AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE

General Editor
E.F. KONRAD KOERNER
(University of Ottawa)

Series IV - CURRENT ISSUES IN LINGUISTIC THEORY

Volume 61

Roberta Corrigan, Fred Eckman and Michael Noonan (eds)

Linguistic Categorization

LINGUISTIC CATEGORIZATION

Edited by

ROBERTA CORRIGAN, FRED ECKMAN
and
MICHAEL NOONAN

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

1989

CATEGORY LEARNING IN A CONNECTIONIST MODEL: LEARNING TO DECLINE THE GERMAN DEFINITE ARTICLE*

ROMAN TARABAN, JANET L. MCDONALD,
and BRIAN MACWHINNEY
Department of Psychology
Carnegie Mellon University

When linguists write a grammar for a language, they do it by formulating a set of rules together with a set of exceptions to these rules. Chomsky & Halle (1968) provided a generative grammar for English derivational and inflectional morphology containing less than a score of rules. But many of the rules in that grammar were quite complex, having numerous subconditions and referring to highly abstract underlying forms. For German morphology, Mugdan (1977) presents an even longer set of rules and branching conditions. For each of their many rules and subconditions, both Chomsky & Halle and Mugdan list a set of exceptions to the rule. Sometimes the exceptions to the rule are also formulated in terms of yet another rule.

If we think of language acquisition as the process of acquiring a grammar, then it would seem logical to think of the child as learning a set of rules together with a list of exceptions to those rules. However, psychologists have often 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 (Ervin-Tripp 1966; Slobin 1971). Nor is there good direct evidence that children pick up rules in any simple manner during learning. How then is it that children come to behave as if they knew the rules, even without learning the rules as discrete entities? Does the mechanism underlying this learning closely resemble the rule itself or does it look like something very different?

In this paper we examine this question in relation to the learning of the declensional paradigm of the definite article in German. This declensional paradigm is dependent on three different categorizations of the German noun: gender (masculine, feminine, neuter), number (singular, plural) and case (nominative, genitive, dative, accusative). The actual marking for these categories is given by the form of the articles and adjectives that precede the noun. The noun itself is marked for number and only infrequently for case. Crossing the categories of gender, number, and case yields 24 possible cells in the full German declensional paradigm. Fortunately for the German child, gender distinctions disappear in the plural, reducing the paradigm to 16 distinct cells. The distribution of German definite articles in this paradigm is shown in Table 1A. German has only six different definite articles (*der, den, dem, des, die, das*). Each article occurs in at least two different cells of the paradigm, so that no article can define 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.

Table 1a. *German definite articles — gender, number and case paradigm*

Case		Gender		Plural
	Masculine	Feminine	Neuter	all genders
nominative	der	die	das	die
genitive	des	der	des	der
dative	dem	der	dem	den
accusative	den	die	das	die

Table 1b. Expanded portion of paradigm

	Masculine Singular				
nominative	"der"	+ definite reference	+ (gender cues)	+ (number cues)	+ (case cues)
genitive	"des"	+ "	+ "	+ "	+ "
dative	"dem"	+ "	+ "	+ "	+ "
accusative	"den"	+ "	+ "	+ "	+ "

Acquisition of this system is not a trivial task. Overtly, there are two major problems to be solved. The first problem is the classification of nouns into the three gender categories. The second problem is the construction of the complete paradigm and the assignment of various forms of the definite article to the different slots in the paradigm. We can think of these two problems as the word class formation problem and the paradigm formation problem. As we shall see, the solution to these two problems cannot be pursued independently, and the solution to the overall problem depends critically on cue learning. The remainder of this paper focuses on cue learning and cue use. We begin with a discussion of the shape of the cues available to the child and then turn to two simulations of the acquisition and use of these cues.

1. Cue learning and cue competition

The simplest way to solve the word class formation problem 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. There is actually a large and powerful set of cues to German gender (Zubin & Köpcke 1981; Zubin & Köpcke 1986; Köpcke & Zubin 1983, 1984). These cues, examples of which are given in Table 2, occur on many different levels, including the phonological, morphological, and semantic levels. 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 monosyllabic nouns that start with the phonological sequence /šC/ (i.e. š + any 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.

Unlike gender, which has no single real-world correlate, the other two dimensions of the paradigm — number and case — are used to express real semantic facts. However, when information about these facts is absent, the child can still use certain other cues to predict number and case. The plurality of German nouns is marked by one of six morphemes shown in Table 3. The distribution of these plural morphemes over the nouns of German is rather complex — 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"

Table 2. *A selection of cues to gender*

	Major associated gender	*Example*	*Translation*
Phonological cues			
Š + consonant	masculine	*der Schrank*	closet
-fricative + t	feminine	*die Nacht*	night
Morphological cues			
-el	masculine	*der Schlüssel*	key
-ling	masculine	*der Feigling*	coward
-e	feminine	*die Sonne*	sun
-ung	feminine	*die Zeitung*	newspaper
-lein	neuter	*das Fräulein*	young woman
-ment	neuter	*das Instrument*	instrument
Semantic cues			
natural gender	masculine	*der Sohn*	son
weather phenomena	masculine	*der Regen*	rain
natural gender	feminine	*die Tochter*	daughter
flowers	feminine	*die Tulpe*	tulip
youth	neuter	*das Kind*	child
metals	neuter	*das Gold*	gold

