is a mass noun and that phrases such as "another sugar" are ungrammatical (Gordon, 1985). From this, it follows that the sentence "I'd like another sugar, please" is also ungrammatical. However, if someone is asking for a small packet of sugar and using the contents of the packet to refer to the whole (metonymy), the extension is quite reasonable and even conventional. Or it could be that someone is working in a chemistry lab analyzing the reactions of various sugars such as fructose, sucrose, and glucose. In this case, "another sugar" refers to another type of sugar. This extensional pathway uses a word to refer to a member of a taxonomic class. One can say that only words like "sugar" can do this because of the special circumstances mentioned. However, even so unlikely a sentence as "I'd like another sand, please" can be interpreted in similar ways. Much like the interpretation of "another sugar" as referring to a packet of sugar, "another sand" might be referring to a bag of sand used either for construction or for sand-bagging a swollen river. Just as one could imagine a chemist working with various sugars, one could imagine a situation where geologists are describing the sand content of a new formation. They have used sieves to sort out the various types of sand in the formation and then placed these sands into jars. One of them asks the other for "another sand" for testing, meaning either another bottle of sand or another type of sand.

Proper nouns can also be converted into common nouns. Usually, we are told that a determiner such as "a" cannot precede a proper noun such as "Reagan." However, there is nothing wrong with a sentence, such as "A wiser Reagan returned from Reykjavik," if "Reagan" is being thought of not just as a single man, but also as a man who can assume various states or values. Virtually any proper noun can be extended in this way. Another extensional path allows conversion of adjectives into nouns, as in the sentence "the green is nicer than the red." This type of conversion works best if new deadjectival nouns can be conceived as members of a collection or ensemble.

Pushy polysemy is also at the heart of co-occurrence learning. It is polysemic processing that allows the "abduction" of semantic facts on the basis of formal regularities. For example, given a sentence such as "the man niffed the plate at the fence," the child can abduce some of the semantics of "niff" on the basis of co-occurrence pattern. The child does this by attend-

ing to the underlying system of connections between semantics and verb frames. This system tells us that "niff" takes a subject and an object and that the action of the subject on the object is like that in "hit" and "slam." The importance of a mechanism of this type has been stressed by MacWhinney (1987), Maratsos and Chalkley (1980), Bowerman (1982), and Schlesinger (1977). There is evidence that even very young children are able to infer the class of a word from co-occurrence data. For example, Katz, Baker, and Macnamara (1974) found that, beginning around 17 months, girls who were given a proper name for a doll learned this name better than girls who were given a common noun. In the proper noun frame, girls were told that the doll was called "Zav"; in the common noun frame they were told that the doll as "a zav." Thus, even at this early age, children seem to realize that names with articles are common nouns and names without articles are proper nouns. This ability to infer the semantics of words on the basis of co-occurrence continues to develop. Werner and Kaplan (1950) were able to show in their classic "corplum" experiment that by age 8, children could acquire many aspects of the semantics of abstract nouns from highly abstract sentence contexts.

The connectionist model of McClelland and Kawamoto (1986) does a good job of simulating this abductive learning. Words that behave formally like other words begin to be treated like those words. For example, in a sentence such as "the doll hit the ball" the simulation has a tendency to begin to attribute animacy to "the doll" on the basis of its status as the subject of "hit." In fact, this learning is not unproductive, because in both fantasy and fiction dolls are often treated as animate.

The eventual goal of this type of a connectionist analysis of verbs and case roles is to go from a set of semantic features to a set of valence descriptions. Such a system will allow for the emergence of generalities on the basis of semantic features, while still tolerating exceptions for high frequency items (Stemberger & MacWhinney, 1986). Within an inheritance network that processes valence and polysemy, more detailed features of the predicate may activate more detailed features of the valence description. For example, if the predicate is "big," the feature [+measurable] is activated for the argument. Of course, virtually any object can be treated as measurable, but the point is that the presence of the word "big" would force focus on the size properties of its argument.

