Models of Language Acquisition

Inductive and Deductive Approaches

Edited by

PETER BROEDER
AND JAAP MURRE

2000

OXFORD
UNIVERSITY PRESS

2

Lexicalist Connectionism

BRIAN MACWHINNEY

2.1. Introduction

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

During the 1960s and 1970s, scientists took this common sense idea of a linguistic rule and reworked it into a basic principle underlying artificial intelligence (AI), Chomskyan theoretical linguistics, and cognitive psychology. By viewing the brain as a computer, they began to think of the mind as a system for transforming symbolic strings according to well-specified rules. The vision of human language as a system of formal rules was an important ingredient underlying two decades of work in linguistics and cognitive science. This work led to the emergence of complex and impressive systems of rules and symbols based on what I have called the 'Big Mean Rules' (MacWhinney 1994) and the 'Big Mean Flowcharts' which were systems designed to learn the Big Mean Rules.

In recent years, the empirical underpinnings of these great symbolic systems have become increasingly shaky and vulnerable. Two basic observational problems faced by all of these analyses are the fact that no developmental psychologist ever observed a child learning a rule and that no neuroscientist ever traced the neural substrate of either a rule or a symbol. Beginning in the 1970s and continuing up to the present, attempts to provide the necessary demonstrations of the psychological reality of rules in adults (Fodor, Bever, and Garrett 1974; Ohala, J.J. 1974*a*,*b*; Ohala, M. 1974; Trammell 1978; Linell 1979; Jaeger 1984) have yielded uniformly disappointing results. More recently, systems of rules have been supplemented with innate parameters, triggers, and constraint satisfaction hierarchies designed to minimize the size of the core rule component. However, these new conceptual devices have ushered in a new set of empirical worries, as attempts to match the parameters, triggers, and constraints to the facts of language learning have failed to yield consistent results (Meisel 1995).

Given these doubts and empirical failures, it made sense for researchers to begin to explore alternatives to symbols and rules. In the late 1980s, work in connectionist modelling (Rumelhart and McClelland 1986) began to challenge the necessity of linguistic rules and categories, focusing attention instead on models based on simple, observable cues and connections between these cues. These new models sought to correct a fundamental, fatal flaw inherent in symbolic models: the problem of excessive descriptive power.

The great power of AI systems derives from the computational architecture of the von Neumann serial computer and the application of this architecture to human cognition by Simon, Newell, and their followers (Newell and Simon 1972; Klahr and Kotovsky 1991). In the digital computer, once a symbol is stored in a location in computer memory (RAM), it then becomes fully accessible to any rule that wants to access it. It is this availability of symbols in a uniformly addressable memory that provides the von Neumann computer architecture with symbol passing, generativity, and scalability. When this hardware architecture is supplemented by a high-level AI programming language, such as ACT-R (Anderson 1993), the expressive capacity of the system is virtually limitless. A modeller can take a few symbols, concatenate them into rules and, magically, the computer will conjure up a working model of mental processing.

However, these AI models are at the same time both too powerful and too weak. They are too powerful in that they allow one to model the learning of things that could never in reality be learned. At the same time, they are too weak in that they fail to generalize properly across language types and patterns. Moreover, attempts to identify a uniquely correct model without adding further constraints have been shown to be impossible in principle (Anderson 1978). Neural nets (Hopfield 1982; Kohonen 1982; Grossberg 1987) limit this descriptive power by imposing two stringent limitations on computational models: a prohibition against *symbol passing* and an insistence on *self-organization.*

Neural networks require that the computations involved in the models echo the connectionist architecture of the brain. The basic constraint involved here is the prohibition against symbol passing. The clearest example of symbol passing is a simple pair of rewrite rules such as the ones that expand a sentence in a noun phrase and verb phrase:

1. S $\rightarrow$ NP + VP
2. VP $\rightarrow$ V + NP

Here, the symbol 'VP' generated by rule 1 is passed on to rule 2 for further expansion. In fact, rewrite rules are nothing much more than symbol passing devices, whose chief function is to pass symbols about until finally sounds and words are activated (MacWhinney and Anderson 1986).

However, neuroscience has shown that the brain cannot pass symbols. This is because it cannot use memory addresses to bind variables, since there is no neural mechanism that can assign an absolute 'address' to a particular neuron (Squire 1987). Neurons do not send Morse code, symbols do not run down synapses, and brain waves do not pass phrase structures. Unlike the computer, the brain has no general scheme for register assignment, data pathing, or memory addressing. Moreover, the individual components of the neural system do not have the reliability of the electrical components of a standard digital computer (von Neumann 1956). In general, the brain provides no obvious support for the symbol passing architecture that provides the power underlying the von Neumann machine. Instead, computation in the brain appears to rely ultimately on the formation of redundant connections between individual neurons.

By itself, the requirement that computation be performed locally without symbol passing or homunculi is not enough to fully constrain descriptive power. One could still hand-wire a connectionist system to perform a specific function or behaviour. To do this, one actually has to 'go inside the system' and set a value for individual nodes by hand. By detailed weight setting and the use of gating and polling neurons, virtually any function can be wired into a neural net (Hertz, Krogh, and Palmer 1991). An early example of a fully hand-wired connectionist architecture was Lamb's (1966) stratificational grammar. Although Lamb explicitly warned the reader that the labels on individual nodes were not 'really there', he never provided a mechanism for getting a network to take on the shape required by his insightful linguistic theory.

More recently, we have seen hand-wired connectionist models in areas such as speech errors (Stemberger 1985; Dell 1986; MacWhinney and Anderson 1986), ambiguity resolution (Cottrell 1987), and lexical activation Marslen-Wilson 1987; Norris 1994). The 'implementational' approach to hand-wiring spares the modeller the tedium of hand-wiring by running the wiring procedure off symbolic templates. For example, Touretzky (1990) has shown that there are techniques for bottling the full power of a LISP-based production system architecture into a neural net. These demonstrations are important because they show how difficult it is to control excessive modelling power. However, they tell us little about how language is implemented in the brain.

In order fully to constrain descriptive power, modellers must match the constraint against symbol passing with the requirement that networks be *self-organizing.* The notion of self-organization can be best understood in terms of simple physical systems. Perhaps you have noted that each of the cells of a bee's honeycomb assumes a clear hexagonal shape. Is this shape carefully hand-crafted cell by cell by each and every honeybee? No, it simply emerges through self-organization as the best solution to the problem of packing a certain quantity of honey cells into a certain volume.

