

# Language Learning: Cues or Rules?

Brian MacWhinney  
Jared Leinbach  
Carnegie Mellon University

Roman Taraban  
University of Massachusetts

Janet McDonald  
Louisiana State University

This is a final draft of:

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

Abstract:

Child language researchers have often taken gender and case paradigms to be interesting test cases for theories of language learning. In this paper we develop a computational model of the acquisition of the gender, number, and case paradigm for the German definite article. The computational formalism used is a connectionist algorithm developed by Rumelhart, Hinton, and Williams (1986). Three models are developed. In the first two, various cues to gender studied by Köpcke and Zubin (1983, 1984) are entered by hand. In the third, the simulation is given only the raw phonological features of the stem. Despite the elimination of the hand-crafting of the units, the third model outperformed the first two in both training and generalization. All three models showed a good match to the developmental data of Mills (1986) and MacWhinney (1978). Advantages of a connectionist approach over older theories are discussed.

When generative grammarians write a grammar for a language, they do it by formulating a set of rules together with a set of exceptions to these rules. In their classic work, "The Sound Pattern of English" (SPE), Chomsky and Halle (1968) provided a fairly complete generative grammar for English derivational and inflectional morphology. One of the goals of that work was to express the grammar in a maximally compact form, thereby capturing the maximum number of "linguistically significant generalizations." As a result, the rules of SPE were highly symbolic, having numerous subconditions, alternative environments, and variables. Moreover, the forms upon which those rules operated were abstract, often with no direct relation to any actually occurring phonological form. Although modern generative phonology no longer subscribes to the formulations of SPE, it still accepts the view of grammar as a set of rules based on often highly abstract symbols.

The status of rules in descriptions of children's language has often been called into question. For example, both psycholinguists such as Ervin-Tripp (1966) and Slobin (1971) and linguists such as Hockett (1968) and Givón (1984) have pointed out that there is no direct evidence that language users actually manipulate rules and rule symbols in their heads in the same way that rules are processed in a linguist's grammar. Nor is there direct evidence that children pick up rules

during actual conversational interactions. How then is it that children come to behave as if they knew the rules, even without learning rules as discrete entities? Does the mechanism underlying this learning resemble the rule itself or does it look like something very different?

Our goal here is the articulation of an approach to morphological processing and structure that does not depend on rules. This approach views morphological form as arising from a competition between a large set of phonological, syntactic, and semantic cues. The competition is expressed computationally in the form of a connectionist network (Rumelhart and McClelland, 1987), rather than through the formalisms of generative grammar. A connectionist architecture is selected as a particular computational formalization of a general model of language processing and acquisition we have called the “Competition Model” (MacWhinney and Bates, in press). As a testing ground for the claims of the Competition Model, we have chosen to model the learning of the German declensional paradigm. Because of the complexities of gender, number, and case assignment in German, child language researchers have often (Braine, 1987, MacWhinney, 1978; Maratsos and Chalkley, 1980; Maratsos, 1982 and Pinker, 1984) viewed the acquisition of German declensions as a challenge to language learning theory.

Rumelhart and McClelland (1987) used neural networks to model the ways in which English-speaking children learn to form the past tense of the verb. The Rumelhart and McClelland verb-learning model was able to produce overextensions such as “drawed,” as well as correct formations such as “fell” and “ran.” However, the target structures in English are fairly simple. There are only a few dozen irregular verbs and the paradigm only involves three forms - the present, the past, and the perfect. In German, there is no regular or default pattern and the paradigm has 16 cells for the article alone and separate paradigms for the adjective and the noun. Moreover, whereas English appears to have some simple phonological cues that predict irregular past tense forms (Bybee and Slobin, 1982), it is generally believed that the gender of German nouns is entirely arbitrary (Maratsos and Chalkley, 1980) and cannot be predicted by any set of cues or combination of cues.

Both generativists and connectionists agree that grammar must be learnable. In the generative framework, one first formulates a descriptively adequate grammar such as that of SPE. Next, one attempts to formulate a learning model that guarantees that the target structures are learnable. In constructing a learning account, the generativist is often forced to attribute to the learner a fair amount of abstract innate knowledge (Wexler and Culicover, 1980; Pinker, 1984). The postulation of innate structures is often taken to be a goal in its own right and is even used as proof of the correctness of the structural analysis. The connectionist framework we will adopt takes a different approach to learning. In that framework, learning and processing are treated directly in the same computational architecture. As the network learns, its processing abilities develop. There is no separation between learning, structure, and processing. The emphasis is upon maximizing the contribution of the learning algorithm and minimizing the recourse to innate abilities.

Our goal in formulating a process model for the acquisition of German declension is not simply to raise questions regarding the value of the generative approach. We are interested in constructing a concrete cue-based alternative to that approach. Of course, we want an approach that will learn the target structures. The real test of this alternative will be its ability to simulate German children’s language learning (MacWhinney, 1978; Mills, 1986). We want to see if the model learns as the child learns. Does the model make the same errors the child makes and is the order of correct mastery of forms in the simulations like the one we find in the child?

## Declension in German

The declensional paradigm in German is configured around three morphosyntactic dimensions.

1. **Number.** As in English, nouns and pronouns in German can vary in *number*, since they can be either singular or plural. For example, the word for “student” is *Student* and the

plural form is *Studenten*. Changes in number are also marked on the article or other modifier, so that the singular form “the student” *der Student* becomes plural *die Studenten*.

2. **Case.** The second dimension along which nominals may vary is case. Both nouns and pronouns can be in the nominative, the accusative, the genitive, or the dative case. For example, the nominative singular form of “student” is *der Student* and its accusative form is *den Student*. Typically, subjects are in the nominative, direct objects are in the accusative, and indirect objects are in the dative. The genitive is used primarily to mark possession. Prepositions can take either the accusative or the dative and sometimes the genitive. Typically, the dative is used when the verb is static and the accusative is used when the verb expresses motion.
3. **Gender.** The third dimension is gender. Nouns can be masculine, feminine, or neuter. A male student is *der Student* and a female student is *die Studentin*. The choice of *der*, *die*, or *das* in the nominative reflects choice of one of the three genders for the noun. To some observers (Maratsos and Chalkley, 1980) the assignment of nouns to genders has seemed entirely arbitrary. To others, like Mugdan (1977), the assignment has seemed rule-governed, but exceedingly complex.

The noun itself is primarily marked only for the dimension of number. Gender is not marked on the noun and case is only marked on the noun for the genitive singular and the dative plural of certain nouns. The article does the main work of marking gender, number, and case. Simplifying the situation quite a bit, let us focus on the way in which nominal marking is achieved by the selection of the correct form of the definite article. Theoretically, a complete cross of the categories of gender, number, and case yields 24 possible cells for the full German declensional paradigm. Fortunately for the German child, gender distinctions for the definite article disappear in the plural, reducing the paradigm to 16 distinct cells. The complete paradigm for the German definite article is shown in Table 1. Although there are 16 cells in the paradigm, there are only six different forms of the definite article (*der*, *den*, *dem*, *des*, *die*, *das*). Each form of the article occurs in at least two different cells of the paradigm, so that no form defines a unique combination of gender, number, and case. For example, the article *der* can mark the masculine nominative singular, the feminine genitive singular, the feminine dative singular, or the genitive plural.