Table 3. *Cues to number*

	Examples		
Plural morphemes	*Singular*	*Plural*	*Translation*
1. -e	*Tag*	*Tage*	days
2. -(e)n	*Blume*	*Blumen*	flowers
3. -er	*Kind*	*Kinder*	children
4. -s	*Radio*	*Radios*	radios
5. 0	*Zimmer*	*Zimmer*	rooms
6. ¨	*Bruder*	*Brüder*	brothers
Combination of plural morphemes			
1 & 6 ¨e	*Hand*	*Hände*	hands
3 & 6 ¨er	*Mann*	*Männer*	men

is *die Hüte*. However, 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).

Cues to case in German occur on the morphological, syntactic and semantic levels. Morphological cues include the *-(e)s* ending added to singular masculine and neuter nouns 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, noun-verb agreement (i.e. a noun that agrees in number with the verb is a potential subject (nominative case), while one that disagrees must be in a different case), 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").

In light of what has been said so far about cues, the cues supporting selection of the masculine articles in Table 1a might be depicted as shown in Table 1b. This represents the masculine category as a set of forms, with each form associated with a meaning and a particular pattern of gender, number and case cues. The simple claim that we wish to make is that learning the definite article system involves learning the relationships between the forms of the article and the gender, number and case cue contexts in which they are used.

Although the general claim that we are making is simple, the underlying cue-to-category relationships are not simple. This is because the mapping between cue and category is many-to-many. Any particular cue can map to many categories. For example, 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 phonological /šC/ cue, which is indicative of masculine gender, and also has the semantic cue of alcoholic beverage, also indicative of masculine gender. However, because cues are not completely reliable, multiple cues on an item need not agree about the categorization. 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 & Kliegl 1984; MacWhinney 1987; Bates & MacWhinney 1987; McDonald, 1989). 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 secondly, with its premise that cues contribute to the categorization process in proportion to their relative informational value.

The Competition Model provides a number of properties of cues and a means of analyzing and talking about them. Four properties that were used to assess cue weights in our simulations are *availability*, *reliability*, *validity*, and *conflict validity*. First, any particular cue may or may not be present on a particular item. This property has been defined in the Competition Model as the *availability* of the cue. For example, although some feminine nouns are marked by the *-e* ending, not all feminine nouns contain this cue. *Availability* specifies how often a cue appears. Second, even when a cue is present, it may not always indicate the same category. Therefore, it will not be perfectly *reliable* in indicating the correct category. For example, 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. Third, the *validity* of a cue (Brunswik 1956) specifies how often a cue is both present to signal a categorization choice and also correct about the choice that it signals. It thereby combines the properties of availability and reliability. Finally, there is a property of *conflict validity*, which specifies how often a cue is present and correct in cases containing conflicting cues. That is, cues are not described simply in terms of the distributions between associated categories (reliability and validity), but also in terms of the conflicting cues with which they cooccur (conflict validity).

One purpose of the simulations was to see whether these properties of cues from the Competition Model could be used to describe the *strength* of cues in the simulation. This was done with a number of empirical results as background. 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 or to level all articles to a single gender. Mills (1986) reports that children often pursue both of these strategies. 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. Also, children show some tendency to generalize the feminine gender. Mills (1986) examined the role of the Köpcke-Zubin cues in the acquisition of German gender and found strong evidence for their use. 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. MacWhinney also found that children between the ages of 4 and 6 were able to make correct use of the morphological marking *-erei* 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. A general finding was that cue strengths depended on the *validity* of a cue, for morphological categories (MacWhinney 1978) and syntactic categories (MacWhinney; Bates & Kliegl 1984 McDonald 1986). There was also evidence that the relation between cue validity and cue strength could change during the course of acquisition, with initial stages dependent on *cue validity* and later stages on *reliability* (Sokolov 1988) or on *conflict cue validity* (McDonald 1986; McDonald & MacWhinney 1987). These empirical results will be used as a further test of the simulation results.

2. Paradigm formation

The language learner must eventually extract the paradigm given in Table 1. Control of the paradigm is indicated by performance on novel instances. If part of the paradigm were extracted, the child could use it in the following way: if a new noun appeared with the article *den* and appeared to be the object of the verb, then it would be reasonable to assume that it is a masculine noun. The child could then take that noun and use it with the article *der* in the nominative. The inference would be even more reliable if the noun were referred to by the masculine pronoun *ihn*. It was this latter type of inference that MacWhinney (1978) examined in his study of the acquisition of German gender. He found that 4-year-old children were able to make reliable use of the pronoun as a cue to the gender of nonce words. If he told children "I am picking him (*ihn*) up in my hand," children knew 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. As MacWhinney (1978) has argued, German-speaking children are able to assign words to classes by making inferences based on the declensional paradigm.