Neural networks acquire their shape through a similar, albeit more complex, self-organizing process. For example, the algorithm underlying the Kohonen feature map network works in a way that encourages units that behave similarly to move toward spatially similar places on the feature map. It is this property of neural nets that makes them particularly interesting to the developmental psychologist and which also poses the greatest challenge to detailed modelling work. When the prohibition against symbol passing is combined with the demand for self-organization, the class of potential models of language learning becomes extremely limited. In fact, there is currently no detailed model of language acquisition that can satisfy these two criteria. Is this evidence that the criteria are too strict? I think not. Rather it is evidence that we can use these criteria to constrain our search for a truly plausible model of language acquisition. More importantly, it appears that those models which come closest to satisfying these criteria are also the same models that display further interesting and important properties, such as category leakage (McClelland and Kawamato 1986), graceful degradation (Hinton and Shallice 1991; Harley and MacAndrew 1992; Marchman 1992), and property emergence (MacWhinney *et al.* 1989).

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

2.2. Grand Pretensions, Modest Reality

Like the symbolic paradigm before it, the connectionist paradigm seeks to provide a general model of human cognition. Because it has staked out such a wide territory, connectionism is committed to providing an account of all of the core issues in language acquisition, including grammatical development, lexical learning, phonological development, second language learning, and the processing of language by the brain. Despite these grand pretensions, the reality of connectionist modelling is more sober and modest. In fact, much of the work to date has focused on the learning of narrow aspects of inflectional morphology in languages like English and German. While limited, work in this area has taught us a great deal. This chapter sketches out the achievements of connectionist models in this well-researched area and then examines how we can move from these preliminary achievements to a fuller, more explanatory, unified approach to the core issues facing language acquisition theory.

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

Like all connectionist models, this model based its performance on the development of the weights on the connections between a large collection of 'units'. The pattern of inputs and the connections between units was designed to implement the pattern of an autosegmental grid that has been developed in phonological theory (Goldsmith 1976; Nespor and Vogel 1986). The idea is that each vowel or consonant sound is a bundle of features that sits inside a slot within the framework or grid of the syllable. Words, in turn, are formed from combinations of syllables in a metrical grid. The MacWhinney–Leinbach model used 12 consonantal slots and six vowel slots and allowed for words of up to three syllables. The segments of the stem were filled into this grid in right-justified fashion (MacWhinney 1993), as in this example for the word 'bet':

Right justified: CCC VV CCC VV CC$\underline{b}$ V$\underline{E}$ CC$\underline{t}$

A further syllable was reserved for the suffix in the output. Each of the slots was in turn composed of a group of feature units. Since each of these feature units was bound to its particular slot, we can think of each unit as a slot/feature unit. For example, the first consonantal slot in the representation for 'bet' would have active units for the labial, consonantal, and voiced features required for the sound /b/. Each of the consonantal slots had ten units and each of the vowel slots had eight units. The network is displayed in Figure 2.1.

The complete training corpus used 6,949 different verb forms, derived from the 2,161 highest frequency verbs in English (Francis and Kucera 1982). Of these 2,161 verbs, 118 were irregulars and 2,043 were regulars. The frequency with which a given form was included in the training epochs was determined by its

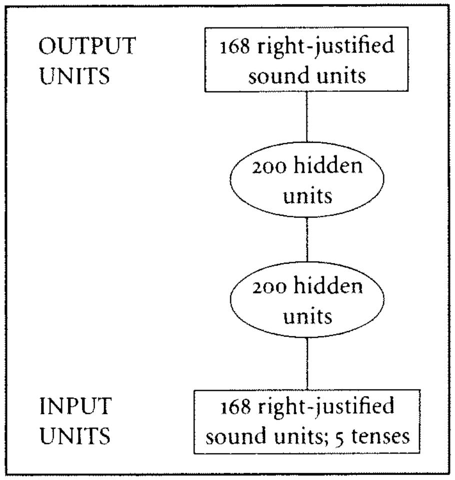

FIGURE 2.1. Neural network model for English conjugation

frequency in the Francis and Kucera (1982) word-frequency list. The highest frequency verbs were included most often. Learning in the model was controlled by the backpropagation algorithm (Rumelhart, Hinton, and Williams 1986).

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

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

> a person who has not studied German can form no idea of what a perplexing language it is . . . Every noun has a gender, and there is no sense or system in the distribution; so the gender of each must be learned separately and by heart. There is no other way. To do this, one has to have memory like a memorandum book. In German, a young lady has no sex, while a turnip has. Think what overwrought reverence that shows for the turnip, and what callous disrespect for the girl.

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

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

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

This paradigm is rife with homonymy. The six forms of the definite article (*der, die, das, dem, des, den*) must cover the sixteen cells in the paradigm. This is done by having a single form cover several meanings. For example, the article *der* can mean either masculine singular nominative, feminine singular genitive, feminine singular dative, or plural genitive.

In order to select the correct form of the definite article, the language learner has to know three things about the noun—its case, its number, and its gender. Number is the easiest category, since it bears a straightforward relation to real-world properties. Case is somewhat more abstract, but it can generally be figured out through a combination of cues from the verb, related prepositions, and some word-order patterns. However, there is little in the external situation that can help the child figure out the gender of a noun (Maratsos and Chalkley 1980). It is possible that the noun's gender could be simply memorized or even inferred on the basis of its use within the paradigm. However, recent work by Köpcke and Zubin (Köpcke 1982, 1988; Köpcke and Zubin 1983, 1984, 1997; Zubin and Köpcke 1981, 1986) has shown that Mark Twain's view of gender as arbitrary and unpredictable is incomplete and partially incorrect.

In fact, Köpcke and Zubin have shown that there are dozens of phonological cues that can be used to predict the gender of a German noun. For example, almost all nouns ending in *-e* are feminine, as in *die Sonne, die Ente,* and *die Tante.* Almost all nouns beginning with *dr-, tr-,* and *kn-* are masculine, as in *der Knecht, der Trieb,* and *der Drang.* There are dozens of other cues like these. In addition to these purely phonological cues, there are derivational endings such as *-chen, -lein, -ett, -tum, -ei,* and so on, each of which reliably specifies a particular gender.

