

# LANGUAGE ACQUISITION ACCORDING TO CONNECTIONISTS: MODELS OF LEARNING MORPHOLOGY

Jenny Hayes

Department of Psychology, University of Hertfordshire  
J.Hayes@herts.ac.uk

*The challenge for supervised neural net models of morpho-syntax has been to demonstrate that language learning that appears to entail a data base of rules and exceptions can be simulated without the need for these structures to be present. This article reviews connectionist models of morpho-syntax which have attempted to meet this challenge. The article begins with a background description of how connectionist models work and then proceeds to explain the way in which both static and sequential models of the acquisition of morpho-syntax have been developed. That static connectionist models have been able to simulate the development of both verbal and nominal morphology is discussed in the context of how these models question the dual mechanism model of morphological processing (Pinker, 1999). The role of sequential connectionist models in understanding of the treatment of plural nouns in compound words is considered in detail. Finally, how morpho-syntax might be learnt is considered in an overview of developmental models.*

**Keywords:** connectionist modelling, morpho-syntax, language acquisition, supervised neural network models

## 1. INTRODUCTION

There has been an assumption that learning a language involves learning a series of rules and exceptions to those rules. This view has been challenged by connectionist modellers who have been able to demonstrate that behaviour, which appears to entail a database

of rules and exceptions, can be simulated without these structures being present. Thus, connectionist modellers have questioned the dual mechanism model of morphosyntactic processing; presented evidence that morpho-syntax can be learnt simply by exposure to the linguistic input and identified mechanisms by which language acquisition might proceed. Several models that have investigated morpho-syntactic aspects of language processing in supervised connectionist architectures are reviewed here together with an overview of how connectionist models are constructed and trained.

---

Address for correspondence:

Jenny Hayes  
Department of Psychology  
University of Hertfordshire  
Hatfield Campus  
College Lane  
AL10 9AB  
Hatfield, AL10 9AB, U.K.

## 2. THE CHALLENGE FOR CONNECTIONIST MODELS OF MORPHO-SYNTACTIC ASPECTS OF LANGUAGE PROCESSING

Classical (symbolic) models assume that human cognition includes the capacity to use stored mental rules to process input from the environment (Fodor & Pylyshyn, 1988). Implicit in Chomsky's (1959) idea, that children use some innate, language specific mechanism to uncover the underlying grammatical rules of their native language is the notion that there are rules there to be discovered. This belief that language learning involves the acquisition of a series of rules (Chomsky; Fodor & Pylyshyn; Pinker, 1999) is driven in part by examples of overregularisation errors produced by children learning the past tense of English verbs (saying *eated* rather than *ate*) or the plural of English nouns (saying *mouses* rather than *mice*). The child will not hear these overregularisation errors in the language they hear yet they still make them. Overregularisation errors have been seen as evidence for the misapplication of the stored rule "to form past tense add -ed to the verb stem" or "to form the plural add -s to the noun stem". Children would be expected to make errors as they learn a skill. That children can make subject and verb agree in a sentence they have never heard before is also thought to be evidence of possession of a general rule about the relation between subject and verb that can be applied to any sentence.

Rumelhart and McClelland (1986) and other connectionist modellers (e.g. Elman, Bates, Johnson, Karmiloff-Smith, Parisi and Plunkett, 1996) while agreeing that it may be possible to describe language in rule-like terms, argue that there might not actually be any rules available for the child to represent. Rumelhart and McClelland and many oth-

er connectionist modellers have been able to simulate rule like behaviour in artificial neural networks that have no specific knowledge of the rules of grammar. Connectionist models do not have explicit declarative rules of the kind "to form past tense add -ed to the verb stem" or "to form plural add -s to noun stem". Neither do they have a specific memory store to accommodate the list of exceptions that would be required to override the application of this rule to irregular verbs (verbs such as *eat* that do not form the past tense by adding -ed to the verb stem or nouns such as *mouse* that do not form the plural by adding -s to the noun stem). In connectionist models knowledge about all verbs, both regular and irregular, is stored in the same general matrix of information. The connectionist view is that general associative memory processes are used to learn language. These processes are guided by the fact that language appears in highly regular patterns (Saffran, 2001) and the way learning proceeds is influenced by the frequency with which linguistic items appear in the input during the acquisition process.

The challenge for connectionist models of cognition has been to demonstrate that behaviour, which appears to entail a database of rules and exceptions, can be simulated without these structures being present. Connectionist models have been used to investigate many areas of cognition e.g. reading aloud (Harm, McCandliss, & Seidenberg, 2003; Harm & Seidenberg, 1999, 2001); syntactic processing (Seidenberg, 1997, 1999; Seidenberg & Elman, 1999); spoken word recognition, (Elman & McClelland, 1986; Christiansen, Allen, & Seidenberg, 1998) and lexical access (Dell, Chang & Griffin, 1999; Dell, Schwartz, Martin, Saffran, & Gagnon 1997; Dell, Burger & Svec, 1997).