Thus, there is good evidence for paradigm-based inference in German children. But there is a problem with MacWhinney's account. In order to perform paradigm-based inference, the child must have extracted the paradigm and the ways in which this is done are not entirely clear. On one view, a paradigm is a fairly formal structure. It is an n-dimensional matrix in which the values along each dimension represent values along some domain of linguistic categorization. This is the structure that was used in the descriptive linguistics of the 1950s (Hockett 1954) under the name of an "item-and-arrangement grammar." 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 which are then three columns in a paradigm. Pinker (1984) presents an algorithm for the splitting of rows and columns that is 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.

Thus, the acquisitional work provides us with a paradox. On the one hand, it is clear that German children can make inferences that appear to be based on something like a paradigm. On the other hand, we really have no idea how they might learn the paradigm. In the current paper we explore a new approach to the problem of the learning of word classes and paradigms. We will argue that this new approach preserves the insights of MacWhinney (1978) and subsequent work, while avoiding the major pitfall of that earlier study by incorporating the ground-breaking work on gender cues by Köpcke & Zubin. The new approach provides an account of cue-based learning and inference without having to propose *ad hoc* strategies for paradigm formation. 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 on the basis of associations between cues*.

3. Learning in a connectionist architecture

The new approach we will examine was inspired by a simulation of the acquisition of the English past tense by Rumelhart and McClelland (1986), who modeled the developmental data of Bybee and Slobin (1982) and Kuczaj (1977). The Rumelhart and McClelland model, as well as the model that we used for our simulations, are most accurately described in terms of the computational framework of connectionist networks.

Connectionist networks typically consist of large numbers of densely interconnected elements (represented in Figure 1 as circles) that operate 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 (represented as lines between the circles in Figure 1). A network can have internal layers of nodes (the middle layer in Figure 1), in addition to an input layer (the lowest layer in Figure 1) and an output layer (the topmost layer in Figure 1). The number of nodes at each level, as well as the pattern of interconnections between nodes, are free parameters that must be set in advance of a simulation. On any given trial of the network, each connection in the network assumes a scalar value

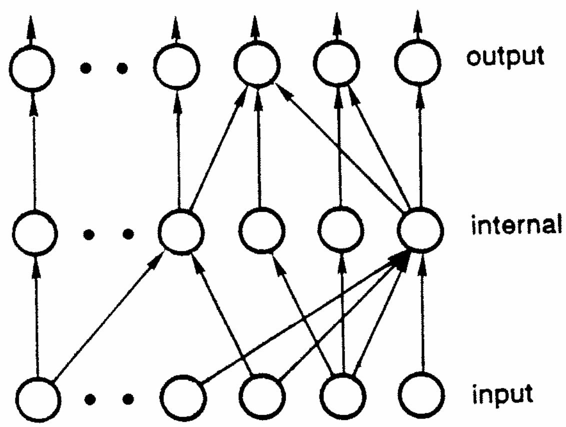

Figure 1. *A connectionist network.*

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 two simulations 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, either in a sentential context (Simulation I) or by itself (Simulation II). We have already described what these cues were like and we will discuss them in more detail as we describe each simulation 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 the German articles. There were as many 'correct' patterns as there were acceptable articles for the simulation (six in Simulation I: *der*, *den*, *dem*, *des*, *die*, *das*; and three in Simulation II: *der*, *die*, *das*), which were each represented by a unique output unit 'on' and the remaining output units 'off.'[1] Thus, correct performance required a stringent pattern of activation over the output units. In addition, the patterns of activation on the internal layer(s) 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 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 same learning algorithm, which is the Back Propagation algorithm of Rumelhart, Hinton & Williams (1986). The learning was therefore consistent for all connections in the network, with no *ad hoc* intervention into the learning process.

In this paper 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.

4. Simulation I

The first simulation 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.

The network was repeatedly trained on a training 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. Knowledge of the paradigm in Table 1 and of word classes gained by this learning was then assessed with two tests. In one test, we saw how well the network learned the nouns in the case contexts they appeared in in the training set and also how accurately it assigned articles to these nouns when they appeared in case contexts that the noun did not appear in in the training set. 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 totally new words.

4.1 *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.

Each of the 102 nouns in the training set was coded for the presence or absence of 39 gender cues. Each token of a noun appeared with the same cues. These phonological, morphological and semantic cues, given in Table 4, were taken primarily from the work of Zubin & Köpcke (Zubin & Köpcke 1981, 1986; Köpcke & Zubin 1984), with some additions from a German grammar (Lederer, Schulz & Griesbach 1969). In addition, each of the 102 nouns was given a unique code by turning on a different combination of 4 out of 11 additional input nodes. These additional nodes were meant to distinguish nouns that were identical on all of the 39 cues. This unique code can be thought of as a crude way of encoding all the remaining information not encoded in the original 39 cues.

Each of the 305 tokens in the training set was presented in a case context. These contexts were selected to approximate the frequencies with which the various cases occur in German: 41.6% nominative case, 24.1% accusative case, 24.9% dative case and 9.4% genitive case (Meier 1967). Twenty unique case contexts were prepared: 8 nominative, 5 accusative, 5 dative and 2 genitive. These case contexts consisted of 19 cues, which included 2 case endings (i.e. *-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). The 20 case contexts were randomly assigned to tokens of a noun, 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.