MacWhinney *et al.* (1989) constructed a series of models of the acquisition of this complex system of German case–number–gender marking. The first model dedicated a series of nodes to the cues enumerated by Köpcke and Zubin, along with a series of nodes for case and number cues. The second model made no explicit coding of the Köpcke–Zubin cues, instead simply encoding the phonological form of the base in the manner of the MacWhinney–Leinbach model for English. Much to our surprise, the network with no hand-coding of features outperformed the hand-crafted network in terms of both learning and generalization. These results provide nice support for the view of connectionist networks as providing emergent self-organizing characterizations of linguistic systems. Similar results for hand-wired vs. emergent solutions are reported by Daelemans, Gillis, and Durieux (1993) for the learning of Dutch stress by a connectionist network. The successful German simulation without the hand-crafted input units is displayed in Figure 2.2.

The input to the network was a pattern across the 143 phonological units to represent the noun stem and the eleven phonological units to represent suffixes attached to the noun. In addition, there were five semantic units representing inherent gender and seventeen cues that provided a distributed pattern of surface structure information helpful in determining the case for the noun. However, the actual identity of the case to be used was not given. Hidden units were attached separately to the two major types of inputs and then connected to a further set of 'collector' units. This combination of 'feeder' and 'collector' hidden units is frequently used for input sets that differ markedly in type.

The network was trained with 2,000 German nouns from all cells in the paradigm. It learned the training set completely. When tested with 200 new nouns, the system was able to guess the gender of the new words with 70 per cent accuracy. This compares with a level of 80 per cent accuracy (Köpcke, personal communication) that could be expected from a native German speaker.

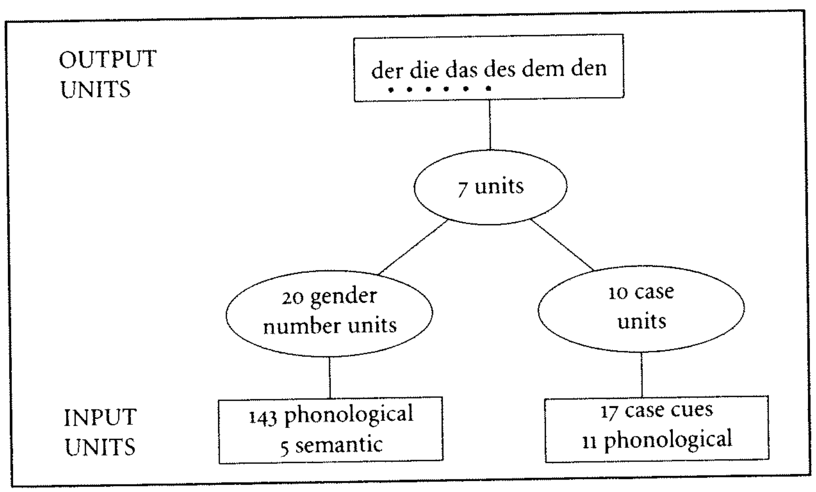

FIGURE 2.2. Neural network model of German declension

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

When evaluating the success of these connectionist models of language acquisition, it is important to consider the extent to which symbolic models are able to address similar problems. For the learning of English verb morphology, Ling (1994) and Ling and Marinov (1993) present a model that performs about as well as the MacWhinney–Leinbach model for English. Although Ling's model is based on a conventional symbol-passing architecture, it uses an input-driven induction algorithm, thereby avoiding problems with hand-wiring. However, the detailed feature combinations constructed by Ling's pattern associator provide no clear representation of rules and would probably not be accepted as a full symbolic model by many linguists and psycholinguists. Nonetheless, the comparisons conducted by Ling show that the competitive testing of symbolic and connectionist models can be quite instructive.

2.3. Lexical Items: an Achilles Heel?

Despite its basic successes, there are several properties of the MacWhinney *et al.* models that should give us serious cause for worry. Fortunately, as so often happens, weaknesses and failures may actually be more instructive than successes, as long as we are willing to learn from our failures. In regard to inflectional learning, the weaknesses we can detect appear in a similar garb for both English and German.

2.3.1. *Problem #1—Homophony*

Because these models perform a conversion from one phonological form to another phonological form without using discrete representations for lexical items, they run into serious problems with homophonous forms. Consider the three ways in which we can form the past tense of 'ring' in English. We can say 'the maid wrung out the clothes', 'the soldiers ringed the city', or 'the choirboy rang the bell'. These three different past tense forms all have the same sound /rIN/ in the present, but each takes a different form in the past.

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

2.3.2. *Problem #2—Compounds*

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

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

Of course, this problem would go away if the model knew how to treat a compound as two individual words. But most of the models we have discussed have no concept of a 'word' and therefore no way of understanding what it might mean for 'undergo' to be potentially two separate words.

2.3.3. *Problem #3—Derivational Status*

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

German provides even clearer examples of the importance of derivational status. All German nouns that derive from verbs are masculine. For example, the noun *der Schlag* 'blow, cream' derives from the verb *schlagen* 'to hit'. However, there is no way of indicating this in the model, since it has no concept of words and no idea about how one word could derive from another. Of course, we could simply hand-wire a feature called [+derived], but this would be in clear violation of the principles we discussed earlier.

Second, because affixes are lexical items and because the model has no concept of lexical items, it cannot distinguish between true phonological cues such as final *-e* or initial *kn-* and derivational markers such as *-chen* or *-ett*. This leads to some very clear confusions. For example, masculines such as *der Nacken* 'neck' and *der Hafen* 'harbour' end in phonological /en/, whereas neuters such as *das Wissen* 'knowledge' and *das Lernen* 'learning' end in the derivational suffix *-en*. Confusion of these two suffixes leads to inability to predict gender correctly for new nouns ending in *-en*.

2.3.4. *Problem #4—Early Irregulars*

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

2.3.5. *The Core Problem*

These four weaknesses we have discussed can be linked to a single core problem: the absence of any way of representing lexical items. Lexical items are meant to include anything that might somehow be entered into a systematic dictionary as a form–function relation. This would include all manner of content words and functions words, as well as productive inflectional affixes. Because these neural network models have no lexical items, they are forced to rely on sound features as the only way to determine inflectional morphology. It would be a mistake to imagine that the sound form of words has no impact on inflection and derivation. In fact, it seems that what really happens during both production and comprehension is that both the sound and meaning of stems and affixes are available in parallel, although the time course of their activation may vary (Kawamato 1993).