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

Acquisition of this system is not a trivial task. Overall, this type of learning can be viewed as a three-dimensional word-class formation problem (Levy, Schlesinger, & Braine, 1987). The three dimensions to be controlled are gender, number, and case. In production, control of this system involves correct selection of article and noun markings. On the first dimension, nouns must be placed into one of three gender classes. On the second dimension of number, the child must decide on semantic grounds whether a noun should be singular or plural. If it is to be plural, the child must choose from one of eight pluralization types. On the third dimension of case, the various cues and configurations in the sentence must be grouped together so that they correctly select the case of the noun. In comprehension, the child’s task is to use the various forms of the definite article and the markings on the nouns as cues to the correct assignment of the noun to a particular gender, a particular number, and a particular case. Let us look in more detail at each of these three dimensions of this word-class formation problem.

### Cues to gender assignment

The simplest way to solve the word-class formation problem for gender is for the learner to find a set of reliable cues that tells him when to assign a noun to a certain class. Maratsos and Chalkley (1980) have argued that German gender is so arbitrary that no set of cues would allow a child to assign a noun to its gender class. Why, for example, is “fork” feminine (*die Gabel*),

“knife” neuter (*das Messer*) and “spoon” masculine (*der Löffel*)? In fact, recent work has shown that, while the German gender system is complex, it is not as arbitrary as it appears on first analysis. In a series of research reports, Klaus-Michael Köpcke and David Zubin (Zubin and Köpcke, 1981, 1986; Köpcke and Zubin, 1983, 1984; Köpcke, 1982) have conducted a broad survey of various types of German nouns and found that there is a large and powerful set of cues to German gender. Using these cues, Köpcke (1982) was able to correctly assign gender to 90% of the 1466 monosyllabic words listed in the first volume of the Duden (Grebe, 1973). The work of Köpcke and Zubin for German is parallel in many ways with that of Tucker, Lambert, and Rigault (1977) on the prediction of gender in French. Both research groups have found that there are indeed a large number of morphological and phonological cues predicting gender.

The most important cues discovered by Köpcke and Zubin are given in Table 2 along with a few additional cues taken from a German grammar (Lederer, Schulz and Griesbach, 1969). There are 38 cues in all. Of these, 15 are phonological, 18 are morphological, and 5 are semantic. Some of these cues to gender are absolute. For example, if a word has a diminutive ending (i.e. *-lein* or *-chen*), the noun is guaranteed to be of neuter gender. Other cues are more probabilistic in nature. For example, although nouns that start with an “sh” followed by a consonant tend to be masculine, there are words (e.g. *die Stadt*, *das Spiel*) that violate this mapping. The use of this kind of cue will not guarantee a correct gender classification, but it will improve the chances of a correct classification.

Type		Gender	Example	English
Phonological	umlauting	Masculine	Der Arger	Anger
	ir- or dr-	Masculine	Der Trieb	Force
	CV-	Masculine	Der Tabak	Tobacco
	CCV-	Masculine	Der Klub	Club
	CCCV-	Masculine	Der Strich	Stroke
	-VC	Masculine	Der Beamier	Official
	-VCC	Masculine	Der Hahn	Rooster
	-VCCC	Masculine	Der Markt!	Market
	monosyllabic	Masculine	Der Akt	Nude
	shC-	Masculine	Der Schrank	Closet
	-el	Masculine	Der Schliissel	Key
	-n	Masculine	Der Zahn	Tooth
	-fricative + 1	Feminine	Die Nacht	Night
	-e	Feminine	Die Sonne	Sun
	-<e>s	Non-feminine	Das Glas	Glass
Morphological	-ling	Masculine	Der Feigling	Coward
	-ent	Masculine	Der Patient	Patient
	-er	Masculine	Der Reiter	Rider
	-eur	Masculine	Der Redakteur	Editor
	-ei	Feminine	Die Malerie	Painting
	-je	Feminine	Die Phantasie	Fantasy
	-ik	Feminine	Die Polemik	Polemic
	-in	Feminine	Die Studentin	Co-ed
	-ion	Feminine	Die Portion	Portion
	-ilat	Feminine	Die Realitat	Reality
	-sis	Feminine	Die Basis	Basis
	-ung	Feminine	Die Zeitung	Newspaper
	-lein	Neuter	Das Fraulein	Young woman
	-ment	Neuter	Das Instrument	Instrument
	-ell	Neuter	Das Tablet!	Tablet
	-chen	Neuter	Das Madchen	Girl
	-en	Neuter	Das Sehen	Seeing
	-um	Neuter	Das Gymnasium	High School
Semantic	Natural male	Masculine	Der Sohn	Son
	Natural female	Feminine	• Die Tochter	Daughter
	Young being	Neuter	Das Kind	Child
	Superordinate	Neuter	Das Tier	Animal

### Cues to selection of a plural marker

Unlike the gender dimension, which has no single real-world correlate, the dimension of number maps directly onto salient features of the external world. The decision to treat a noun as singular or plural involves none of the complexities of the decision to treat a noun as either masculine, feminine, or neuter. Whereas gender is marked only on the modifiers and never on the noun, number is marked most clearly on the noun itself. But this marking is not simple (Köpcke,

1988), since there are eight different ways to mark the plural. The actual choice of one of these eight forms is governed by a set of cues that are almost as complex as those governing gender assignment. For example, the plural of *die Flut* “flood” is *die Fluten*, the plural of *das Gut* “estate” is *die Güter* while the plural of *der Hut* “hat” is *die Hüte*. There are some regularities in the assignment of these plural morphemes to a word based on the suffixes and prefixes on the stem, the mutability of the stem vowel, and the gender and animacy of the noun (Köpcke, 1988). Table 3 illustrates the eight possible ways in which nouns may be pluralized.

Change	Singular	Plural	Translation
-e	Tag	Tage	days
-(e)n	Blume	Blumen	flowers
-er	Kind	Kinder	children
-s	Radio	Radios	radios
0	Zimmer	Zimmer	room
	Bruder	Bruder	brothers
"-e	Hand	Hande	hands
"-er	Mann	Manner	men

### Cues to case assignment

Like gender, case is marked primarily on the modifiers of the noun. Some nouns take the -s ending for the genitive singular and some take the -n ending for the dative plural, but often there is no information on the noun that indicates its case. Instead, case is mostly marked by the choice of one of the various forms of the articles or adjectives modifying the noun. Cues to case in German occur on the morphological, syntactic and semantic level. Morphological cues include the -(e)s ending added to singular masculine and neuter words in the genitive case, and the -n ending added to plural nouns in the dative case. Other morphological and syntactic cues include accusative, dative and genitive prepositions, subject-verb agreement, and word order. Semantic cues include verb meaning and semantic roles. These cues to case may be simple--i.e. a word following the dative preposition *mit* is always in the dative case, or they may depend on a combination of factors. For example, some prepositions may take the accusative case or the dative case depending on whether the verb in the sentence is a verb of motion (e.g. *Ich lief unter die Brücke* “I ran underneath the-ACC bridge”) or a static verb (e.g. *Ich stand unter der Brücke* “I stood under the-DAT bridge”).

### Cues and Categories

Both word classes and syntactic classes are linguistic categories. From the viewpoint of the Competition Model (Bates and MacWhinney, 1982, 1987; MacWhinney, in press), categories involve many-to-many mappings between forms and the cues that predict the use of these forms. Any particular cue can map to many categories. For example, in German, a final -e can be the pseudo-derivational marker of feminine gender, one of the markers of plurality, or the first person singular present marker on verbs. In addition, many cues can map to the same category. For example, the endings -e, -ung, -ie, as well as natural feminine gender all map to feminine gender. Because many cues can map to the same category, more than one cue can be present in a given instance. For example, *der Schnaps* “schnapps” begins with the “sh plus consonant” cue, which is indicative of masculine gender. *Schnapps* also has the semantic cue of alcoholic beverage, which cues masculine gender. However, because cues are not completely reliable, an item may have one set of cues pointing to one gender and another set of cues pointing to another gender. For example, *der Junge* “boy” has the final -e ending that strongly indicates feminine gender, but has the semantic cue of natural masculine gender. Because of the many-to-many mapping, the “informational value” of a cue depends on the strength of its association to alternative categories. For example, if a cue is associated with masculine, feminine, and neuter genders an equal number of times, it is not a very useful cue for gender.