### 3. HOW CONNECTIONIST MODELS WORK

#### *Neurally inspired*

Connectionism (also known as “parallel distributed processing or “neural networks”) is an attempt to design computer models inspired by how the brain might process information. The brain is thought to consist of a large number of simple processors called neurons that are densely interconnected in a complex network. These neurons appear to work simultaneously and cooperatively to process information. Connectionist models attempt to simulate these properties. They consist of large numbers of simple processors called units that are densely interconnected in a complex network. They operate simultaneously and in cooperation with each other. At the beginning of training the weight of connection between any two units is random. The network is trained on a representation of stimuli that the human brain has to process. This could be a representation of a human face and a representation of the name of the person with that face. The weights connecting units that represent certain faces and units that represent certain names will be strengthened as the network is exposed to particular faces being linked to particular names during training. Conversely weights will be weakened between units representing faces and names that are not paired together in the training set.

#### *Parallel processing and distributed representations*

Connectionist models involve parallel processing usually of distributed representations. This is a departure from classical models of cognition in which computation is serial and knowledge representation is local.

Serial processing means that each process is carried out in sequence and the outcome of one process affects the processing of the next stage. Parallel processing means that more than one process is carried out concurrently. Localist representation means that each stimulus item is represented by one token and this token is stored in a space in the model totally independently from all other tokens. In localist models, for example, information about the word *dog* is stored in one place and information about the word *cat* is stored in another place. Dictionaries and telephone directories use localist representation. Localist coding is sometimes referred to as “grandmother cell representation” because it would suggest that we have a cell tuned uniquely to each possible pattern, including a cell for detecting our grandmother. Distributed representation (as used in connectionist models) means that there is no one place where a particular stimulus item can be located. Knowledge about *dog* is distributed across many units in a connectionist model. Both *dog* and *cat* might be connected to a unit “*has fur*” but only *dog* would be linked to “*barks*” and only “*cat*” would be linked to “*meows*.” Distributed representations do not directly correspond with the individual features of the stimuli being encoded. Instead, patterns of activity in many (often all) of the cells in a network collectively encode the stimulus. It may be possible to determine the elements of the stimuli which each cell in the model is preferentially responsive to but the point is that there is no one to one mapping between elements of the stimuli and the cells in the neural net. This means that the relationship between external stimuli and the internal representations in a neural net are difficult to interpret unambiguously and therefore difficult to analyse. However, distributed representations are more economical (require

fewer neurons) and they permit greater generalization than localist representations. Furthermore, distributed representations permit access to an item even when full knowledge of that item is not available to the enquirer. In this way connectionist models are like human memory in that they are content addressable. Using distributed representation even if the word *dog* has been forgotten (or it is presented in noise) it can be triggered (just as it could from a person) by clues such as "*it barks*" *it has fur*". Access would not be possible from a localist system (a dictionary) if the word string *dog* had been forgotten. Localist representations also make it difficult to respond to in-between stimuli, e.g. if there is a white car detector and a black car detector, a grey car will not be recognized. Although in the early PDP models representations were by definition distributed, some researchers e.g. Gary Dell (Dell, Chang & Griffin, 1999; Dell, Schwartz, Martin, Saffran, & Gagnon 1997; Dell, Burger & Svec, 1997) have developed connectionist models using localist coding schemes. It is now established that the distinction between localist and distributed coding is not a dichotomy, rather completely distributed and completely localist schemes lie at extremes of a continuum of the 'coarseness' of coding used by connectionist modellers.

Although connectionist models are neurally inspired, it is important to note that it is not claimed that they are realistic models of the brain. Connectionist modellers claim only that their models might provide a useful starting point for understanding how cognitive computations might be performed (Christiansen & Chater, 1999). Influential work that has led to our current understanding of the role connectionism might play in understanding cognition include Rosenblatt, (1958), Amari, (1967), Grossberg, (1967), Kohonen, (1972, 1989), Oja,

(1982), Sejnowski, & Rosenberg (1986), Haykin, (1999).

#### 4. TYPES OF CONNECTIONIST MODEL

The main division between types of neural net models is between models which learn in a supervised or unsupervised environment. Models that are driven by supervised learning proceed under the control of a "teacher signal" that gives the network feedback about its performance. In contrast unsupervised systems require no feedback. Unsupervised systems organise themselves on the basis of the statistical properties in the input, irrespective of whether their outputs have the desired consequences for later stages of analysis.