One way of addressing this problem is to mix both sound features and meaning features without providing any explicit representation of lexical items. Attempts to achieve lexical access without lexical representations have been partially effective in models of reading (Kawamato 1993; Plaut *et al.* 1996) and spelling (Seidenberg and McClelland 1989). It makes sense to use non-lexical representations for this task, because orthographic–phonological correspondences typically make little reference to lexical items. However, these models run into more serious problems (Cottrell and Plunkett 1991; Hoeffner 1992), when dealing with language learning and word production. Models of the Hoeffner type display this problem most clearly. They learn to associate sound to meaning and store these associations in a distributed pattern in the hidden units. This approach works well enough until the model is given more than about 700 forms. At this point, the large pool of hidden units is so fully invested in distinguishing phonological and semantic subtypes and their associations that there is simply no room for new words. Adding more hidden units does not solve this problem, since all the interconnections must be computed and eventually the learning algorithm will bog down. It would appear that what we are seeing here is the soft underbelly of connectionism—its inability to represent Islands of Stability in the middle of a Sea of Chaos. Perhaps the problem of learning to represent lexical items is the Achilles heel of connectionism.

2.4. A Solution to the Lexical Learning Problem

Given the seriousness of these problems and the extent to which they have limited the full effectiveness of connectionist models for English and German, we decided to explore alternatives to fully distributed representations. The core assumption in our new approach is that the lexical item serves a central controlling and stabilizing role in language learning and processing. We can refer to this

revised approach as *lexicalist connectionism*. Predecessors to lexicalist connectionist models can be found in localist connectionist models of the type developed by Stemberger (1985) and Dell (1986), where a central role is given to the lexical item. However, because of their localist node-based architecture, these models were forced to rely on hand-wiring.

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

Schematically, one can think of the map as a 2-D compression that maps a set of multidimensional inputs onto a map in the way suggested by Figure 3. In this map, the inputs being mapped are auditory forms. Learning involves the strengthening of weights between particular inputs and units on the map. This can be done in strict accord with established biological principles of lateral inhibition and the redistribution of syntactic resources (Kohonen 1982) using a computationally efficient algorithm that is faithful to these biological principles (Miikkulainen 1990).

Word learning involves development of one map for auditory forms and another for semantic forms. These two maps are then mutually associated using Hebbian learning. With this algorithm, we found that a network with 10,000 nodes can learn up to 6,000 lexical associations with an error rate of less than

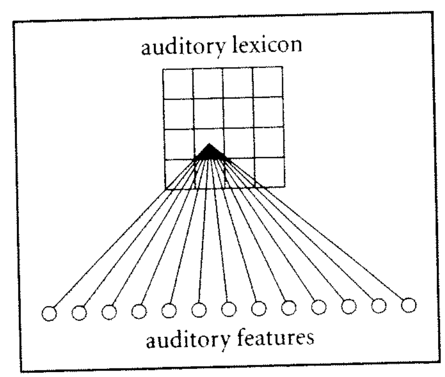

FIGURE 2.3. Self-organizing feature map for auditory patterns

1 per cent. In this implementation, we used four floating-point numbers to represent sound and four additional floating-point numbers to represent meaning. The shape of these eight numbers for each item was generated randomly. At the beginning of learning, the first input vector of eight numbers led by chance to somewhat stronger activation on one of the 10,000 cells. This one slightly more active cell then inhibits the activation of its competitors. This is done using a 'Mexican hat' function which allows nearby neighbours to maintain some activation, while strongly inhibiting cells a bit farther away. As a result of this pattern of activation and inhibition, inputs that are close in feature space end up activating cells in similar regions of the map. Once a cell has won this particular competition, its activation is negatively damped to prevent it from winning for all of the inputs. Then, on the next trial, another cell has a chance to win in the competition for the next sound–meaning input pattern. This process repeats until all 6,000 sound–meaning patterns have developed some 'specialist' cell in the feature map. During this process, the dynamics of self-organization ensure that items that shared features end up in similar regions of the feature map.

We were able to follow the development of the feature map by tracking the average radius of the individual items. After learning the first 700 words, the average radius of each word was 70 cells; after 3,000 words, the radius was 8; after 5,000 words the radius was 3; and after 6,000 words the radius was only 1.5 cells. Clearly, there is not much room for new lexical items in a feature map with 10,000 cells that has already learned 6,000 items. However, there is good reason to think that the enormous number of cells in the human brain ensures that the size of the initial feature map is not an important limiting constraint on the learning of the lexicon by real children.

2.5. Using Lexical Representations

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

2.5.1. *Learning Inflectional Morphology*

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

Before this training is completed, we introduce a few generalization trials. These trials take verbs that have been learned as bare-stems, but not yet as inflected forms. For example, the network already knows 'jump', but has not yet learned 'jumped'. At this point, the network engages in the process of 'masking' (Cohen and Grossberg 1987; Grossberg 1987; Carpenter, Grossberg, and Reynolds 1991; Burgess and Hitch 1992; Burgess 1995) which leads to the suppression of 'jump' and the isolation of 'ed' as the not-yet-recognized form. The suffix is then learned as a new lexical item that associates *-ed* with the meaning of the past tense. In the terms of MacWhinney (1987*a*), learning involves associating the 'unexpressed' with the 'uncomprehended'. Learning of the other inflectional suffixes proceeds in a similar way. When the system is asked to produce the past tense of a new verb, masking is used in reverse. For example, the past tense of 'jump' is produced by first activating 'jump' and then masking the semantics and phonology of 'jump', reactivating the network with 'past' and then retrieving *-ed*.

This approach to the problem of learning the English past tense solves a number of problems faced by earlier nonlexical models. First, the model succeeds in capturing both rote lexicalization and combinatorial lexicalization within a single connectionist model. Rote forms are picked up directly on the feature map. Combinatorial forms are created by the isolation of the suffix through masking and the use of masking in production. Second, the model no longer faces the earlier problems that stemmed from a lack of lexical items. Homophony is not a problem, because the various homophonic meanings of 'ring' are now representationally distinct. The model can locate 'go' inside 'undergo' and *Mutter* inside *Grossmutter* because it now has lexical items and can use masking. In German, derivational suffixes like *-chen* can be used as cues to gender because these suffixes now have their own representational status. Finally, the model no longer has any problem with the early acquisition of irregulars such as 'went'. Since the learning is grounded now on lexical items, these high frequency forms are some of the first forms learned.