Cue use requires coordinating multiple cues and weighting them appropriately. These observations about processing have been made for the German article and for many other cue-to-category relationships in language in the Competition Model (MacWhinney, Bates and Kliegl, 1984; MacWhinney, 1987; Bates and MacWhinney, 1987; McDonald, in press). This model has guided the simulations that we present in this paper -- first with its premise that the relationship between cues and categories is many-to-many, and second with its premise that cues contribute to the categorization process in proportion to their relative informational value.

The Competition Model holds that cue acquisition and cue strength depend on four basic properties of cues: detectability, task frequency, availability, and reliability. In order for a child to learn to use a cue, the child must first be able to detect the presence of the cue. Without being able to hear the difference between the dative article *dem* and the masculine accusative article *den*, the child cannot even begin to gather further evidence to make this case distinction. Once a cue is detected, its initial acquisition is affected by the overall frequency of the category to which it relates. The problem of marking the subject of a verb occurs much more frequently in discourse than the problem of marking the possessor of an object. Because of this, the task of marking the nominative case in German occurs much more frequently than the task of marking the genitive case. This higher task frequency for the nominative gives the child more experience with the various cues that predict its form. The more often a child needs to process a certain type of cue or assign a particular category, the sooner he will learn the cue-category relationship. Within a particular category, such as the German nominative, the frequencies of particular cues also vary. In the terms of the Competition Model, we say that these cues then vary in their *availability*. The higher the availability of a marking, the earlier will be its acquisition, all other things being equal. The child will learn to use the *-e* marking for feminine gender in German before the *-nis* marker simply because the former is far more available as a marker of noun gender. The dimensions of detectability, frequency, and availability are important in understanding differences in order of acquisition across domains. Even when a cue is highly frequent and highly available, it may not always correctly indicate the same category. Although the presence of the *-e* ending is highly likely to indicate a feminine noun, there exist both masculine words--e.g. *der Junge* "boy"--and neuter words--e.g. *das Ende* "end"--that contain this ending. Reliability specifies how often a cue is associated with each category--e.g. masculine, feminine, and neuter--in those instances when the cue is present.

### **Empirical data on the learning of German declension**

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 agree in large measure and also match well with non-experimental observations such as those of 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.** The empirical literature indicates that children first achieve correct mastery of the use of the nominative case. In particular, use of the nominative for the accusative is frequently reported (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 attention to other ways of controlling gender categorization. One simple way of solving the problem is to simply omit the article. Early on, omission of the article is very common. 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 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. Schnewly (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. A general finding of the Competition Model is that cue strengths depended on the *reliability* of a cue, both for morphological categories (MacWhinney, 1978) and syntactic categories (MacWhinney, Bates, and Kliegl, 1984; McDonald, 1986). There is also evidence that the relation between cue validity and cue strength changes during the course of acquisition, with initial stages dependent on cue availability and detectability and later stages on reliability (Sokolov, in press), or on conflict cue validity (McDonald, 1986; McDonald and MacWhinney, 1987).
7. **Children can use cues to infer word class.** 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 general form of the models

The general architecture of the networks we used in these simulations is based on the “back propagation” algorithm of Rumelhart, Hinton, and Williams (1986). Like other connectionist models, our models consisted of a large number of densely interconnected elements operating in parallel. Hereafter we will refer to these elements, or processors, as “nodes.” The network’s knowledge is contained in the strength of the connections between the nodes in the network. Our networks were composed of several internal layers of nodes, in addition to an input layer and an output layer. The number of nodes at each level and the pattern of interconnections between nodes were free parameters that we set in advance of a simulation. On any given trial of the network, each connection in the network assumes a scalar value that serves as the input to the next node. This value is the product of the current activation level of the node on the input side of the connection and the strength or “weight” associated with the connection. In turn, the activation level on each node is a function of the sum of all of the inputs coming into it.

In the three models to be described, German nouns were presented to the network one at a time. The input layer encoded the presence or absence of cues associated with a particular noun and its sentential context. We will discuss the exact nature of these cues in more detail as we describe each model below. Each node on the input layer described a single cue. If the cue was present for a particular noun, the input node was fully activated, and if it was not present, the node remained “off.” The words, therefore, were represented as sets of cues. The activation of the input layer produced activation on the internal layer(s), which in turn produced activation on the output layer. Patterns of activation values on the output layer represented each of the six German definite articles.

The activation level for a node is actually represented by a positive real number between 0.01 and 0.99, so we had to define the level at which a node would be considered “on.” We did this as follows. If an output node was supposed to come on for a given learning trial and had activation greater than .5, the unit was considered to be on. If activation was less than .5, the unit was considered to be off. If the unit was on and was supposed to be on, the trial was considered a hit. If the unit was on but was supposed to be off, the trial was considered a false alarm. If an output unit was supposed to be off and was off, the trial was considered to be a correct rejection. If the unit was off but was supposed to be on, the trial was a miss. Thus, correct performance required a stringent pattern of activation over the output units. In addition, the patterns of activation on the internal layers could be examined. We can think of these internal layers as forming a useful internal representation of the input. In our simulations, we would expect these to correspond to the grammatical categories that describe the German nouns presented to the network.

The beginning of a run of a simulation corresponded to the beginning of a learning sequence. At this point, all the weights on the connections were assigned small random weights. A training set for the simulation, which is described for each simulation below, 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 was therefore consistent for all connections in the network, with no *ad hoc* intervention into the learning process.

We develop a connectionist architecture as a model of the acquisition of the declension of definite articles in German. The input to the network consists of the kind of input considered to be available to a learner and the desired output is the correct form of the German article. We show not only that such a network can learn to correctly assign definite articles to a set of training items, but also that it forms consistent internal representations of its knowledge and is able to generalize this knowledge to new instances.

## Model 1

The first model attempted to mimic the acquisitional situation of a young native German learner. Because we lack detailed data on the input to German-speaking children, we were forced to select words from a frequency count of a spoken German corpus of over 80,000 words (Wängler, 1963). This corpus is based on adults speaking to other adults rather than to children, so frequencies and actual vocabulary items may differ from that heard by children. As a result, the input to the simulation can only be viewed as an approximation to the actual input received by the German-speaking child. As we begin to obtain increasingly accurate data about the actual shape of the input to the German-speaking child in the context of developing databases such as the CHILDES system of MacWhinney and Snow (1985), we will be able to make the match between the simulation and the real acquisitional situation increasingly accurate.