The architecture chosen for the network consisted of two internal layers in addition to the input and output layers. We chose to use two internal layers for reasons that are not theoretically important to our results, and we have evidence that our results would have been similar with a single internal layer. The network learned according to the procedure outlined above. The goal of the learning sequence was to adjust the weights so that, after repeated exposures to the input, the combination of input cues present on each word would come to reliably activate the associated definite

Table 4. *Gender and number cues used in Simulation I*

Phonological or word structural cues
umlauting
tr- or dr- initial segment
š + consonant- initial segment
-fricative + t ending segment
one initial consonantal phoneme
two initial consonantal phonemes
three initial consonantal phonemes
one final consonantal phoneme
two final consonantal phonemes
three final consonantal phonemes
monosyllabic word

Endings

-e	-ei	-el	-en
-ent	-er	-(e)s	-ett
-eur	-ie	-ik	-in
-ion	-ität	-ling	-ment
-n	-sis	-um	-ung
-ur			

diminutive endings *-lein* or *-chen*

Semantic and derivational cues
natural masculine gender
natural feminine gender
young or immature being
superordinate
noun derived from a verb
noun created from a number

article as the output. We could test the network at any level of mastery using the two tests described above; here we will limit our discussion of these tests to performance after the initial learning set was completely mastered.

5. Results for training set items in Simulation I

This simulation was run 20 different times always using the same learning 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. We are concerned here with patterns of performance, although it may be that a learning rate (a parameter in Back Propagation networks) could be found that corresponds to performance at various ages.

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 to us 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 — 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 fifteen 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 9.4% 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, in addition, be more 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 for feminine singulars and for plurals is in the articles for dative. Given that singulars are more highly represented, there was a tendency to treat some instances of plurals as 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.

5.1 *Generalizing the paradigm to old nouns in new contexts*

On the basis of the results so far, one could argue that the network simply develops 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 three contexts in this test. If a word had occurred with only one 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 this 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, with 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. In fact, it assigned the correct definite article to 92% of these new items, on average (range = 89-95%), computed over five consecutive simulation runs.

Many of the errors that the network made were caused by ambiguity in the information it was originally given in the training set. This ambiguity was caused by the paradigm overlap discussed above. For example, if a noun occurred in the training set in the nominative with *die*, accusative with *die*, and genitive with *der*, but did not occur in the dative case, 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.

5.2 *Generalizing the paradigm to new nouns*

Productive language use depends on detecting some regularity in the structure of a language, and then applying it to novel instances. The paradigm illustrated in Table 1b suggests that the regularity detected is at the level of cues for gender, number and case, in accord with the Competition Model. There is an alternative model, though, for the generalization results above. This is the Maratsos & Chalkley (1980) model, which suggests that the regularity is at the cue level for number and case, but at the word level for gender. On that model, correlated patterns of *form* + *meaning* + *number* + *case context* linked to sets of nouns that share these patterns underlie knowledge of the paradigm. That is, children learn sets of correlated patterns centered around nouns. On the Maratsos & Chalkley account, knowledge of a novel noun plus article in one of the contexts is sufficient to give knowledge of the others. For example, if a child knew that *Mann* "man", along with other words, takes *der* in a nominative context, *den* in an accusative context, *dem* in a dative context, and *des* in a genitive context, and also knew that *Lehrer* "teacher" takes *den* in an accusative context, the child could provide the correct articles for *Lehrer* in the remaining case contexts.

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 & 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 new words. Given the tenets of the Competition Model, it was expected that generalization would be most successful with those cues that were high in information value.

The results of the test showed that the network was able to successfully generalize the cues to new words. Using the weights from the 5 consecutive runs used for the previous test, we found that the network assigned the correct definite article and activated no incorrect article on 61% of the new instances, on average (range = 59-63%). Since there are six definite articles, chance performance on this task would be 16% correct, if one considered only the output unit that had to come on, and less than 2% correct if one considered the entire activation pattern for all six output units. 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.

5.3 *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%). Table 5 shows the average proportion of tokens in each case that were errors (i.e. errors/total number of tokens in that case), for 4 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 shows the lowest proportion of errors, while the genitive case, which is the least represented, shows the highest proportion of errors. The accusative case falls

into this pattern as well, although the dative case does not. We interpret this as an availability (frequency) effect for case learning. As discussed above, genitives are the latest learned items, which is consistent with error rates at this much earlier stage of learning, and which is the outcome of the availability of genitive tokens during learning.

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 level of activation required to be considered 'on.' The proportion of the total error that was omissions is presented by case in Table 5.

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 5. 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 the feminine paradigm does.

Finally, 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

Table 5. *Simulation I: Percent errors in early learning ('Case' indicates the correct case for the article)*

	Case Nominative	Accusative	Dative	Genitive
Total proportion of errors	42%	49%	26%	70%
Omission	34%	44%	9%	40%
Incorrect Article				
der	2%		13%	18%
die	4%	4%		
das	2%	1%		
dem			4%	12%
den				
des				

71%. Less than half of these errors were omissions; all of the remaining errors were consistent with the feminine singular paradigm.