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

2.5.2. *The Logical Problem*

The ability to produce 'went' by rote and '*goed' by combination within the same lexical feature map also allows us to begin work on a general solution to the so-called 'logical problem of language acquisition' (Wexler and Culicover 1980; Baker and McCarthy 1981; Gleitman, Newport, and Gleitman 1984; Pinker 1984, 1989; Morgan and Travis 1989; Gleitman 1990). In the case of the competition between 'went' and '*goed', we expect 'went' to become solidified over time because of its repeated occurrence in the input. The form '*goed', on the other hand, is supported only by the presence of the combinational *-ed* form. Figure 2.4 illustrates this competition. This particular situation is an example of what Baker calls a 'benign exception to the logical problem'. Later, we will examine some non-benign cases.

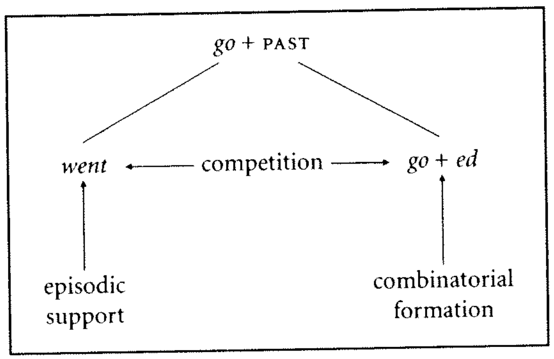

FIGURE 2.4. The competition between regular and irregular inflections

2.5.3. *Masking and Buffering*

The masking mechanism underlies not only inflectional extraction, but also syntactic processing more generally. In order to process sentences, we need to have some process that deactivates each lexical item immediately after it is activated. The trace of this masked item must then be stored provisionally in some separate form apart from the main lexicon. The simplest way to do this is to activate a second copy of the original item (Burgess and Hitch 1992). This could be done by creating a complete secondary copy of the primary lexicon. However, even this complete duplication of the lexicon would only guarantee memory for two words at a time. A more flexible system would convert the initial lexical representations to some other pattern. There have been several suggestions regarding the nature of this short-term verbal memory.

1. As soon as words are linked together into conceptual clusters, they can be used to activate a unique underlying meaning that no longer requires verbal storage.
2. Before this linkage occurs, words may be obtained in a phonological loop (Baddeley 1986). This immediate rehearsal requires that words be present in a primarily articulatory form (Gupta and MacWhinney 1994).
3. It is also possible that some additional mechanism operates on lexical items to encode their serial occurrence without reference to either meaning or sound. This could be done in terms of some additional episodic, possibly hippocampal, mechanism that stores activation levels of words prior to masking. A system of this type is close to the Competitive Queuing mechanism proposed first by Grossberg and then again by Houghton.

Further experimental work will be needed to understand more closely which of these three mechanisms is involved at which point in the storage of short-term verbal memories. What is important for our current simulations is only the fact that there is evidence that neural (Gupta and MacWhinney 1997) mechanisms are available to support masking in the lexicon.

2.5.4. *Argument Frame Extraction*

Earlier versions of the Competition Model (MacWhinney 1988) presented a system for the control of syntax through lexical argument (or 'valency' or 'dependency') relations. From a connectionist viewpoint, the masking process is what triggers the acquisition of argument relations. To illustrate this process, consider an example in which the child already knows the words 'Mommy' and 'Daddy', but does not know the word 'like'. Given this state of lexical knowledge, the sentence 'Daddy likes Mommy' would be represented in this way:

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

For the first and last phonological stretches, there are lexical items that match. These strings and the semantics they represent are masked. The unknown stretch is not masked and therefore stimulates lexical learning of the new word 'likes'.

The core of the learning for 'likes' is the association of the sound /laIk/ to the meaning 'like'. In addition to this association, the central association feature map now constructs links not only to sound and meaning, but also to argument relations. The initial argument frame for the word 'likes' is [[arg1, preposed, 'Daddy'][arg2, postposed, 'Mommy']]. Further exposures to sentences such as 'Daddy likes pancakes' or 'Billy likes turtles' will soon generalize the dependency frame for 'likes' to [[arg1, preposed, human][arg2, postposed, object]].

The implementation of the acquisition of argument frames follows the logic developed by Gupta and MacWhinney (1992) for the acquisition of German declensional marking. That model used a SOFM for the extraction of cooccurrence patterns between articles and nouns. Using these patterns, the full shape of the German declensional pattern emerged inside the SOFM. Nodes in the map took on the role of associating particular constellations of case and number marking on the article with one of the three grammatical genders of German. This system was then linked to a backpropagation system that responded to additional phonological cues to gender. The general shape of this type of model is given in Figure 2.5 which includes the basic components of Figure 2.2 along with additional maps for argument structures and articulatory control.

The argument frame feature map is intended to capture a few basic grammatical categories which will then be related to particular lexical items. However, these argument frames will also activate patterns in the meaning and auditory form feature maps.

The feature map is designed to include two basic effects. One is the activation of the correct argument frame for a specific lexical item. The other is the activation of argument frames for semantically related groups of words or lexical 'gangs'. Words that have similar meanings will tend to activate similar argument structures.

Lexical gang effects help us address some remaining aspects of the 'logical problem of language acquisition'. Bowerman (1988) cites overgeneralization errors such as '*I poured the tub with water', '*I unsqueezed the ball', and '*I recommended him a soup' as potential evidence for the logical problem. However, recovery from these errors can be viewed as similar to recovery from errors such as '*goed'. Figure 2.6 illustrates the situation.