LEXICAL DISAMBIGUATION PATTERN	SEMANTIC CUES	PHONOLOGICAL CUES	MORPHOLOGICAL CUES	EXPLICIT CASE CUES
--------------------------------------	------------------	----------------------	-----------------------	-----------------------

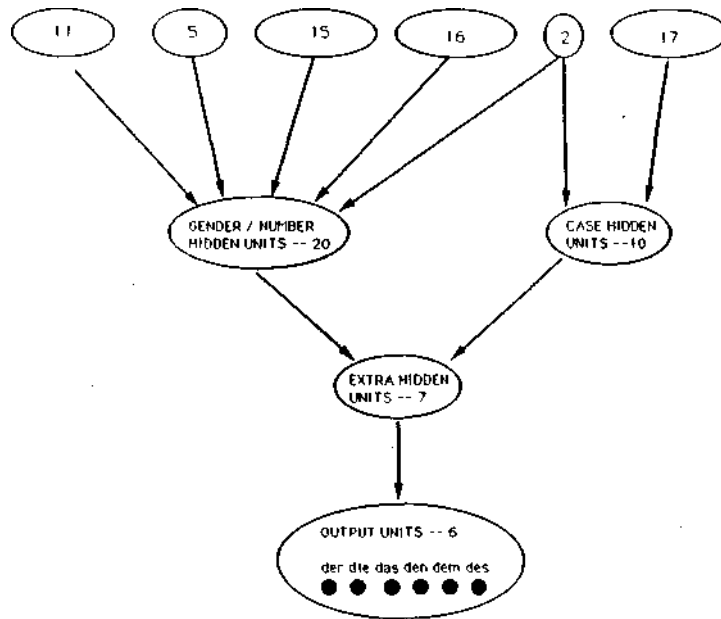


Figure 1: *Architecture for Model 1*. The circles indicate pools of units. The arrows represent interconnections between all the units in one pool and all the units in another pool. The top row of square describes the nature of the input units in the circles directly below. The three pools of hidden units are for gender/number cues, case cues, and the conjunction of gender, number, and case. The output units are for the six forms of the definite article.

Figure 1 displays the architecture of the network for Model 1. The network was repeatedly trained on a input set consisting of German nouns and their associated definite articles, as described above. Nouns appeared in the training set as a function of their real-world frequency. The tokens for any particular noun appeared in a subset of case contexts for the nominative, accusative, dative, and genitive cases. We assessed knowledge of the paradigm in Table 1 by seeing how well the network learned the nouns in new contexts. This corresponds to a test of the kind of paradigm inferencing evidenced by children in MacWhinney (1978). In a second test, we assessed how well the network induced “rules” for gender, case, and number by testing the network with new words.

### The training set

The training set consisted of 102 different nouns that had between 15 and 166 occurrences in the corpus of Wängler (1963). The relative frequency of occurrence for these 102 nouns was preserved by entering each noun into the training set at one tenth of the frequency with which it occurred in the Wängler corpus. Thus, a noun that occurred 80 times in the corpus occurred 8 times in the training set. Most nouns had only one (11 nouns), two (48 nouns), or three (23 nouns) occurrences; the remaining 20 nouns had from four to 17 occurrences. This yielded a total of 305 tokens in the training set.

### Input units

Each of the 102 nouns in the training set was coded for the presence or absence of each of the 38 gender cues of Table 2. Each token of a noun appeared with the same cues. In many cases, several nouns received the same coding in regards to the features of Table 2. In order to allow the network a chance of treating each lexical item differently, each of the 102 nouns was additionally given a unique feature code that was composed by turning on a

different combination of 4 out of 11 additional input nodes. These nodes had no correspondence to any real properties of the nouns. They were used simply to allow the network to fully distinguish individual noun tokens.

Each of the 305 tokens in the training set was presented in a case context. These contexts were represented by a set of 19 cues included in the input for each token: the 2 case endings (e.g. -s for masculine and neuter nouns in the genitive, -n for plural nouns in the dative), 7 prepositions, 7 word order configurations (NNV, NVN, VNN, NN, first noun, second noun, third noun), and 3 verb types (verb of motion, copular verb, and plural verb). Twenty such contexts were used with 8 for the nominative, 5 for the accusative, 5 for the dative, and 2 for the genitive. This distribution was chosen to approximate the frequencies with which these cases actually occur in German (41.6% nominative, 24.1% accusative, 24.9% dative, and 9.4% genitive) (Meier, 1967). The 20 contexts were then randomly assigned to noun tokens, with the restriction that the same context could not be repeated for the same noun. Given that the most frequent noun appeared with 17 tokens, as discussed above, no noun in the training set appeared in all 20 possible case contexts.

## Training Results

The goal of the training was to adjust the weights so that, after repeated exposures to the input, the cues for each word would come to reliably activate the associated definite article as the output. The simulation was run 20 different times always using the same training set. The results of these 20 simulation runs were all quite similar. Learning proceeded smoothly to a level of mastery. In 13 of these, all the items in the learning set were mastered within 100-200 epochs (passes through the learning set). Thus, the network was able to use a set of input cues based on the noun stem and the noun's use in a sentence to correctly select one of the six forms of the definite article. At this point in our work, we cannot make a direct comparison between a particular epoch and a child's age.

On seven of the simulation runs, one or two items remained unlearned long after all the other items in the set were learned. Four of these intractable items were *Bild* "picture," *Brief* "letter," *Hunger* "hunger," and *Stück* "piece." It is not entirely clear what the particular source of the difficulty was for these items, so we will withhold speculation. For three other items, *Junge* "boy" in the dative case and *Ende* "end" in the nominative and accusative cases, it is somewhat easier to understand why the network had difficulties. *Junge* is a masculine singular noun that is an exception to two patterns. First, it contains the *-e* ending, which is a cue that reliably predicts feminine gender both in the language and in the items in the training set. Second, *Junge* belongs to a small class of weak masculine nouns that take an *-n* ending in cases other than the nominative. This *-n* ending is characteristic of the plural in feminine nouns, and is identical to the ending all plural nouns must take in the dative. In one run the network failed to activate any article for *Junge* to the activation level required to be considered 'on,' and in two other runs it activated the dative plural (an incorrect response), which is a reasonable error. *Ende* is a neuter singular noun that is another exception case, containing the *-e* ending, which is highly indicative of the feminine gender. After all the other items had been learned in three of the simulation runs, the network continued to incorrectly treat this noun as feminine, assigning the definite article *die* to its two occurrences--one in the nominative and one in the accusative case, and only after many additional epochs of learning was the correct article *das* assigned to *Ende* .

In order to look at late-learned items more carefully, we stopped learning in 3 consecutive runs at a point at which about 15 words remained unlearned. The errors fell into two general categories: weak learning for nouns in the genitive case; and errors associated with paradigm overlap. In half of the cases, the noun failed to activate an article. These will be referred to as cases of 'omission' (or 'misses'). These nouns were all in the genitive

case. This case, as stated above, represented only 10% of the case contexts. The genitive case, therefore, was sparsely represented in the learning set, and the results here suggest a frequency effect for this case. Because the network is exposed to the case relatively fewer times compared to other cases, it takes the network longer to learn these words. There may also be specific difficulties associated with genitive cues, although we cannot comment on these at present.

The majority of the remaining errors involve paradigm overlap, either between feminine singulars and plurals, or between masculine and neuter singulars. Table 1 shows that the only difference between the articles that mark the feminine singular and those that mark the plural is in the dative. Given that singulars are more highly represented, there was a tendency to treat some instances of plurals as feminine singulars. Feminine plurals in the dative that were assigned the singular dative *der* on one or more of the test runs were *Leute*, *Fragen*, *Schuhe*, and *Minuten*. Instead of treating the dative plurals *Kinder* and *Männer* as feminine singulars, they were assigned masculine singular dative *dem*, consistent with their strong cue for masculine, which is the *-er* ending. Table 1 also shows that masculine and neuter singulars differ only in the nominative and accusative cases. Therefore, we might expect this overlap in the paradigm to cause some shifting between these genders, which it did: *Stück* and *Ding*, which are neuter, were twice treated as masculines, and *Hunger*, which is masculine, was twice treated as a neuter.