Kohonen artificial neural networks (KANNs) (Kohonen, 1989) for instance fall into the category of "unsupervised learning" because the multivariate algorithm used seeks out "clusters" in the data (Everitt, 1993). Unsupervised learning allows the network to group items together on the basis of their perceived closeness in  $n$ -dimensional hyperspace (where  $n$  is the number of variables or observations made on each item). During the training process the network is presented with each input pattern in turn, and all the nodes calculate their activation levels on the basis of the Euclidean distance (straight line distance between two points) between them and the input vector in  $n$ -dimensional space. Thus a node whose weight vector closely matches the input vector will have a small activation level, and a node whose weight vector is very different from the input vector will have a large activation level. The node in the network with the smallest activation level is deemed to be the "winner" for the current input vector. The winning node and some of the nodes

around it are then allowed to adjust their weight vectors to match the current input vector more closely. The nodes included in the set which are allowed to adjust their weights are said to belong to the "neighbourhood" of the winner. The size of the winner's neighbourhood is varied throughout the training process. Initially all of the nodes in the network are included in the neighbourhood of the winner, but as training proceeds the size of the neighbourhood is decreased linearly after each presentation of the complete "training set" (all the items being analysed), until it includes only the winner itself. The amount by which the nodes in the neighbourhood are allowed to adjust their weights is also reduced linearly throughout the training period. The factor which governs the size of the weight alterations is known as the learning rate. The effect of the "learning rule" (weight update algorithm) is to distribute the neurons evenly throughout the region of  $n$ -dimensional space populated by the training set (Hecht-Nielsen, 1990; Kohonen, 1989). The neuron with the weight vector closest to a given input pattern will win for that pattern and for any other input patterns that it is closest to. Input patterns which allow the same node to win are then deemed to be in the same group, and a map of their relationship can be drawn with a line enclosing them. By training with networks of increasing size a map with several levels of groups or "contours" can be drawn. Construction of these maps allows close examination of the relationships between the items in the training set. Unsupervised models have been used to investigate many areas of language acquisition such as syntactic processing (Scholtes, 1992a, 1993) and semantics and pragmatics (Honelka, 1997, 2000; Honelka, Pulkki and Kohonen, 1995; Honelka and Vepsäläinen, 1991). Morphological processing, the focus of this article,

has also been investigated using unsupervised neural net models, Creutz & Lagus, (2002).

Supervised models incorporate a teacher signal that produces an error signal if for instance the wrong past tense is output in response to a verb stem and the weights in the model are adjusted. The success of a model is judged by its error. That is, how much difference there is between the actual output the model produces and what the target output should be. This difference is squared (to remove negative values) and then the square root is calculated and this figure, termed "root mean square" is usually quoted as the error term in connectionist papers.

Supervised connectionist models that have been used by those interested in language acquisition have been of two main types. These are models that can learn static patterns and models that can learn sequential patterns.

#### *4.1 Models for static patterns*

Static models learn that a particular input should be paired with a particular output. For instance static models of the English past tense can learn to output the "ed" suffix when a regular verb is input, but not to output it when an irregular verb is presented. To understand how a static connectionist model works it is useful to take the example of a model that learns to pair certain sounds with certain patterns of letters (a model that learns to simulate reading aloud). Connectionist models have an input layer where the stimulus is presented to the network. In this case the stimulus would be a representation of a word. The network would be required to output the correct pronunciation for that word. At first (early in training) the network would find it difficult to produce the right sound output to correspond with each word input and would

produce a random output (similarly to a child learning to read). The network is then corrected on what the correct pronunciation should be. The network learns from being corrected and changes the weights of the connections between the input and output units in such a way that the next time the word is presented at input it will produce an output which is closer to the correct pronunciation. Thus, over the period of training, a successful model builds a set of connections that will be able to pair the correct output with each input. Figure 1 shows a typical architecture of a static model. During training, similar input patterns become represented in a similar form in the layers (termed hidden layers) between the network's input and output layers thus facilitating the network's ability to generate the correct output for any input, including previously unseen items.

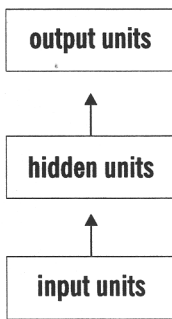


Figure 1. Typical architecture of a static connectionist model

*Implications of connectionist modelling in relation to a dual or single route mechanism*

A number of static models have been influential in demonstrating that both verbal and nominal morphology in English can be learnt using single route architectures with only an associative memory system and without the need for declarative rules of the kind "to form past tense add -ed to the verb stem" or to

make plural add -s to the noun stem". As such static models have been able to offer alternative explanations for many lines of evidence put forward in support of a dual route model of morphology. Pinker (1991, 1994, 1999) and others (e.g., Marcus, Brinkmann, Clahsen, Wiese & Pinker, 1995) have developed the dual mechanism model that attempts to unite the classic symbolic view of language with associative memory based accounts of language processing. The dual mechanism model proposes that items of regular inflectional morphology are rule governed but less systematic features of language such as irregular verbs and nouns are learned and represented using associative memory systems. Thus, irregulars are learned on a case by case basis. However, they are not simply learned as separate examples by rote memory systems and stored as unique, isolated items. Instead, items which share phonetic similarity (e.g. *sing/ sang/ sung; ring/ rang/ rung*) appear to have overlapping representations (Chandler, 1993).