In this case, the construction 'recommend Y X' represents an ungrammatical phrase such as '*recommend the boy the book' and the construction 'recommend X to Y' represents a grammatical phrase such as 'recommend the book to the boy'. An overgeneralization such as '*recommend the boy the book' arises from generalization out of phrases such as 'give the boy the book' or 'send your client a check'. Because the verb 'recommend' shares many semantic features with transfer verbs such as 'give' and 'offer', it becomes a part of a lexical gang and is subject to overgeneralization or what connectionists call lexical gang

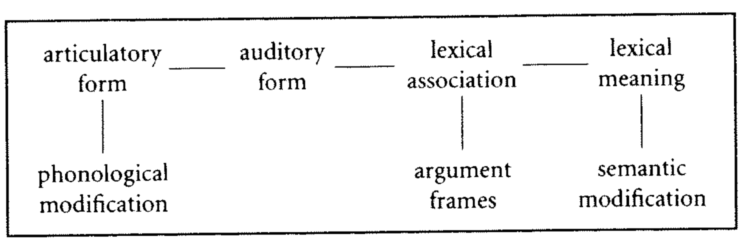

FIGURE 2.5. A connectionist account of the shape of linguistic representations

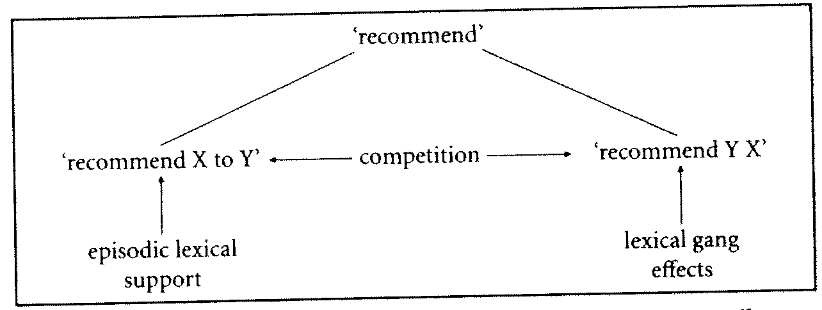

FIGURE 2.6. Competition between lexical argument frames and lexical gang effects

effects. In order to correct this overgeneralization, the child has simply to strengthen the alternative frame for the verb 'recommend'. To do this, it is necessary to focus on the actual use of the verb in the input and to record the specific episodes in which it is used in the 'recommend X to Y' form.

The Competition Model (MacWhinney 1988) treated the argument structures for inflectional morphemes as similar to the structures for verbs and prepositions. For the English past tense, the masking and extraction of the phonological form /Id/ from the verb 'jumped' produces a frame specifically linked to the sound and meaning of 'jump'. Over time, the meaning of the verb to which the past tense suffix is attached becomes generalized to any verb, and the sound of the suffix becomes generalized so that the /id/ form of the morpheme requires final dental consonants, whereas the /d/ and /t/ forms require final non-dentals and vowels.

2.5.5. *Modification Systems*

In addition to the basic maps for lexical associations and argument frames, Figure 2.3 can be supplemented with systems for phonological and semantic modification. Phonological modification operates to enforce general phonological patterns when words are linked together. Semantic modification works to adapt meanings when words are linked together. Often this linkage is predictable and obvious. However, sometimes words are placed together even at the expense of standard argument frames. Because connectionist systems are constraint satisfaction systems, rather than rule systems, they can deal with partial violations in the combinations of words. Consider a combination like 'another sand'. Typically, the word 'another' requires a count noun and 'sand' is a mass noun. However, when the listener is confronted with this particular combination, it is still possible to retrieve an interpretation by treating 'sand' as a count noun. This can be done by thinking of bags of sand, types of sand, alternative meanings of the word 'sand', or even the act of applying sandpaper to something. MacWhinney

(1989) talks about these semantic extension effects in terms of a process of 'pushy polysemy'.

2.6. Discussion and Conclusions

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

Viewing these efforts from the viewpoint of cognitive neuroscience, we can say that our efforts have concentrated on ways in which small areas of cortex can compute specific linguistic relations. But we know that language is processed in separate cortical and subcortical areas that must work in a concerted fashion across large distances, relying on connecting axonal pathways and cortical–subcortical connections. Understanding how divergent areas constitute functional neural circuits in which information is transferred without symbol passing is a major task facing connectionist modelling. There may still be an Achilles heel in the connectionist approach that will doom this whole process to failure. But, for now, we can look at the progress we have made as grounds for cautious optimism.

References

ANDERSON, J. (1978). 'Arguments Concerning Representations for Mental Imagery'. *Psychological Review*, 85, 249–77.

—— (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum.

BADDELEY, A. (1986). *Working Memory*. Oxford: Oxford University Press.

BAKER, C. L. and MCCARTHY, J. J. (eds.) (1981). *The Logical Problem of Language Acquisition*. Cambridge, Mass.: MIT Press.

BOWERMAN, M. (1988). 'The "No Negative Evidence" Problem'. In J. Hawkins (ed.), *Explaining Language Universals*. London: Blackwell, 73–104.

BURGESS, N. (1995). 'A Solvable Connectionist Model of Immediate Recall of Ordered Lists'. In G. Tesauro, D. Touretzky, and J. Alspector (eds.), *Neural Information Processing Systems*. San Mateo, Calif.: Morgan Kaufmann.

—— and HITCH, G. (1992). 'Toward a Network Model of the Articulatory Loop'. *Journal of Memory and Language*, 31, 429–60.

CARPENTER, G., GROSSBERG, S., and REYNOLDS, J. (1991). 'ARTMAP: Supervised Real-time Learning and Classification of Non-stationary Data by a Self-organizing Neural Network'. *Neural Networks*, 4, 565–88.

COHEN, M. and GROSSBERG, S. (1987). 'Masking Fields: A Massively Parallel Neural Architecture for Learning, Recognizing, and Predicting Multiple Groupings of Patterned Data'. *Applied Optics*, 26, 1866–91.

CORRIGAN, R., ECKMAN, F., and NOONAN, M. (eds.) (1989). *Linguistic Categorisation*. Amsterdam: John Benjamins.

COTTRELL, G. W. (1987). 'Toward Connectionist Semantics'. *TINLAP III*, 1–6.

—— and PLUNKETT, K. (1991). 'Learning the Past Tense in a Recurrent Network: Acquiring the Mapping from Meaning to Sounds'. In K. Hammond and D. Gentner (eds.), *Proceedings of the 13th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, 109–15.

DAELEMANS, W., GILLIS, S., and DURIEUX, G. (1993). 'The Acquisition of Stress: A Data-oriented Approach'. *Computational Linguistics*, 20, 421–51.

DELL, G. (1986). 'A Spreading-activation Theory of Retrieval in Sentence Production'. *Psychological Review*, 93, 283–321.