## Generalization Results

It could be 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 used. The first test checked how well the network was able to induce the case 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. That is, if a word had been paired with 17 of the 20 case contexts in the training set, it occurred with the remaining 3 contexts in this test. If a word had occurred with only 1 of the case contexts in the training set, it was paired with the remaining 19 case contexts in this test. This yielded a total of 1735 items for the generalization test. Each item in this test consisted of a combination of gender, number, and case cues that were being presented to the network for the first time. The test thus allowed us to see how well the network had learned the case paradigm and whether it could generalize case information to nouns that it had never seen in the test contexts.

This test set was given to the network after it had achieved 100% performance on the training set. During testing, weights in the network were not altered. The results of this test were excellent. On the five generalization runs, the model had an average success rate of 92%. The chance level here would be 16%. This high level of performance provides strong evidence that the network was able to successfully generalize parts of the overall paradigm to the 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. For example, if a noun occurred in the training set with nominative and accusative *die* and 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 noun or whether it was a plural noun, since the plural takes the same articles as the feminine singular in these cases. The ambiguity is even more confounded, since one of the most frequent cues to feminine, final *-e*, 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* -- i.e. 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, even in the face of evidence in either the nominative or the accusative as to the gender of the noun.

We next conducted a second generalization test designed to further examine the productivity of the cues to gender assignment. From the results of the first test, it was not clear whether the gender and number cues were simply acting as a code for a word and activating its cooccurrence pattern, as we might expect according to the Maratsos and Chalkley (1980) model, or whether the network was learning something about the underlying cue structure of a word, according to the Competition Model. In order to examine this issue, 48 nouns next highest in frequency in the Wängler corpus (range 10-14 occurrences) were coded for their word constant cues, and placed in each of the twenty case contexts, for a total of 960 items. This test, more stringent than the first test, allowed us to see how well the various gender and number cues generalize to totally new words of unknown gender. Given the tenets of the Competition Model, it was expected that generalization would be most successful when the gender cues that are present are high in reliability and frequency.

The results of the test showed that the network was able to successfully generalize the paradigm to new words. On average across five runs, the network assigned the correct definite article and activated no incorrect article on 61% of the new instances. Since there are six definite articles, chance performance on this task would be 16% correct. The actual rate of correct answers is clearly much higher than that expected from chance alone. Many of the errors showed that the network incorrectly inferred the gender or number of a noun, and then systematically assigned articles based on this inference. For example, the neuter noun *Kleid* was assigned *der* in the nominative case, *den* in the accusative, *dem* in the dative and *des* in the genitive, indicating that it thought *Kleid* was a masculine noun. Other nouns that in one or more runs consistently followed the wrong gender in all cases, or in three out of four cases, were the plurals *Blumen* and *Augen*, which were treated as singular feminines; the neuters *Papier*, *Glück*, *Licht*, *Heft*, and *Fräulein*, which were treated as feminines; the masculines *Pfennig*, *Anfang*, and *Westen*, which were treated as feminines; and the neuters *Beispiel*, *Krankenhaus*, and the feminine *Zahl*, which were treated as masculines.

### Comparison to the Developmental Literature

In order to examine the degree to which the performance of the network corresponds to early learning in the child, we ran a number of simulations to a point at which one-half of the total error at the beginning of a training sequence was eliminated. At this point the network had an average error rate of 42% (range = 41-45%). Some of the results of these runs are given in Table 4. The numbers in Table 4 indicate the percentages of total usages that involved errors for particular case forms. These numbers do not sum to 100% across either rows or columns.

TABLE 4 PERCENT ERRORS IN EARLY LEARNING IN SIMULATION I

Nominative	Accusative	Dative	Genitive
42	49	26	70

34	44	9	40
•2	0	13	18
4	4	0	0
2	1	0	0
0	0	4	2

---

Let us look at the extent to which the simulation matched the six major phenomena noted in the developmental literature.

1. **Early acquisition of the nominative.** Table 4 shows the average proportion of tokens in each case that were errors for four consecutive tests of early learning. It appears that errors at this early stage are a function of the frequency of the case. The nominative case, which is most frequently represented in the training set, is shows the second lowest proportion of errors. This finding is in line with the predictions of the Competition Model.
2. **Delayed acquisition of the genitive.** The genitive case, which has the lowest frequency of occurrence both in real input and in the training set, is also the case that shows the highest proportion of errors. Here, again, the predictions of the Competition Model for the role of frequency of occurrence are supported.
3. **Children often omit the article.** Mills (1986) reports that early on in acquisition, children tend to omit articles. The network also exhibited this behavior. On average, 73% of the errors made by the network were omissions -- i.e. none of the articles reached the activation. The proportion of the total error that involved omissions is presented by case in Table 4.
4. **Children often overgeneralize one gender.** Second, Mills (1986) reports that in the early stages of acquisition children tend to overgeneralize feminine articles. This tendency also occurs early on in the simulations. This is evident in the error patterns for articles, shown in Table 4. The majority of the errors follow the feminine singular paradigm (*die* in the nominative and accusative, *der* in the dative and genitive). Neither the masculine nor neuter paradigms fit the errors as well as does the feminine paradigm.
5. **Children make early use of the -e cue.** MacWhinney (1978) and Mills (1986) report that children acquire the connection between the *-e* ending and feminine gender early on in acquisition. The network is also quick to acquire this correspondence. All the feminine singular items containing the *-e* ending were consistently assigned feminine articles even at this early point in learning. In addition, the error rate for the masculine and neuter nouns that had the *-e* ending was 71%. Less than half of these errors were omissions; all of the remaining errors were consistent with the feminine singular paradigm.
6. **Children make use of highly reliable cues.** In a similar vein, we found no errors at all for words marked with highly reliable cues to gender such as *-chen* and *-um*.
7. **Children can use cues to infer word classes.** The ability of the model to generalize the paradigm to new cases for old words indicates that it was able to use the exemplars given to determine gender. Although the model used no formal inferential logic, it was able to behave as if it was making this inference.

It is important for a model to match empirical data. However, it is more exciting when it makes predictions that have not yet been tested in empirical research. Our simulations made a distinct set of predictions of this type. A behavior that was exhibited by the model at all stages of learning before complete mastery was the confusion of definite articles when there is substantial paradigm overlap. This was especially serious for early learning. For example, every feminine plural in the dative case was either omitted or, more frequently, was assigned the feminine dative singular article *der*. This was true for *Leute*, *Fragen*, and *Minuten*. As far as we know, there is not yet any child acquisition data regarding this particular behavior. However, from the general analysis of paradigm acquisition in Slavic outlined by Slobin (1973), one would expect children to have problems with overlaps of this type. Given that paradigm overlap causes clear problems for the network, it would be revealing to see if it also causes problems for German children.

## Model 2

Model 1 was a success in many ways. It not only achieved perfect performance on the training set, it also showed generalization behavior quite comparable to that of a young German child. Model 1 succeeded in achieving many of the goals we had for this line of research. However, one can reasonably raise at least two objections against Model 1. First, one might argue that much of the power of the first model lay in the hand-crafting of a set of 38 phonological and semantic cues. This problem will be addressed in Model 3. A second, perhaps even more glaring, problem with Model 1 involves the use of the 11 arbitrary features which were used to reduce confusion between lexical items. As Pinker and Prince (1988) have noted, a major complaint that can be raised against connectionist models of morphological processing such as that of Rumelhart and McClelland (1987) is their inability to correctly represent those properties that are distinct to particular lexical items. One could defend the use of the 11 arbitrary features on the ground that that are place holders for a more complete semantic representation that is simply too time-consuming to produce. In any case, we need to understand more clearly how much of the correct behavior of the model in Model 1 is due to its use of the Zubin-Köpcke cues and how much is due to its reliance on the 11 arbitrary features.