A behavior that was exhibited by the model at all stages of learning before complete mastery is 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 can tell, there is no child acquisition data on this particular behavior. Given that paradigm overlap causes problems for the network, it would be interesting to see if it also causes problems for children.

6. Simulation II

Overall, the behavior of the network in Simulation I was quite impressive. The network demonstrated an ability to fill in missing paradigm slots for nouns that did not appear in all cases during the training sequence. It was also able to transfer its knowledge with moderate reliability to totally new nouns. If this performance is interpreted using the Competition Model, then the knowledge of the network is in the information value of the cues, and we should be able to probe this knowledge using the information metrics of the Competition Model. The second simulation was designed to get a more detailed picture of what factors determine the strength of cue usage during acquisition at various levels of mastery. This simulation simplifies the acquisitional problem to that of gender acquisition. We track the development of cue strengths with respect to internal representations in the network, and with respect to overt performance, using the information value constructs from the Competition Model.

For this simulation, we used a network with a single internal layer, in addition to the input and output layers. Every element in the input layer connected to each of three internal units, and each of the internal units connected to each of three output units. The input layer was used to represent phonological, morphological and semantic cues associated with the gender of German nouns. These cues, as in the first simulation, came from the work of Zubin & Köpcke (Zubin & Köpcke 1981, 1986; Köpcke & Zubin 1984), and from Lederer et al. (1969).

As in the previous simulation, there was a single node on the input

layer for each particular cue that we used. The output layer was used to represent the three nominative singular articles *der*, *die*, or *das*. The articles were coded by an output pattern of one unit on and the other two off. These patterns of activation for the outputs were decided in advance and were the 'desired' outputs during the learning phase, as in Simulation I. Since we will be concerned with the representations on the internal layer in this simulation, we should note that no 'desired' output was associated with these units. Patterns of activation over these internal units evolved without tutelage over the course of learning as a result of weight changes based on the Back Propagation algorithm. There were a total of 150 nouns in the training set; 75 of these came from the examples provided by Zubin & Köpcke for their cues, and another 75 were drawn for the most part from

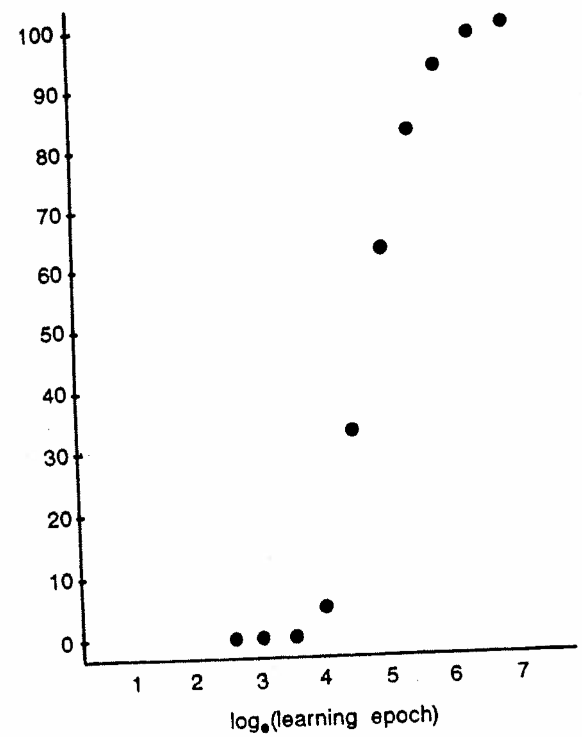

Figure 2. *Percent correct of learning set as a function of log of learning trials.*

an article picked at random in the Medicine section in *Der Spiegel*, under the constraint that no two words were chosen that had exactly the same cues but different outputs associated with them.

6.1 *Basic results*

The results of this simulation were straightforward. An examination of Figure 2 shows that accuracy increased monotonically with the natural log of the training epoch, where one epoch is a single presentation of each word in the training set in random order, until the network achieved 100% performance.[2]

At the point when the network achieved 100% performance on the training set, the pattern of activation over the three internal units was examined as each input pattern was presented to the network. Three distinct patterns were evident. For masculine nouns, the pattern of activation was unit 1_{on} unit 2_{on} unit 3_{off}, for feminine nouns it was unit 1_{off} unit 2_{on} unit 3_{on}, and for neuter nouns it was unit 1_{on} unit 2_{off} unit 3_{on}.

6.2 *The impact of cue validity on internal representations*

A major reason for conducting this simulation was an interest in an answer to the question, 'What is the relationship between internal representations and the cues that produced them?' In answering this question, we used constructs from the Competition Model, which were interpreted as follows for this simulation:

- cue availability: the percentage of times that a cue is present in the training set
- cue reliability: a cue's contribution to each of the internal gender representations (at the point of mastery), computed only over those cases in which the cue is present
- cue validity: the multiplicative product of availability and reliability

We also looked at cooccurrence relationships among the cues, that is, the behavior of a cue in the context of the other cues with which it occurs. Below we relate these co-occurrence relationships to another construct of the Competition Model, *conflict validity*.