FODOR, J. A., BEVER, T. G., and GARRETT, M. F. (1974). *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.

FRANCIS, W. and KUCERA, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.

GLEITMAN, L. (1990). 'The Structural Sources of Verb Meanings'. *Language Acquisition*, 1, 3–55.

GOLDSMITH, J. (1976). 'An Overview of Autosegmental Phonology'. *Linguistic Analysis*, 2, 23–68.

GROSSBERG, S. (1987). 'Competitive Learning: From Interactive Activation to Adaptive Resonance'. *Cognitive Science*, 11, 23–63.

GUPTA, P. and MACWHINNEY, B. (1992). 'Integrating Category Acquisition with Inflectional Marking: A Model of the German Nominal System'. *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, 253–8.

—— —— (1994). 'Is the Articulatory Loop Articulatory or Auditory? Re-examining the Effects of Concurrent Articulation on Immediate Serial Recall'. *The Journal of Memory and Language*, 33, 63–88.

—— —— (1997). 'Vocabulary Acquisition and Verbal Short-term Memory: Computational and Neural Bases'. *Brain and Language*, 59, 267–333.

HARLEY, T. and MACANDREW, S. (1992). 'Modelling Paraphasias in Normal and Aphasic Speech'. *Proceedings of the 14th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, 378–83.

HERTZ, J., KROGH, A., and PALMER, R. (1991). *Introduction to the Theory of Neural Computation*. New York: Addison-Wesley.

HINTON, G. and SHALLICE, T. (1991). 'Lesioning an Attractor Network: Investigations of Acquired Dyslexia'. *Psychological Review*, 98, 74–95.

HOEFFNER, J. (1992). 'Are Rules a Thing of the Past? The Acquisition of Verbal Morphology by an Attractor Network'. *Proceedings of the 14th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

HOPFIELD, J. J. (1982). 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities'. *Proceedings of the National Academy of Sciences*, 79, 2554–8.

JAEGER, J. J. (1984). 'Assessing the Psychological Status of the Vowel Shift Rule'. *Journal of Psycholinguistic Research* 13, 13–36

KAWAMOTO, A. (1993). 'Non-linear Dynamics in the Resolution of Lexical Ambiguity: A Parallel Distributed Processing Account'. *Journal of Memory and Language*, 32, 474–516.

KIM, J., PINKER, S., PRINCE, A., and PRASADA, S. (1990). 'Why No Mere Mortal Has Ever Flown out to Centre Field'. *Cognitive Science*, 15, 173–218.

KLAHR, D. and KOTOVSKY, K. (1991). *Complex Information Processing: The Impact of Herbert A. Simon.* Hillsdale, NJ: Erlbaum.

KOHONEN, T. (1982). 'Self-organized Formation of Topologically Correct Feature Maps'. *Biological Cybernetics*, 43, 59–69.

KÖPCKE, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache.* Tübingen: Niemeyer.

—— (1988). 'Schemas in German Plural Formation'. *Lingua*, 74, 303–35.

—— (1996). 'Prinzipien für die Genuszuweisung im Deutschen'. In E. Lang and G. Zifonun (eds.) *Deutsch—Typologisch.* Berlin: Walter de Gruyter.

—— and ZUBIN, D. A. (1983). 'Die kognitive Organisation der Genuszuweisung zu den einsilbigen Nomen der deutschen Gegenwartssprache'. *Zeitschrift für germanische Linguistik*, 11, 166–82.

—— —— (1984). 'Sechs Prinzipien für die Genuszuweisung im Deutschen: Ein Beitrag zur natürlichen Klassifikation'. *Linguistische Berichte*, 93, 26–50.

LAMB, S. (1966). *Outline of Stratificational Grammar.* Washington, D.C.: Georgetown University Press.

LINELL, P. (1979). *Psychological Reality in Phonology: A Theoretical Study.* Cambridge: Cambridge University Press.

LING, C. (1994). 'Learning the Past Tense of English Verbs: The Symbolic Pattern Associator vs. Connectionist Models'. *Journal of Artificial Intelligence Research*, 1, 209–29.

—— and MARINOV, M. (1993). 'Answering the Connectionist Challenge'. *Cognition*, 49, 267–90.

MCCLELLAND, J. L. and RUMELHART, D. E. (eds.) (1986). *Parallel Distributed Processing.* Cambridge, Mass.: MIT Press.

—— and ELMAN, J. (1986). 'Interactive Processes in Speech Perception'. In McClelland and Rumelhart (eds.), 58–121.

—— and KAWAMOTO, A. (1986). 'Mechanisms of Sentence Processing: Assigning Roles to Constituents'. In McClelland and Rumelhart (eds.), 272–325.

MACWHINNEY, B. (1978*a*). 'The Acquisition of Morphophonology'. *Monographs of the Society for Research in Child Development*, 43, Whole no. 1, 1–233.

—— (1978*b*). 'Conditions on Acquisitional Models'. *Proceedings of the 1978 Annual Conference.* New York: Association for Computing Machinery.

—— (1988). 'Competition and Teachability'. In R. Schiefelbusch and M. Rice (eds.), *The Teachability of Language.* New York: Cambridge University Press, 63–104.

—— (1989). 'Competition and the Lexical Categorization'. In Corrigan, Eckman, and Noonan (eds.), 195–242.

—— (1993). 'Connections and Symbols: Closing the Gap'. *Cognition*, 49, 291–6.

—— (1994). 'The Dinosaurs and the Ring'. In R. Corrigan, S. Lima, and M. Noonan (eds.), *The Reality of Linguistic Rules.* Amsterdam: Benjamins, 283–320.

—— and ANDERSON, J. (1986). 'The Acquisition of Grammar'. In I. Gopnik and M. Gopnik (eds.), *From Models to Modules.* Norwood, NJ: Ablex, 3–25.

—— and LEINBACH, J. (1991). 'Implementations are not Conceptualizations: Revising the Verb Learning Model'. *Cognition*, 29, 121–57.

—— —— TARABAN, R., and MCDONALD, J. L. (1989). 'Language Learning: Cues or Rules?' *Journal of Memory and Language*, 28, 255–77.