Making valence descriptions subject to semantic features of the predicate and its argument has some further interesting consequences for extensional uses of verbs. For example, the first argument of the verb "polish" is usually an animate actor and the second argument is usually an inanimate object. However, when an inanimate occurs as the first argument in pre-predicate position, as in "this table polishes easily," its presence forces the verb to take on the features [+potential] and [+state] and to drop the feature [+activity]. This general change can apply to any action verb, such as "this

phone dials easily" or "this micro programs easily." Such forms can be produced and comprehended without any prior experience with them, indicating that the valence descriptions involved cannot be frozen forms, but must arise from some general process. In fact, this general process is exactly what is captured by the valence description network.

Nouns have valence descriptions that simply require them to be the argument of other predicates. Thus, all nouns expect to be either the argument of some verb or preposition. However, common nouns have an additional expectation of being the first argument of a modifier with the feature [+delimiting], such as "another," "one," "a," or the plural suffix. Thus, one cannot say "I like dog" without treating "dog" as a mass noun. To treat it as a mass noun, it would have to be thought of in terms of, say, "dog meat."

CLUSTERING

The view of language processing sketched out so far is well within the scope of the types of issues that can be dealt with by current connectionist models. Although many of these phenomena will require very large and complex models, there is little in what we have said so far that lies beyond the scope of connectionist processing. However, there is a fundamental issue that has been carefully concealed under the rug up to this point—the problem of the relation between polysemic processing and the formation of larger structural units. When the predicate "another" combines with the argument "beer," it forms a new structural unit that can then be further combined with additional predicates or can be referred to anaphorically. Here it will be referred to as a *cluster*. Clustering is the fundamental nonconnectionistic process that lies at the heart of the competition model approach to syntactic processing, as sketched out in MacWhinney (1987, 1989). Clustering takes an argument and a predicate, merges their semantic features, and outputs a new syntactic and semantic unit. Consider how clustering works to process a simple sentence, such as "the cat is on the mat." First, "cat" links with "the" to form a new cluster. Then "the cat" links with "is" to form a partially saturated verb. Then "on" links with "the mat" to form an adverbial phrase which then attaches to "is" and the processing is complete. The final clustered structure is:

$$((\text{the} \rightarrow \text{cat}) \leftarrow \text{sat} \rightarrow (\text{on} \rightarrow (\text{the} \rightarrow \text{mat}))).$$

This account of the processing of "the cat is on the mat" ignores any possible competitions for attachments and assumes that each lexical item assumes its default polysemic value. At various points in the left-to-right processing, there are often words not yet attached. The processor must be able to

store these words temporarily. It must also be able to take the output of the polysemic processors and pass them on to new competitions. All that is really required here is the ability to keep track of lexical items and new clusters.

CONCLUSIONS

This chapter has presented detailed findings from connectionist models of morphological processing along with more speculative claims about the design of a general connectionist system. Although connectionist accounts of language processing are very new, they offer a variety of advantages over earlier noninteractionist accounts. One major strength of connectionist approaches is the ability of networks to learn in a general way. The same basic mechanism of learning on error can be used to model the acquisition of declension in German or Hungarian, the development of lexical fields, and the learning of case role frames in English verbs of transfer. The other important property of connectionist networks is their ability to enforce mutual constraint satisfaction. In the area of sentence processing, one can see how pushy polysemy interacts with attachment competition. Although current connectionist models require an external process to keep track of the identity of the competing lexical forms, it may be that future models will be able to express even control processes within a uniform connectionist architecture.