Each weight connecting a cue to the internal units was correlated with its associated cue validity and cue reliability values, with the results plotted in Figure 3. The rightmost points indicate the correlation values when the

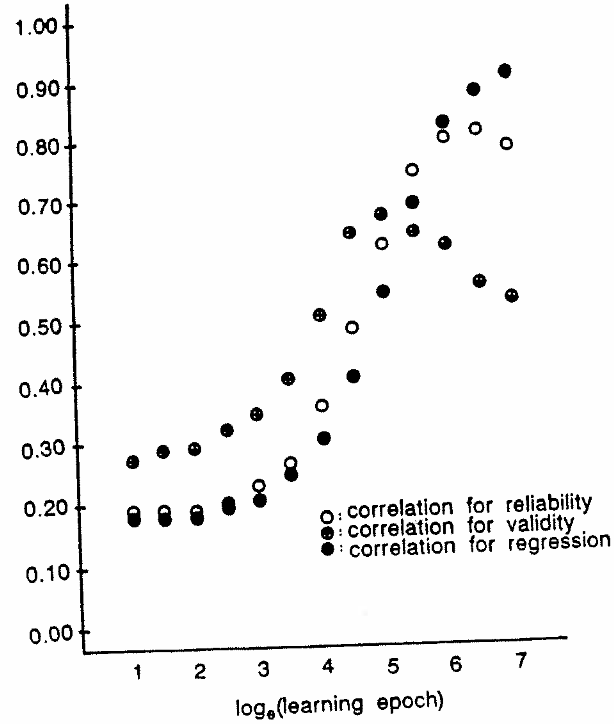

Figure 3. *The correlation of network weights and validity, reliability and regression values (y axis) at various points in learning (x axis).*

network had achieved 100% performance, and the points to the left of these indicate the correlation values for earlier levels of performance (The correlation values for earlier levels of performance are for the existing weights at a particular level of performance and the gender representations at the 100% mastery level). The general pattern that is evident here is that cue validity, which takes into account the frequency of a cue as well as its reliability is more strongly correlated to the weights — the knowledge that the cues carry — early in learning, but as learning progresses, pure reliability becomes a stronger predictor of the weights. This suggests that, with continued learning, the network becomes increasingly sensitive to the inherent usefulness of a cue for predicting the right answer, regardless of how often the particular cue might be used. This general pattern has in fact been reported in some recent empirical work (Kail 1989; Sokolov 1988).

One type of information that simple cue reliability measures do not include is cooccurrence information about any particular cue. This type of information could be important for a number of reasons. For example, two cues that cooccur and agree on an output can share the burden of activation, resulting in weights lower than those that a reliability metric would suggest. Alternatively, two cues that cooccur and disagree on an output could result in a less reliable cue taking on a larger burden than the reliability metric would suggest, for the sake of the correct output. The third curve plotted in Figure 4 shows the relationship of regression values for each cue and the actual cue weights in the network. These regression values — ßs in standard terminology — were computed using the cues as independent variables and the gender representation associated with a set of input cues as the dependent variable (the gender representations were the internal representations described above). The ßs were intended to capture the co-occurrence relationships in the training set. Figure 3 shows that reliability is a better predictor of weights for a long stretch of learning, but that late in learning the ßs become better predictors. Cue co-occurrence, especially cases where cues are in conflict, has been shown to be important in the later stages of acquisition in other linguistic and non-linguistic tasks (McDonald 1986, McDonald & MacWhinney 1987).

6.3 *The impact of cue validity on new words*

These results showed that we could interpret the relationship between cues and their internal representation using 'information' metrics from the Competition Model. We then asked whether these metrics could be used effectively to predict the outputs for novel test words. These test words were selected based on the reliability and cooccurrence relationships of their cues in words in the training set. It was predicted that article assignment by the network would generally follow the most reliable cues. Two sets of eight words were formed. In the first set, the words contained one or more cues whose reliabilities favored the correct gender, or contained two cues whose reliabilities conflicted with respect to the gender, but in which the cue with stronger reliability favored the correct gender. It was predicted that words in this first group would receive correct article assignment. In the second group, the word had one or more cues whose reliabilities agreed on the gender of the noun, but this gender was actually incorrect; or, the word had cues whose reliabilities conflicted with respect to the gender, and in which the cue with the stronger reliability favored the incorrect gender. It was predicted that words in this group would receive incorrect article assignment. Article assignment by the network, shown in Table 6, matched the prediction of reliability exactly. All words in group 1 were given correct articles; all words in group 2 received incorrect articles.

Overall, this simulation was a success. An examination of the internal representations showed that the network was representing the categorical constructs that characterized the training set — i.e. masculine, feminine, and neuter genders. The correlation results suggest that basic constructs from the Competition Model can be used to describe the basis on which these internal representations of categories develop over time. The test on novel words confirms performance predictions that one would make after applying these same constructs to the linguistic input available to the network. This simple test confirmed our belief that the cues would be generalizeable to different words, and were not bound to the context in which they were learned.