A series of connectionist models that have been able to learn items of inflectional morphology are described below. In each of these models any learning that takes place is driven by associative memory processes learning patterns and frequencies recoverable from the input. As such, these models raise questions for the viewpoint that inflectional morphology is mediated by a dual mechanism involving both symbolic and associative learning.

*The role of input frequency in learning by single route models*

Learning majority default systems

Inflectional systems such as English past tense, where regular (default) morphology has a much higher type frequency than irregular

morphology are referred to as majority default systems. Rumelhart and McClelland (1986a), who were the first to develop a neural network able to demonstrate that a single mechanism might be sufficient to learn an aspect of morphology, used a static connectionist model and modelled the English past tense (a majority default system). Their model learnt to output the "ed" suffix when a regular verb is input, but not to output it when an irregular verb is presented. Rumelhart and McClelland were able to model the three stages that children demonstrate, to some degree or other, in acquiring the past tense of some English verbs (Bowerman, 1982; Brown, 1988; Marcus, Pinker, Ullman, Hollander, Rosen and Xu, 1992). In the first stage, children only use a few past tense verbs and these tend to be mainly high frequency irregular verbs on which they make few errors. In the second stage children begin to use many past tenses, the majority of which are regular verbs. At this stage children appear to have learned rules which guide their behaviour, in that they make overregularizations on irregular verbs which they could use correctly at stage one (e.g., using 'buyed' instead of 'bought' as the past tense of the verb *to buy*). This, it is argued, must be due to rule learning because the children never hear these overregularisations in the input they receive. At stage three they stop making overregularization errors and are able to use the correct form of irregular and regular verbs. This has been described as the U-shaped profile of learning. By altering the balance between the frequency of regular and irregular verbs in the input to their model, at various stages of learning, Rumelhart and McClelland demonstrated evidence of these three stages of learning in their model. Initially the network was trained on eight highly frequent irregular verbs and 2 regular verbs and the net showed performance similar to children at stage 1. The training set was then

changed to 420 medium frequency verbs (80% of which were regular verbs) and initially the network showed evidence of overregularising irregular verbs but of being able to produce the correct past tense for regular verbs (i.e., similar to stage 2 of child behaviour). Later in training, the network made few errors in forming the correct past tense of the 420 verbs (stage 3 behaviour). Furthermore, when the network was tested on 86 unseen low frequency verbs (80% regular), it demonstrated an ability to generalise to these new forms.

Rumelhart and McClelland's (1986a) model has received much criticism (e.g. Pinker & Prince, 1988) mainly because the increase in training from 10 to 420 verbs is not representative of the exposure to the 2 types of verb that children receive. However, Plunkett and Marchman (1991) addressed this issue and successfully modelled the U shaped profile using a training set in which the size of the vocabulary was held constant at 500 verbs. Plunkett and Marchman trained a network with an architecture similar to that shown in Figure 1 using an artificial lexicon of verb stems and past tenses. The artificial verbs mimicked the phonological patterns found in English verbs. Importantly, Plunkett and Marchman (1991) found that their network could learn irregular past tenses if the type and token ratios approximated those in English. In a later simulation, Plunkett and Marchman (1993) gradually increased the training set from from 20–500 verbs. From the results of this simulation, Plunkett and Marchman concluded that a critical mass of exposure to verbs is needed before the change from rote learning (memory) to system building (rule like behaviour) can occur (Marchman & Bates, 1994; Plunkett and Marchman, 1996). Indicating that exposure to the linguistic input plays a critical role in acquiring

morphology. MacWhinney and Leinbach (1986) and Cottrell and Plunkett (1991) also produced successful models of past tense acquisition, having addressed the criticisms levelled at the Rumelhart and McClelland model.

An aspect of the success of the Rumelhart and McClelland (1986) model was that it was able to demonstrate (regardless of whether the frequencies represented were realistic or not) that the frequency of the two types of morphology had a direct effect upon learning about regular and irregular morphology in English. However, Prasada and Pinker (1993) have argued that the fact that connectionist models rely so heavily on the balance of frequency between regular and irregular morphology, is actually a disadvantage of this approach to language learning.