REFERENCES

- Anderson, R., & Ortony, A. (1975). On putting apples into bottles—a problem of polysemy. *Cognitive Psychology*, 7, 167–180.
- Anderson, R., Pichert, J., Goetz, E., Schallert, D., Stevens, K., & Trollip, S. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior*, 15, 667–679.
- Anderson, R., Reynolds, R., Schallert, D., & Goetz, E. (1977). Frameworks for comprehending discourse. *American Educational Research Journal*, 14, 367–382.
- Bates, E., & MacWhinney, B. (In press). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of language processing*. New York: Cambridge University Press.
- Bates, E., MacWhinney, B., Caselli, C., Devescovi, C., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child development*, 55, 341–354.
- Berko, J. (1958). The child's learning of English morphology. *Word*, 14, 150–177.
- Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: the state of the art*. New York: Cambridge University Press.
- Braine, M. D. S. (1971). The acquisition of language in infant and child. In C. Reed (Ed.), *The learning of language*. New York: Appleton-Century-Crofts.

- Braine, M. D. S. (1987). What is learned in acquiring word classes—a step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- Gordon, P. (1985). Evaluating the semantic categories hypothesis: The case of the count/mass distinction. *Cognition*, 20, 209–242.
- Katz, N., Baker, E., & Macnamara, J. (1974). What's in a name? A study of how children learn common and proper nouns. *Child Development*, 45, 469–473.
- Köpcke, K., & Zubin, D. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für germanistische Linguistik*, 11, 166–182.
- Köpcke, K., & Zubin, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*, 93, 26–50.
- MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, 42, 1–122.
- 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: Vol. 2, Theoretical issues*. Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (1989). Competition and teachability. In M. Rice and R. Schiefelbusch, (Eds.), *The teachability of language*. Baltimore: Brooks-Cole.
- MacWhinney, B. (In press). Competition and connectionism. In B. MacWhinney and E. Bates (Eds.), *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
- MacWhinney, B., & Bates, E. (In press). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
- MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, 28, 255–277.
- Malkiel, Y. (1968). The inflectional paradigm as an occasional determinant of sound change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics*. Austin: University of Texas Press.
- Maratsos, M. (1982). The child's construction of grammatical categories. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.
- Maratsos, M., & Chalkley, M. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's language: Vol. 2*. New York: Gardner.
- McClelland, J., & Elman, J. Interactive processes in speech perception: The TRACE model. In J. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing: Vol. 2. Psychological and biological models*. Cambridge: MIT Press.
- McClelland, J., & Kawamoto, A. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing*. Cambridge: MIT Press.
- Mills, A. (1986). *The acquisition of gender: A study of English and German*. Berlin: Springer.
- Park, T.-Z. (1981). *The development of syntax in the child with special reference to German*. Innsbruck: Amoe.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge: MIT Press.
- Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a

- parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.
- Rumelhart, D., Hinton, G., & Williams, E. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1, Foundations*. Cambridge: MIT Press.
- Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2, Psychological and biological models*. Cambridge: MIT Press.
- Rumelhart, D., & McClelland, J. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- Schlesinger, I. M. (1977). *Production and comprehension of utterances*. Hillsdale, NJ: Erlbaum.
- Schneuwly, B. (1978). *Zum Erwerb des Genus in Deutschen: eine mögliche Strategie*. Manuscript, Max-Planck Institut, Nijmegen.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson and D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart, & Winston.
- Small, S., Cottrell, G., & Tanenhaus, M. (Eds.). (1988). *Lexical ambiguity resolution*. San Mateo, CA: Morgan Kaufmann.
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-660.
- Taraban, R., McDonald, J., & MacWhinney, B. (In press). Category learning in a connectionist model: Learning to decline the German definite article. In R. Corrigan (Ed.), *Linguistic categorization*. Amsterdam: Benjamins.
- Tucker, R., Lambert, W., & Rigault, (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Mouton.
- Wängler, (1963). *Rangwörterbuch hochdeutscher Umgangssprache*. Marburg: Elwert.
- Werner, H., & Kaplan, (1950). The development of word meaning through verbal context: An experimental study. *Journal of Psychology*, 29, 251-257.