Table 6. *Test words for Simulation II*

Group 1	
das Raubzeug	varmint
die Stute	mare
die Krähe	crow
die Raupe	caterpillar
die Injektion	injection
der Stromstoss	electrical shock
die Lehre	lecture
die Praktik	practice
Group 2 (errorful responses in parentheses)	
der (das) Liter	liter
der (das) Meter	meter
das (der) Bier	beer
die (der) Tunte	male effeminate homosexual
das (die) Weib	woman (derogatory)
das (die) Mensch	woman (derogatory)
das (die) Reff	crone
der (das) Gegenstand	object

7. Discussion

We can discuss the results of these simulations in terms of four major issues: the use of rules in learning models, the relation of the simulation to the developmental data, the role of lexical items in the model, and directions for future research.

7.1 *Rules vs. networks*

We began this research with an interest in examining alternatives to a rule-based account of language learning. Although rules are useful ways of characterizing the competence of the adult speaker of a language, it is often difficult to find developmental evidence for the acquisition of rules by children. The model we have constructed provides an alternative way of thinking about language learning. The results of the two simulations show that, given the information available in the German language, it was possible for our network to accurately assign definite articles to training words in various case contexts, to generalize the case paradigm to these same words placed in new case contexts, and to generalize gender and number cues to new words. There is evidence here that the network achieves this by representing categorical information internally, and activating these 'categories' in the course of performance. This performance may look rule-like, but it is clearly not generated by explicit rules. Rather, the ability of the network to learn the system for article declension in German is a result of its ability to take a set of reliable cues and combine them in complex ways to predict the correct output. Without good cues, this model could not have exhibited the productive behavior that it did.

The cues we used are all cues that are eventually available to the child in the input. We say "eventually" because we realize that these cues themselves are not givens, but are classes that are constructed out of smaller building blocks. For example, the child is not born with detectors for /šC/ clusters. However, this cue becomes available to the child during the process of the acquisition of phonology. Similarly, the child is not born with detectors for the presence of the diminutive suffix *-chen*. Rather, he must learn the association between the phonological features composing /çən/ and the meaning features of the diminutive.

7.2 *The model and the developmental data*

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.

However, the model went far 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.

Beyond mastering the problem of German definite article assignment, the simulations presented here are in accord with more general patterns in the empirical literature. As illustrated by the second simulation, the strength of the connection between a cue and a category changes during the learning process — cue strength first follows cue validity, then reliability, and finally a construct closely allied with conflict validity. Similar shifts have been found in other cue-category learning situations (McDonald 1986, McDonald & MacWhinney 1987, Sokolov, 1988).

7.3 *The role of lexical items*

Recently, Pinker and Prince (1987) have presented a criticism of the Rumelhart and McClelland (1986) simulation of the acquisition of the past tense in English. It is beyond the scope of the current article to treat their criticisms in their entirety. However, it is interesting to consider one of the most important of the Pinker-Prince criticisms in the light of the current study. This criticism has to do with the role of the lexical item in connectionist architectures. Neither the current simulation nor the earlier model of Rumelhart and McClelland makes reference to the lexical item as a distinct entity. In the Rumelhart and McClelland simulation, there was no way to distinguish between forms such as "ringed" and "wrung," because the present tense forms "ring" and "wring" had similar phonological representations, and there were no semantic representations. The current simulation avoids this problem because it introduces a set of semantic features which provide a rough coding of the identity of the lexical item. Thus, this simulation 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 semantic features. Such learning would take a number of iterations, but it should eventually be possible.

We do not believe that the issue of the role of lexical items in the acquisition of morphology is yet settled. The problems the simulation had in acquiring forms like *Junge* and *Ende* may point to a role for rote. It may still be important to think of morphological processing in terms of an interplay between the separate mechanisms of rote and analogy, as suggested by MacWhinney (1978). However, 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 remains a topic for future research.

7.4 *Directions for future research*

There are still other directions for future research. Within the space of possible connectionist architectures, we have sampled only a small subset. Our simulations have failed to test for the effects of varying important parameters such as the learning rate. Our coding of the semantic features of nouns is clearly incomplete. A richer set of codes may lead to even more naturalistic learning patterns. The selection of a training corpus in future work should be based upon samples of the speech directed toward young children, rather than just estimates deriving from frequency lists. At the same time we need to conduct detailed studies of the various cue validity measures in similar texts. Important cues, such as the use of pronouns as identifiers of gender (MacWhinney 1978), have not yet been used. Eventually, we will want to study the learning of the entire system of declension for adjectives, pronouns, and other determiners. This long list of tasks awaiting our attention shows how exciting it will be to work out all the implications of connectionist approaches to language learning.

Acknowledgement

We would like to thank Klaus-Michael Köpcke, James McClelland, Roberta Corrigan, Michael Noonan, and Fred Eckman for comments on an earlier version of this paper, and James McClelland for providing us with the Back Propagation software.

Notes

*Janet L. McDonald is currently at Louisiana State University, Department of Psychology, Baton Rouge, LA 70803.