MARATSOS, M. and CHALKLEY, M. (1980). 'The Internal Language of Children's Syntax: The Ontogenesis and the Representation of Syntactic Categories'. In K. Nelson (ed.), *Children's Language* ii. New York: Gardner, 127–214.

MARCHMAN, V. (1992). 'Constraint on Plasticity in a Connectionist Model of the English Past Tense'. *Journal of Cognitive Neuroscience*, 5, 215–34.

MARSLEN-WILSON, W. D. (1987). 'Functional Parallelism in Spoken Word-recognition'. *Cognition*, 25, 71–102.

MEISEL, J. (1995). 'Parameters in Acquisition'. In P. Fletcher and B. MacWhinney (eds.), *Handbook of Child Language.* Oxford: Blackwell, 10–35.

MIIKKULAINEN, R. (1990). 'A Distributed Feature Map Model of the Lexicon'. *Proceedings of the 12th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum, 447–54.

—— and DYER, M. (1991). 'Natural Language Processing with Modular Neural Networks and Distributed Lexicon'. *Cognitive Science*, 15, 343–99.

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

MORGAN, J. and TRAVIS, L. (1989). 'Limits on Negative Information in Language Input'. *Journal of Child Language*, 16, 531–52.

NESPOR, M. and VOGEL, I. (1986). *Prosodic Phonology.* Dordrecht: Foris.

NEWELL, A. and SIMON, H. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall.

NORRIS, D. (1994). 'Shortlist: A Connectionist Model of Continuous Speech Recognition'. *Cognition*, 52, 189–234.

OHALA, J. J. (1974*a*). 'Experimental Historical Phonology'. In J. M. Anderson and C. Jones (eds.), *Historical Linguistics* ii. New York: Elsevier, 353–89.

—— (1974*b*). 'Phonetic Explanation in Phonology'. In A. Bruck, R. Fox, and M. La Galy (eds.), *Papers from the Parasession on Natural Phonology.* Chicago: Chicago Linguistic Society, 251–74.

OHALA, M. (1974). 'The Abstractness Controversy: Experimental Input from Hindi'. *Language*, 50, 225–35.

PINKER, S. (1984). *Language Learnability and Language Development.* Cambridge, Mass.: Harvard University Press.

—— (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge, Mass.: MIT Press.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., and Patterson, K. (1996). 'Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-regular Domains'. *Psychological Review*, 103, 56–115.
Rumelhart, D.E. and McClelland, J.L. (1986). 'On Learning the Past Tense of English Verbs'. In Rumelhart and McClelland (eds.), 216–71.
—— Hinton, G.E., and Williams, R. (1986). 'Learning Internal Representations by Backpropagation'. In Rumelhart and McClelland (eds.), 318–62.
St. John, M. (1992). 'The Story Gestalt: A Model of Knowledge-intensive Processes in Text Comprehension'. *Cognitive Science*, 16, 271–306.
Seidenberg, M. and McClelland, J.L. (1989). 'A Distributed, Developmental Model of Word Recognition and Naming'. *Psychological Review*, 96, 523–68.
Squire, L. (1987). *Memory and Brain*. New York: Oxford University Press.
Stemberger, J. (1985). *The Lexicon in a Model of Language Production*. New York: Garland.
Touretzky, D. (1990). 'BoltzCONS: Dynamic Symbol Structures in a Connectionist Network'. *Artificial Intelligence*, 46, 5–46.
Trammell, R. (1978). 'The Psychological Reality of Underlying Forms and Rules for Stress'. *Journal of Psycholinguistic Research*, 7, 79–94.
Twain, M. (1935). 'The Awful German Language'. *The Family Mark Twain*. New York: Harper & Brothers.
von Neumann, J. (1956). 'Probabilistic Logistics and the Synthesis of Reliable Organisms from Unreliable Components'. In C. Shannon and J. McCarthy (eds.), *Automata Studies*. Princeton: Princeton University Press.
Wexler, K. and Culicover, P. (1980). *Formal Principles of Language Acquisition*. Cambridge, Mass.: MIT Press.
Zubin, D.A. and Köpcke, K.M. (1981). 'Gender: A Less than Arbitrary Grammatical Category'. In R. Hendrick, C. Masek, and M. Miller (eds.), *Papers from the 17th Regional Meeting*. Chicago: Chicago Linguistic Society, 439–49.
—— —— (1986). 'Gender and Folk Taxonomy: The Indexical Relation between Grammatical and Lexical Categorization'. In C. Craig (ed.), *Noun Classes and Categorisation*. Amsterdam: John Benjamins, 139–80.

3

Are SRNs Sufficient for Modelling Language Acquisition?

NOEL SHARKEY, AMANDA SHARKEY and STUART JACKSON

3.1. Introduction

Connectionist models of natural language processing have attracted considerable interest because of their simple learning mechanisms and their representational capabilities. Not surprisingly, much of the focus of recent research has been concerned with modelling aspects of language acquisition (e.g. Rumelhart and McClelland 1986; Elman 1990, 1991; Plunkett and Marchman 1991), and the findings have challenged old conceptions about the need for the learning of rules and the types of representations that are required for language development. It is early days yet, just over a decade (see Sharkey and Reilly 1992), and there appears to be reason for optimism about the prognosis for connectionist accounts of learning and representation. In this chapter, we assess whether this optimism is appropriate for some of the broader issues of language learning.

One of the problems faced by the modellers of the early and mid 1980s was how to represent the temporal characteristics of language with artificial neural nets. The artificial neural nets used had only feedforward connections and thus could not preserve a memory of prior inputs. Although some methods were used to get around problems of sequentiality such as assemblies or frames (Hinton 1981; McClelland and Rumelhart 1981), moving input windows (Sejnowski and Rosenberg 1986), and Wickelfeatures (Rumelhart and McClelland 1986), these amounted to using the input space as fixed width memory buffer.

A seemingly natural step forward has been to augment feedforward network architecture with recurrent or feedback links that preserve a *fading* memory of the past. One of the most popularly used recurrent nets for modelling language, developed by Elman (1988), is the Simple Recurrent Net (SRN) as illustrated in Figure 3.1. The task of the SRN is to learn to predict the legal successors of the current input (e.g. as in grammar recognition). This approach has yielded many successes both in simple grammar learning (for example Servan-Schreiber, Cleeremans, and McClelland 1991), and in creating appropriately structured lexical representations (for example Elman 1990). Moreover, the SRN has been