To examine this issue, a second model was constructed that differed from the first in only one respect. In this second model, the 11 arbitrary feature units were removed. This made it possible for two different nouns to have identical representations. This would be the case for any nouns that happened to have identical gender and number cues. Of the 102 nouns used, 53 showed this type of overlap. In most cases, only a pair of nouns had the same feature representation. However, in some clusters there were up to 11 nouns with the same representation. Of these nouns, however, only 16 had a different gender and/or number from the norm for their group. Presumably, it is only this group of 16 nouns that could cause problems in training the network.

When the network was trained without the 11 arbitrary units, performance dropped significantly. The number of tokens missed (from the set of all nouns in all case contexts) rose from 202 for Model 1 after 50 epochs to 564 for Model 2 after 50 epochs and from 164 for Model 1 after 100 epochs to 563 for Model 2 after 100 epochs. Without the help of the 11 extra units, progress in learning basically came to a halt after only 50 epochs with still over 560 tokens being missed. These results indicate that the improved performance allowed by the extra 11 units corresponds to the network simply “memorizing” a mapping from these tags to the appropriate articles (they are random so there is no correlation between them and article type that the network can learn). Even nouns that did not require these units for disambiguation (all but 16) are free to use them in this way. The improvement that could have been gained by getting all tokens of the 16 ambiguous nouns

right as opposed to getting them all wrong, accounts for about 80% of the improvement that the extra 11 units provided. Thus the disambiguation units were providing some additional cues to the network apart from their role in lexical disambiguation. To the degree that learning was based on the use of these extra units, one can say that the network was not employing a very pure version of the cue-based process that it was designed to simulate. However, without these additional disambiguating units, performance was unsatisfactory. This problem of the selected cues not being sufficient for the task of determining an article for a given noun was the motivation for running Model 3.

### **Model 3**

A third model was constructed that was designed to address the two major limitations in Model 1. This model made no use of the 11 arbitrary disambiguating units in the first model. In addition, instead of relying on a hand-crafted coding of the Köpcke-Zubin features, it used the raw phonological form of the stem. All gender and number cues other than the five real semantic cues of Table 1 were eliminated. In their place, the network relied on a complete phonological representation of each noun.

### **Design of Model 3**

The design of Model 3 is given in Figure 2. There were three types of input units. Both the five semantic cues and the 17 explicit case cues were exactly as they had been in Models 1 and 2. The hand-crafted cues of Models 1 and 2 were removed. In their place, we used 130 units to represent the full form of the noun in actual phonological features. These 130 units were distributed over 13 slots with 10 features in each slot. The 10 features were standard phonological distinctive features such as [+labial], [+coronal], [+voice], [+high], etc. Diphthongs and affricates were coded as pairs of phonemes. Using these features, we found a unique 10-unit feature code for each German phoneme. The 13 slots we used were divided across five positions:

1. up to three consonants in the initial consonant cluster,
2. up to two vowels in the post-initial vowel nucleus,
3. up to three consonants in the medial consonantal cluster,
4. up to two vowels in the pre-final vowel cluster, and
5. up to three consonants in the final consonantal cluster.

The middle syllable of trisyllabic nouns was not coded. In the case of shorter words, some of the slots were sometimes left vacant. For example, if the word was a monosyllable, the medial cluster was left vacant and the pre-final and post-initial vowel nuclei were identical. Initial clusters were filled from the front, so that a single consonant in initial position would be followed by two empty slots. Similarly, a single consonant in final position would be preceded by two empty slots. In medial position, a single consonant was followed by two empty slots. If there was no final consonant, the final vowel was coded as being in the “pre-final” vowel slot. A single vowel was preceded by an empty slot. This positioning was chosen to maximize the predictiveness of the phonological cues. The 130 phonological units and the 5 semantic units all projected to a set of gender/number hidden units. The explicit case units and the phonological units for the final consonant of the stem all projected to a set of case hidden units. The phonological information was used to code the presence of noun-final markers for the genitive and dative plural. Both sets of hidden units projected to a third set of hidden units that then activated the six output units. Thus, apart from the input units, the design of Model 3 was identical to that of Models 1 and 2.

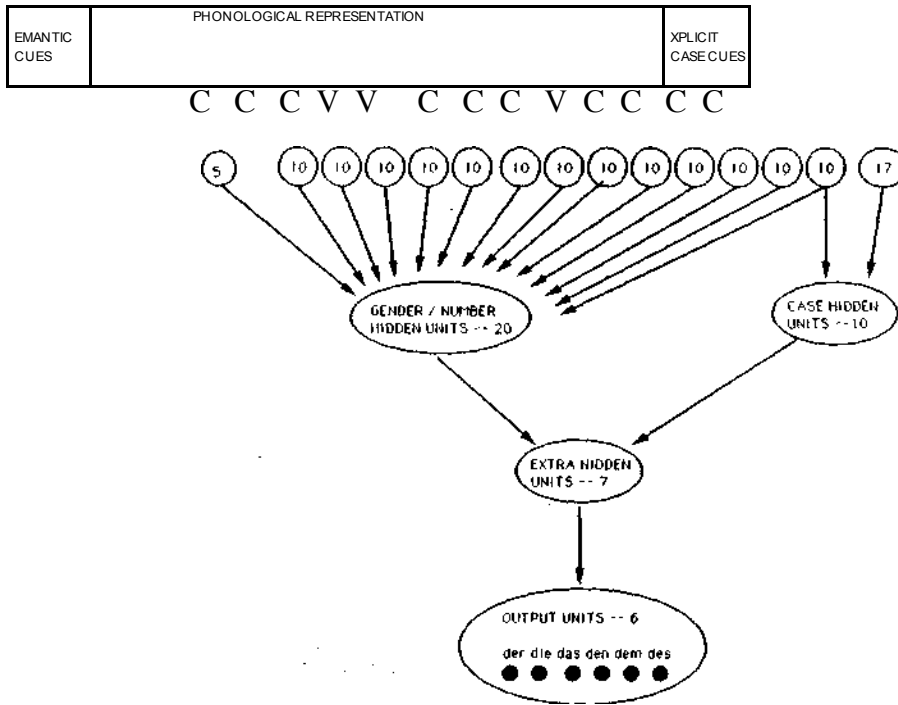


Figure 2: *Architecture for Model 3*. The main difference between this network and that given in Figure 1 is in the design of the input units. The letters at the top of the figure indicate the nature of the input units. The thirteen pools of 10 phonological units represent various positions in the words. The last phoneme of the word is connected to both the gender-number hidden units and the case hidden units, since this sound often bears information regarding case. The hidden units and the output units are the same as in Figure 1.

### Results for Model 3

Despite the larger number of input units that the new network had to encode, performance actually improved. Performance for Model 3 was markedly better than for either Model 1 or Model 2. After 50 epochs of the training set, the new network was missing only one word -- *der Stück* . At that point, it made only 131 errors on the generalization set of 1735 items. After 100 epochs, the new network had fully mastered the training set and made only 111 errors on the set of 1735 items. By comparison, after 50 trials, Model 1 was still missing about 7 training set words and made 202 errors on the generalization set. After 100 epochs, Model 1 made 164 errors on the generalization set. Thus, performance on Model 3 was markedly better than performance on Model 1 for both training and generalization at both 50 and 100 epochs. These results indicate that the hand-crafting of units in these simulations can actually be a disadvantage and that the networks themselves are often the best guides for the extraction of the correct higher-level cues, given an unambiguous set of input units.