#### Learning minority default systems

Prasada and Pinker (1993) argue that Rumelhart and McClelland were only able to demonstrate generalisation in models of the English past tense because of the particular frequency make-up of the English verbs. The default past tense ending (i.e. the regular [-ed] ending) has (by far) the highest type frequency in the input but many regular verbs have low individual token frequencies. Irregular verbs have low a type frequency but many individual verbs have high token frequencies (e.g., *go-went, see-saw*). Prasada and Pinker argue that this distribution pattern allows the networks to construct a system in which irregulars are represented as a series of phonological sub-categories and all other verbs are mediated by a large default category. Thus, an inflectional system such as the German or Arabic plural system which has a default ending, which has both low type frequency and low token frequency could not be modelled. Furthermore, Mar-

cus, et al (1995) have argued that the German plural system is quite arbitrary in that while there are some patterns of gender and phonology which dictate which plural ending is applied to nouns, there are long lists of exceptions to each pattern. However, a series of connectionist models with different architectures developed by Hahn and Nakisa (2000) were able to predict approximately 80% of German plural forms. Thus, they demonstrated that neural net models were able to learn the underlying structure of German plurals. Both single route and dual route computational models were tested and interestingly it was found that dual route models did not show superior performance. By actually building and testing a dual route model, Hahn and Nakisa were able to demonstrate the process that a dual route system would have to undergo in order to produce the correct plural ending. Firstly, any noun selected was "looked up" in the associative memory system to see if the plural form was stored. If the item was not found (various thresholds of activation of the associative memory store were tested before the rule was applied), the rule would be adopted and the default ending was applied. The argument had been that single route models would not be able to cope with a minority default because they work by learning the small number of irregulars and applying the rule to the vast majority of other items. Thus, in the case of a minority default there would be too many items to have to learn and store in an associative mechanism. By instantiating a dual route model, Hahn and Nakisa demonstrated that as dual route models also have a pattern-associator facility, they are just as dependent as single route models on the balance between the frequency of regular and irregular plurals in the input. The rule is only applied once the item has not been found in associative memory, so the pattern associator



element of a dual mechanism model, also has to store a large number of examples where a minority default situation exists. Plunkett and Nakisa (1997) have also successfully simulated learning of the Arabic minority plural system and Daugherty and Hare (1993) and Hare, Elman and Daugherty (1995) have modelled old English verbs, which also have a minority default (i.e. only 17% of items have regular past tenses).

Learning when the type frequency of the irregular category is very low

A further criticism of connectionist models of morphological acquisition centres around Marcus' (1995a) claim that while it was possible to model the acquisition of the past tense of verbs it would not be possible to model the acquisition of the plurals of English nouns. This was because the success of the connectionist models that simulated learning of the past tense was driven by the fact that there were sufficient numbers of irregular verbs to stop items being overregularised at an unrealistic rate. However, Marcus claimed that there may not be a sufficient critical mass of irregular nouns to stop all nouns being regularised by a connectionist model. Like the past tense, English regular plurals involve the addition of a suffix and like many irregular verbs, several irregular plurals are formed by changing the internal vowel in the stem (e.g., *goose* becomes *geese*). Marcus, Brown (1973), and Marchman, Plunkett and Goodman (1997) have reported similar time courses for the acquisition of both types of morphology, and evidence that the U shaped curve of development occurs in both types of morphology. However, as Marcus points out, there are also differences. While there are approximately 100 commonly used irregular verbs (e.g., *go-went, see-saw*), there are only seven frequently used

irregular plurals in English (*man-men, woman-women, child-children, tooth-teeth, foot-feet, mouse-mice, goose-geese*). However, Marchman et al (1997) showed that irregular plurals are frequently exemplified in children's early lexicons. This high token frequency of irregular plurals stops these items being dominated by the far more type frequent regular plurals. Plunkett and Juola (1999) have in fact developed a model of English past tense and plural morphology using a single mechanism connectionist network. Plunkett and Juola's model showed a similar developmental profile as children, in that nouns were learned more quickly than verbs and early performance was characterised by few errors but later performance saw the development of the U-shaped profile for both nouns and verbs.

#### *Models addressing the behavioural evidence for a dual route model*

Thus, connectionist models have been able to learn static patterns of both verbal and nominal inflectional morphology when the default is both the majority or the minority category. They have also been able to learn to produce the correct morphology when the irregular category is very small. These models have all been used to argue for a single route mechanism of inflectional morphology. Other models (examples of which are described below) have addressed some of the behavioural evidence put forward by supporters of the dual mechanism model.

The effects of frequency on irregular (but not regular morphology)

One of the most frequently cited lines of evidence for the dual mechanism model is that irregular morphology (because it is stored with the stem in the lexicon) is subject to fre-

quency effects but regular morphology is not (Pinker, 1991). Daugherty and Seidenberg (1994) have demonstrated that neural nets can also account for this phenomenon. Regular, “rule governed” words have phonetic patterns that are very frequent in the input. In other words, regular verbs have lots of “neighbours” that are similar in sound. However, irregulars have far fewer phonetic neighbours. Thus, performance on irregulars depends far more on how often the language learner is exposed to these items (than is the case for regulars) because the correct past tense cannot be learned from a large number of similar examples.