1. 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 an activation greater than .5, it was considered to be on and a Hit; if it was on but was not supposed to be, a False Alarm. An output node with an activation less than .5 was considered off. If an output unit was supposed to be off and was off, it was a Correct Rejection; if it was off but was supposed to be on, a Miss.
2. The extended training period in this simulation (e^7 = 1097 passes through the training set) was the result of a lower learning rate, compared to Simulation I. As was noted, this is a parameter in the Back Propagation algorithm.

References

Bates, E. & MacWhinney, B. 1987. "Competition, Variation and Language Learning: What is not Universal in Language Acquisition". *Mechanisms of Language Acquisition* ed. by B. MacWhinney. Hillsdale, NJ: Lawrence Erlbaum.

Brunswik, E. 1956. *Perception and the Representative Design of Psychology Experiments*. Berkeley, CA: University of California Press.

Bybee, J.L. & Slobin, D.I. 1982. "Rules and Schemas in the Development and Use of the English Past Tense. *Language* 58. 265-289.

Chomsky, N. & Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Ervin-Tripp, S. 1966. "Language Development". *Review of Child Development Research*, Vol. 2 ed. by M. Hoffman & L. Hoffman. Ann Arbor: University of Michigan Press.

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. *The Cross-Linguistic Study of Language Processing* ed. by B. MacWhinney & E. Bates. Cambridge: New York Cambridge University Press.

Köpcke, K. 1987. "Schemas in German Plural Formation". *Lingua* 74. 303-335.

Köpcke, K., and Zubin, D. 1983. "Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache". *Zeitschrift fur Germanistische Linguistik* 11. 166-182.

———. 1984. "Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation". *Linguistische Berichte* 93. 26-50.

Kuczaj, S. A. 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: Charles Scribner & Sons.

MacWhinney, B. 1978. "The Acquisition of Morphophonology". *Monographs of the Society for Research in Child Development* 39.

———. 1987. "Competition". *Mechanisms of Language Acquisition* ed. by B. MacWhinney. Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B., Bates, E., & Kliegl, R. 1984. "Cue Validity and Sentence Interpretation in English, German, and Italian". *Journal of Verbal Learning and Verbal Behavior* 23. 127-150.

MacWhinney, B., & Snow, C. 1985. "The Child Language Data Exchange System". *Journal of Child Language* 12. 271-296.

Maratsos, M. & Chalkley, M. 1980. "The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories". *Children's Language* ed. by K. Nelson. New York: Gardner Press.

McDonald, J.L. 1986. "The Development of Sentence Comprehension Strategies in English and Dutch". *Journal of Experimental Child Psychology* 41. 317-335.

———. 1989. "Determinants of the Acquisition of Cue-Category Mappings". *The Cross-linguistic Study of Language Processing* ed. by B. MacWhinney & E. Bates. New York: Cambridge University Press.

McDonald, J.L. & MacWhinney, B. 1987. "Levels of Learning: A Comparison of Concept Formation and Language Acquisition." Manuscript, Carnegie-Mellon University.

Meier, H. 1967. *Deutsche Sprachstatistik*. Hildesheim, W. Germany: Georg Olms.

Mills, A. E, 1986. *The Acquisition of Gender: A Study of English and German*. Berlin: Springer-Verlag.

Mugdan, J. 1977. *Flexionsmorphologie und Psycholinguistik*. Tübingen: Gunter Narr.

Pinker, S. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. & Prince, A. 1987. "On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition". Occasional Paper 33, M.I.T.

Rumelhart, D., Hinton, G. & Williams, R. 1986. "Learning Internal Representations by Error Propagation". *Parallel Distributed Processing*, Vol. I ed. by D. Rumelhart and J. McClelland. Cambridge, MA. MIT Press.

Rumelhart, D. & McClelland, J. 1986. "On Learning the Past Tenses of English Verbs". *Parallel Distributed Processing*, Vol. I ed. by D. Rumelhart and J. McClelland. Cambridge, MA: MIT Press.

Schneuwly, B. 1978. "Zum Erwerb des Genus im Deutschen: Eine mögliche Strategie." Manuscript, Max-Planck-Institut, Nijmegen, the Netherlands.

Slobin, D.I. 1971. *Psycholinguistics*. Glenview, IL: Scott, Foresman & Company.

Sokolov, J.L. 1988. "Cue Validity in Hebrew Sentence Comprehension." *Journal of Child Language* 15. 129-156.

Wängler, H.H. 1963. *Rangwörterbuch hochdeutscher Umgangssprache*. Marburg: N.G. Elwert Verlag.

Zubin, D. & Köpcke, K. 1981. "Gender: A Less than Arbitrary Grammatical Category". *Papers from the Seventh Regional Meeting* ed. by R. Hendrick, C. Masek & M.F. Miller. Chicago: Chicago Linguistic Society.

———. 1986. "Gender and Folk Taxonomy: The Indexical Relation between Grammatical and Lexical Categorization". *Noun Classes and Categorization* ed. by C. Craig. Amsterdam: Benjamins.