### Discussion

These simulations open up important new ways of viewing the process of language acquisition. They provide us with new ways of understanding such fundamental concepts as

“rote,” “combination,” and “analogy.” These were the basic concepts in MacWhinney’s (1978) information-processing account of morphological learning in German, Hungarian, and English. Although the MacWhinney account made many successful predictions, it made liberal use of hand-crafted solutions to particular problems. By comparison, the connectionist networks developed here require a minimum of hand-crafting. Do these networks correctly deal with the various phenomena that originally motivated MacWhinney to postulate a system based on rote, combination, analogy, and paradigm-formation?

## **Rote**

MacWhinney (1978, c.f. Peters, 1983) placed heavy emphasis on the importance of rote-memorized forms as the basis of morphological learning. His account viewed the first words as unanalysed associations between sounds and meanings. When a child learned the word “dogs,” it was learned as a unit, not as a combination of “dog” and the plural. MacWhinney viewed these rote forms or “amalgams” as grist for the mill of morphemic segmentation and subsequent rule extraction. He did not view morphological items as separate unique nodes, but rather as patterns of associations between sound and meaning. In that sense, his views are compatible with connectionist accounts. Can connectionist accounts properly represent the important status of lexical items in morphology?

Recently, Pinker and Prince (1987) presented a critique of the Rumelhart and McClelland (1987) simulation of the acquisition of the past tense in English. One of the most important of the Pinker-Prince criticisms had to do with the role of the lexical item in connectionist architectures. Pinker and Prince correctly note that, in the Rumelhart and McClelland simulation, there was no way to distinguish between forms such as “ringed” and “wrung” vs. “wringed” and “rung,” because the present tense forms “ring” and “wring” had similar phonological representations, and there were no semantic representations to provide disambiguation. For German declension, Model 1 avoided this problem simply by introducing a set of arbitrary disambiguating features to provide a rough coding of the identity of the lexical item. Thus, Model 1 should be able to learn to generate both *der See* for “the ocean” and *die See* for “the lake,” since the two forms would differ in several of the 11 arbitrary features. However, Model 3 would have a more difficult time dealing with the two forms of *See*, since it codes lexical items chiefly in terms of phonological features and the two forms of *See* do not differ phonologically. In general, for these models to deal correctly with homophones that belong to different classes, all that is necessary is that there be some way in which the homophones can be coded distinctly. Certainly, real children make these distinctions and there is no reason, in principle, why such distinctions cannot be included in connectionist models.

We do not believe that the issue of the role of lexical rote in the acquisition of morphology is yet settled. The problems Model 1 had in acquiring forms like *Junge* and *Ende* may point to a role for rote. More generally, connectionist models may need to make reference to rote-like associations between patterns of sounds and meanings in order to avoid problems with interference (McCloskey and Cohen, in press) and with the general control of sequential processing (MacWhinney, 1987). On the other hand, the fact that Model 3 acquired these exceptional forms with comparatively little difficulty indicates that these models can simulate rote-like behavior, as long as the input includes sufficient cues to disambiguate forms. What we have shown in these simulations is that a connectionist network appears to be capable of handling both general patterns or analogies and exceptions or rote-learning within a single architecture. How far this unification of rote and analogy can be carried in an architecture that attempts to deal with more aspects of language processing remains a topic for future research.

## **Rules and Analogy**

The two other major processes envisioned by MacWhinney (1978) were “combination” and “analogy.” Combination involved the application of linguistic patterns or “rules” when morphemes are combined as suffixes, prefixes, or infixes. Whenever combination occurred, there was also an opportunity for rules or “patterns” to apply. For example, the rule voicing the /f/ of “wife” was seen as applying when the plural was combined with the stem. Both stems and affixes were stored in a fairly superficial form in terms of independent allomorphs. Although MacWhinney’s rules were simple and probabilistic, they were still independent cognitive entities and not distributed patterns.

MacWhinney’s third process -- analogy -- was much closer in spirit to the type of processing we see in connectionist networks. Analogy was particularly important in the accounting for the experimental data on the acquisition of German gender and number. However, in practice, MacWhinney never found a reasonable, non ad-hoc way of implementing analogy computationally. The program implementing the model was forced to break target stems into initial cluster and rhyming remainders and then randomly reattach initial clusters to the rhyming remainder in order to find a rhyming basis for the analogy. Of course, there was no independent evidence in the developmental data for any process of this type. The problem of implementing analogy is not unique to MacWhinney’s account. Both Hockett (1968) and Bybee (1985) saw lexical processing as essentially analogistic, but were unable to provide a clear specification of the bases for analogies in the general case. Indeed, it is commonly agreed that the major weakness in accounts that rely on analogy as a mechanism is that they fail to tell us which of several possible analogies should apply when. A fundamental problem with MacWhinney’s acceptance of both analogy and combination was the difficulty of deciding when a form was based on rule as opposed to analogy.

Connectionist models, such as the ones developed here, directly address both the problem of separating combination from analogy and the problem of specifying the bases of analogies. They address the problem of distinguishing combination from analogy by treating the two processes as equivalent. All word forms are produced in a single set of connections that simultaneously captures the processes of rote, combination, and analogy. These models address the problem of specifying the bases of analogies by treating coding features as fundamental to the model. Once the phonological features for German phonemes have been specified, most of the major analogy types simply fall out as patterns extracted by the back propagation algorithm.

### **Paradigms**

The fourth major dimension of morphological learning envisioned by MacWhinney was paradigm extraction. Paradigms were treated as a sub-case of the general strategy of combination. The idea was that, when the child comes to producing a form such as *der Mann*, he is combining the article with the noun. His main task in this case is to choose the correct form of the article. In order to do this, he makes reference to something like the paradigm given in Table 1. This is the structure that was used in the descriptive linguistics of the 1950’s (Hockett, 1954) under the name of an “item-and-arrangement grammar” In order to use such a table, the child has to know the gender, number, and case of the noun. The assumption is that number and case are given by other cues and that gender is a feature coded on each lexical item separately. Once a paradigm has been set up and nouns coded for class membership, processing is pretty straightforward. The problem to be solved is how children can 1) acquire the paradigm and 2) mark each noun for its class membership. MacWhinney (1978) thought of the child as learning such structures by repeatedly adding rows and columns to the matrix. The child may begin by thinking that all nouns in German are in one class. With time, however, he breaks up the noun class into three classes that are then three columns in a paradigm. Pinker (1984) presented an algorithm for the splitting of

rows and columns that was essentially the same as that proposed by MacWhinney (1978). The problem with the MacWhinney-Pinker proposal is that the strategies involved in the actual formation of the paradigm are entirely *ad hoc* and have no independent support from acquisitional data. There are also a variety of additional technical problems with the Pinker proposal that have been noted by Braine (1987). Pinker attempted to use paradigm formation as a uniform solution for all of morphological processing. However, as Braine has noted, doing this makes it impossible to properly understand the processes of segmentation and analogy that motivated the fuller structure of the MacWhinney model. For example, Pinker's account relies on the extraction of a single stem allomorph. However, with something like the Latin noun for "soldier," there is both the nominative stem *mil=es* and the oblique stem *milit-* of *militem*. However, a more general problem with the Pinker approach is that it leaves unspecified which paradigms should apply to which forms at which point in either learning or processing. This is the same basic problem with expressing similarity that besieged the MacWhinney approach.

Connectionist models provide an alternative to the strategies for paradigm extraction proposed by MacWhinney and Pinker. The solution establishes no formal classes apart from the patterns of associations within the network. The problem of deciding which of several competing paradigmatic types should apply to new forms is determined entirely by similarity match. What is crucial in the new account is the way that it uses cues both as a basis for direct prediction of gender and as a way of controlling the organization of paradigm-like information. The claim is that *paradigms emerge from associations between cues*.