Evidence that brain injured patients are impaired on the production of either regular or irregular inflections

Joanisse and Seidenberg (1999) addressed the evidence that some brain injured patients seem to be impaired on producing regular morphology, and others seem to be impaired on producing irregular past tenses. This, it is argued, provides evidence for the fact that the two types of morphology are mediated by two separate areas of the brain. Joanisse and Seidenberg showed that damage to either phonological information or semantic information within a single route model can simulate these types of impairments.

Single route static models have been able to simulate the development of both verbal (Rumelhart and McClelland, 1986a; Plunkett and Marchman, 1991) and nominal morphology (Plunkett and Juola, 1999) in English and have been able to find alternative explanations for many lines of evidence put forward in support of a dual route model of morphology in English (e.g. Daugherty and Seidenberg, 1994; Joanisse and Seidenberg, 1999).

#### 4.2 Sequential models

In addition to the use of static models by psycholinguists, many researchers have sought to investigate how the sequential processing of language might be represented in a neural network, a mechanism based on parallel computation. One approach has been to represent time implicitly through its effects on processing rather than explicitly in the architecture of a model. Elman's (1990) simple recurrent network (SRN) uses recurrent links, as first suggested by Jordan, 1986, between the hidden units (layers between the input and the output units) and other units, termed context units, that store representations of prior internal states (weights between connections) of the network. Context units can be thought of as the network's memory stores. As the context units link back to the hidden units, at any point in time the state of the hidden units at the previous time step are used as additional input in the same way that a child processing language would use prior memories. A typical architecture for a SRN is shown in Figure 2.

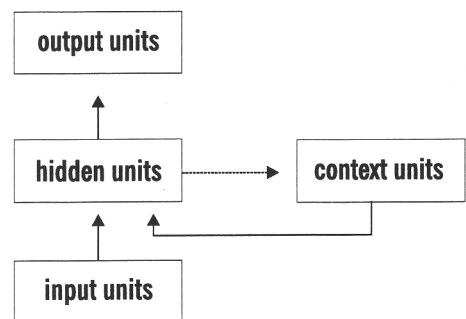


Figure 2. Typical architecture of a sequential connectionist model (dotted lines indicate non-trainable units).

The hidden units output to the output units but also to the context units from where this output is fed back into the hidden layers as additional input. SRN's tend to be used for

prediction tasks where the model is expected to output the next item it expects given the elements it has seen previously. SRNs are self-supervising models, rather than having a teacher signal they work on the basis that language users expect to hear similarities to his or her own constructions in the speech of others. Thus, SRNs predict what word will be heard next in the input (what they would say next) and this is compared with the actual word in the input stream (what the other speaker actually says). If the prediction is accurate then the network will alter its weights to make that response more likely in the future. If the prediction is not confirmed then the weights will be altered slightly to make that response less likely.

A major distinction among sequential models concerns whether or not a model allows feedback from logically later to logically earlier levels of processing. A disadvantage of SRNs is that they do not allow feedback from logically later to logically earlier levels of processing. At any timeslot, an SRN can only produce one output and the state of that output is dependent on the state of the network at that precise timeslot. Thus, SRNs are only capable of predicting what item might come next in the sequence, they are not able to change their prediction subsequently based on items that come later in the input. An SRN might predict a word ending marker after the input *cat* but could not change its prediction to *catapult* if rather than a word ending marker the letter *a* was the next item in the input. In the field of spoken word recognition, the TRACE model (McClelland & Elman, 1986) attempted to deal with the temporal dimension of speech by having many copies of the entire network representing different points in time. By using this large and implausible architecture TRACE can match sets of phonemes to words and can revise its decisions

in the light of subsequent context because it has access to (copies of) the network at all time stages (including previous time stages). Thus, TRACE can change its prediction of a word from *cat* to *catalogue*. It is, however, unintuitive to have a copy of the network available at each time step and SHORTLIST (Norris, 1994) represented an adaptation of TRACE, which only processed a short list of words which were likely to appear next in any given context.

Elman (1990) trained a simple recurrent network to discover word boundaries from a concatenated stream of letters. Thus, Elman sought to investigate whether a network using parallel processing could discover the notion of "word", a sequential pattern. He was interested in whether the concept of words emerge from learning sequential patterns of letters in which word boundaries were not marked. The sequence of letters was formed from a lexicon of 15 words using a sentence generation tool. Two hundred sentences, varying in length from 4 to 9 words were created. The sentences were then concatenated to form a string of words. The words were then broken down into the letters from which they were constructed. The task for the network was to predict the next letter. It was impossible for the network to learn the sequence in the 10 presentations of the data that it received (and indeed it does not). It is characteristic of this type of model that during training the error on predicting the next input does not decrease to any great extent. However, while the error is relatively high at the start of a new word, as more letters are presented to the network the RMS error declines since the word becomes more predictable (See Figure 3.). The network also has high errors on the *y* when *e* is input as part of *the y....* because it has been exposed to the highly frequent pattern *the*.