## Conclusions

The model provided a good match to the currently available data on the acquisition of the declension of the definite article in German. It matched the reported omission of articles in early acquisition and the tendency to overgeneralize the feminine. It showed strong learning of the *-e* cue and later learning of other reliable cues. It showed a clear ability to use its internalized "knowledge" of the paradigm to assign gender to new words. Beyond mastering the problem of German definite article assignment, the simulations presented here are in accord with general patterns in the empirical literature. As discussed in a separate analysis of the Model 1 data by Taraban, McDonald, and MacWhinney (1989), the strength of the connection between a cue and a category changes during the learning process -- cue strength first follows overall cue validity, then reliability, and finally conflict validity. Similar shifts have been found in other cue-category learning situations (McDonald, 1986; McDonald and MacWhinney, 1987; Sokolov, in press).

The model went beyond generating a simple match to already known facts. It also generated a number of predictions that can be tested in future developmental research. It predicts strong difficulties with words like *Junge* and *Ende* that are exceptions to powerful cues. It predicts a confusion between *der* and *den* as markers of the dative plural. Finally, it also predicts fairly uniformly incorrect treatment of new nouns for which it has inferred the wrong gender. The ability of the model to generate clear new predictions for developmental research is an important strength and one not found in earlier accounts.

In terms of acquisitional theory, the model provides the first interesting alternative to the information-processing account of morphological learning presented in MacWhinney (1978) and then 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, we can now think about generalization in terms of cue acquisition. By substituting the phonological cues of Model 3 for the hand-crafted cues of Model 1, the model makes an empirically solid move toward an account that can be generalized across languages.

The success of the current model for this particularly difficult problem in language learning would seem to indicate that claims regarding the insufficiency of connectionist accounts for language learning (Pinker and Prince, 1988) are, to say the least, premature.

## REFERENCES

- BATES, E., & MACWHINNEY, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*, Hillsdale, NJ: Erlbaum.
- BATES, E., & MACWHINNEY, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art*, New York: Cambridge University Press.
- 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.
- BRUNSWIK, E. (1956). *Perception and the representative design of psychology experiments*. Berkeley, California: University of California Press.
- BYBEE, J. (1985). *Morphology*. Amsterdam: Benjamins.
- BYBEE, J. L., & SLOBIN, D. I. (1982). Rules and schemas in the development and use of the English past. *Language*, 58, 265-289.
- CHOMSKY, N., & HALLE, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- ERVIN-TRIPP, S. (1966). Language development. In *Review of child development research*, Ann Arbor: University of Michigan Press.
- GREBE, P. (1973). *Duden Grammatik der deutschen Gegenwartssprache*. Mannheim: Dudenverlag.
- HOCKETT, C. (1968). *The state of the an*. The Hague: Moulon.
- HOCKETT, C. (1954). Two models of grammatical description. *Word*, 10, 210-231.
- KAIL, M. (1989). Cue validity, cue cost, and processing types in French sentence comprehension. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.
- KOPCKE, K. (1988). Schemas in German plural formation. *Lingua*, 74, 303-335.
- KOPCKE, K., & ZUBIN, D. (1983). Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache. *Zeitschrift für germanistische Linguistik*, II, 166-182.

- KOPCKE, K., & ZUBIN, D. (1984). Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*, 93, 26-50.
- KOPCKE, K.-M. (1982). Untersuchungen zum Genusssystem der deutschen Gegenwartssprache. Tübingen: Niemeyer.
- KUCZAJ, S. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16, 589-600.
- LEDERER, H., SCHULZ, D., & GRIESBACH, H. (1969). Reference grammar of the German language. New York, Scribner.
- LEVY, Y., SCHLESINGER, I., & BRAINE, M. (1987). Categories and processes in language acquisition. Hillsdale, NJ: Erlbaum.
- MACWHINNEY, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, 43, whole No. 1.
- MACWHINNEY, B. (in press). Competition and lexical categorization. In R. Corrigan (Ed.), *Categorization*. New York: Benjamins.
- MACWHINNEY, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*, Hillsdale, NJ: Erlbaum,
- MACWHINNEY, B. (1988). The Competition Model and teachability. In M. Rice & R. Schiefelbusch (Eds.), *The teachability of language*. New York: Cambridge University Press.
- MACWHINNEY, B., & BATES, E. (Eds.), (in press) *The cross-linguistic study of sentence processing*. New York: Cambridge University Press.
- MACWHINNEY, B., BATES, E., & KLIEGL, R. (1984). Cue validity on sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, 23, 127-150.
- MACWHINNEY, B., PLEH, Cs., & BATES, E. (1985). The development of sentence interpretation in Hungarian. *Cognitive Psychology*, 17, 178-209.
- MACWHINNEY, B., & SNOW, C. (1985). The child language data exchange system. *Journal of Child Language*, 12, 271-296.
- 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: Volume 2*, New York: Gardner.
- 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.

- MCDONALD, J. (1989). Determinants of the acquisition of cue-category mappings. In *The crosslinguistic study of sentence processing*, New York: Cambridge University Press.
- MCDONALD, J. L. (1986). The development of sentence comprehension strategies in English and Dutch. *Journal of Experimental Child Psychology*, 41, 317-335.
- MCDONALD, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, 8, 379-415.
- MCDONALD, J. L. (1987). Sentence interpretation processes: The influence of conflicting cues. *Journal of Memory and Language*, 26, 100-117.
- MCDONALD, J. L., & MACWHINNEY, B. (1987). Levels of learning: A microdevelopmental study of concept formation. Manuscript, Carnegie Mellon University.
- MEIER, H. (1967). *Deutsche Sprachstatistik*. Hildesheim: George Olms.
- MILLS, A. E. (1986). *The acquisition of gender: A study of English and German*. Berlin: Springer-Verlag.
- MUGDAN, J. (1977). *Flexionsmorphologie und Psycho-linguistik*. Tübingen: Gunter Narr.
- PARK, T.-Z. (1974). *A study of German language development*. Unpublished manuscript, Psychologisches Institut Bern.
- PINKER, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*, Hillsdale, NJ: Erlbaum.
- PINKER, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University 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, 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: MIT Press.
- RUMELHART, D., & MCCLELLAND, J. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processes? In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*, Hillsdale, NJ: Erlbaum.
- SLOBIN, D. (1971). *Psycholinguistics*. Glenview, IL: Scott, Foresman & Company.
- SLOBIN, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart, & Winston.

- SOKOLOV, J. (in press). The development of role assignment in Hebrew. In B. MacWhinney & E. Bates (Eds.), *The cross linguistic study of sentence processing*. New York: Cambridge University.
- SOKOLOV, J. L. (1988). Cue validity in Hebrew sentence comprehension. *Journal of Child Language*, 15, 129-156.
- TUCKER, G., LAMBERT, W., & RIGAULT, A. (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Moulon.
- WANGLER, H. H. (1963). *Rangworlerhiich hochdeut-scher Umgangsprache*. Marburg: Elwert.
- ZUBIN, D. A., & KOPCKE, K. M. (1981). Gender: A less than arbitrary grammatical category. In R. Hendrick, C. Masek, & M. F. Miller, (Eds.), *Papers from the seventh regional meeting*, Chicago: Chicago Linguistic Society.
- ZUBIN, D. A., & KOPCKE, K. M. (1983). Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. In C. Craig (Ed.), *Noun classes and categorization*, Eugene, OR: Proceedings of a symposium.
- ZUBIN, D. A., & KOPCKE, K. M. (1986). Gender and folk taxonomy: The indexical relation between grammatical and lexical categorization. *Typological Studies in Language* (Vol. 7). Amsterdam/ Philadelphia: Benjamins.

(Received July 26, 1988)

(Revision received January 13, 1989)