Elman did not intend this simulation to

be treated as a model of word acquisition. The simulation simply serves as a demonstration of the fact that there seems to be information in the input that could serve as a cue to the boundaries of linguistic input (such as word end), which can be learned. Saffran, Aslin and Newport (1996) found that 8-month old infants could learn transitional properties (word boundaries) from exposure to the input.

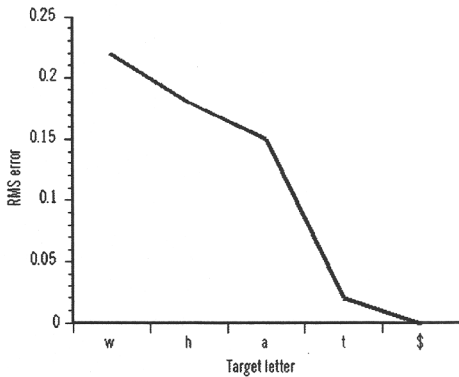


Figure 3. The error between the target and actual output of the network for each letter of the word "what".

Two types of sequential models have been used to investigate syntax. The first type of model attempts to complete the more difficult task of discovering the grammatical type (and function) of each word token (e.g. Elman, 1990). The input to this first type of model is a representation of word tokens and the output is the word token that the network predicts is likely to follow that particular input. A successful network is able to learn that word tokens of particular grammatical types are able to follow tokens of some grammatical types but not others. The disadvantage of this approach is that it uses small lexicons and can only test fragments of grammar. However, models of this kind are able to simulate data collected from human participants on grammatical ratings (Christiansen, 1999;

Christiansen & Chater, 1999), complex grammatical structure (MacDonald & Christiansen, 1999) and sentence comprehension (Tabor, Juliano, & Tanenhaus, 1997). Elman (1990) was also interested in whether it was possible to learn syntactic classes from word order. He trained an SRN to discover syntactic classes from the order in which words appeared in the input. Sentences were generated from a lexicon of 29 items. The sentences were formed using a grammar in which there were subject noun/verb agreements, different verb argument structures (i.e., intransitive, transitive, optionally transitive) and subject object relative clauses (allowing multiple embeddings with complex long-distance dependencies). Sentences were formed by randomly selecting a word that was appropriate for a particular slot in a sentence frame. A localist coding scheme was employed in which each word was represented by an individual code. The coding did not indicate that any word was from the same syntactic category as another word. The network's task was to predict the next word in the sequence. The network was trained on 6 complete repetitions of the training set. The words were input one at a time. For each word there was a limited number of legitimate successors. The network was expected to learn the frequency of occurrence of each of the possible successors. If the network learned these frequencies then words that are likely to occur in similar slots might be expected to be represented in a similar way in the hidden units. A cluster analysis of the way in which the words were represented in the hidden units revealed that the network had discovered that there were several major categories of word types. Two major categories, nouns and verbs, were found. The verb category was further broken down into transitive, intransitive and optionally transitive. The noun category was broken down into animate and inanimate nouns. Thus, from word

order alone the network learned that some verbs need to be followed by a noun but others do not. The network had no semantic representations but the results indicate that an important component of meaning seems to be context (i.e. consistent patterns exist in which certain words frequently co-occur in particular sequences with some words, but not with other words).

The second type of model involves training a network on sentences in which grammatical type is used as the input and the network outputs a prediction of the grammatical type that the next item is likely to come from (e.g. Hanson & Kegl, 1987; Howells, 1988). To test performance, the network is required to assign the appropriate grammatical type to the next word. The advantage of models such as these is that they can be used with large corpora of natural language as the network does not have to remember individual word tokens but rather in which order the various grammatical types in the input must appear to form "grammatical sentences."

In English, the possessive [-s] morpheme is always followed by a second noun but the plural [-s] morpheme is rarely followed by a second noun. Hayes, Murphy, Davey and Smith (2003) developed a model to investigate whether a SRN could learn that in English items from some syntactic categories follow the possessive [-s] morpheme and items from syntactic categories follow the regular plural [-s] morpheme. It was thought that this factor might be implicated in the experimental finding that irregular plurals are included in compounds more frequently than regular plurals (Gordon, 1985). The hypothesis was that the regular plural [-s] morpheme was omitted from a noun when that noun was followed by a second noun (as would be the case in noun-noun compounds) because of competition with the possessive [-s]. Irregular plurals do not end in [-s] and can therefore

be included before a second noun because they do not compete with the possessive. In earlier models, (Hayes et al, 2002) items in the training set were not explicitly coded as being representative of a particular syntactic type (e.g. as being nouns or verbs). Instead, learning about the distinct linguistic functions that the different syntactic types perform emerged during training. However, a disadvantage of these models was that it was only possible to use a small lexicon of words because of the complexity of the learning task. This model was trained on a much larger training set than the earlier models. This simulation sought to reproduce the behaviour of an older child, with a much larger vocabulary, who has knowledge, though perhaps not at a metalinguistic level, of the different functions that are performed by the different syntactic types.

The aim of this simulation was to investigate whether the fact that the possessive [-s] morpheme is always followed by a second noun but the plural [-s] morpheme is rarely followed by a second noun is sufficient to constrain compound formation in English. A simple recurrent network (SRN) was utilised so that at any point in time the state of the hidden units at the previous time step were used as additional input (Elman, 1990). Thus it was expected that the model would be able to learn sequential mappings. The network was trained on a large training set of real child directed speech in which the frequencies with which the various types of morphology occurred were not manipulated in any way. The syntactic type of each word was used as the input to the network. The frequency in which regular and irregular plurals and possessives were included in the training set was determined by the frequency in which they appeared in the child directed speech that was used as the input to the model. Table 1 illustrates that some items appear in sequence with other items in the input (e.g.

possessives are always followed by singular nouns) but other items do not appear in sequence with other items (e.g. regular plurals are not followed by singular nouns). The performance of the network was investigated using a syntactic type prediction task in which one of three syntactic types was input (a possessive, a regular plural or an irregular plural) and the network predicted which syntactic type it expected to see next in the input stream. The ability of the network to learn this task was tested using the same task. The difference (RMS error) between this predicted output and the output for noun, verb, other and word ending was calculated. It was predicted that the RMS error would be high for all items after possessives except nouns. Conversely it was predicted that there would be a high RMS error on predicting a noun after a plural of either kind.

#### *Training set and coding scheme*

The exact composition of the training set is shown in Table 2. Irregular and regular plurals and possessives form less than 1% of the input. Items coded as "others" included anything that was not a noun or a verb (e.g. adjectives, determiners, adverbs and prepositions). 2182 sentences, made up from 9999 words, from the Wells study from the CHILDES corpora (MacWhinney & Snow, 1985) were concatenated and used as input. A sentence ending marker was also included

in the training set. The frequency with which items from various syntactic categories followed irregular plurals, regular plurals and possessives is shown in Table 1.

Possessives were only ever followed by singular nouns in the input. Regular and irregular plurals were followed by a range of items but never by a singular noun. Each item was encoded using a 7 bit vector. Three input units encoded syntactic category (noun, verb, other) and two inputs encoded whether the item was plural or not. Two input units encoded the presence or absence of the [s] morpheme. Thus for both regular plurals and possessives the input units for noun and [s] morpheme present would both be activated. A possessive was disambiguated from a regular plural, however, because the plural input unit was "on" for a plural but "off" for a possessive. Examples of how items from different syntactic categories were encoded is shown in Table 3.

#### *Architecture*

The architecture of the network is shown in Figure 4. The network had 7 input units, 4 hidden units, 7 output units and 4 context units. A simple recurrent architecture was adopted so that at any point in time the state of the hidden units at the previous time step were used as additional input (Elman, 1990). The SRN was trained using a learning rate of 0.1 and a momentum of 0.3.

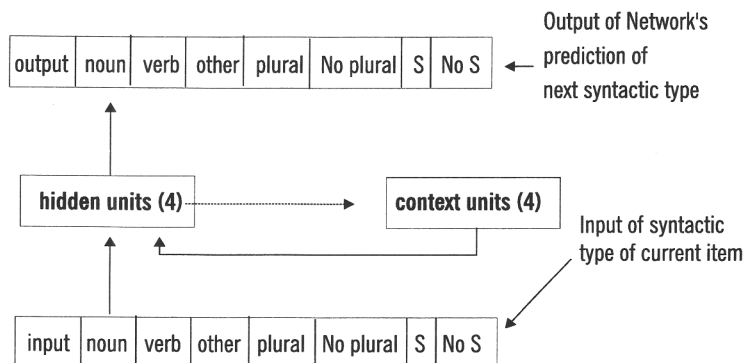


Figure 4. Network architecture

**Task**

In both the training and test phases, the network was required to predict the next input. (i.e. the target output was one time step behind in the input).

**Training**

The network was trained on 10,000 repetitions of the training set. This high number of presentations of the input was necessary because the training set was large and items of particular interest i.e. possessives (0.39%

Table 1. Frequency with which items from various syntactic categories followed irregular plurals, regular plurals and possessives (percentage frequency shown in brackets) in the training set.

Item following plurals(n=95)	plural or possessive Possessives(n=39)	Irregular plurals(n=9)	Regular
Others	3 (33)	40 (42)	0
Sentence ending marker	0	30 (32)	0
Singular nouns	2 (22)	0	39 (100)
Verbs	1 (11)	24 (25)	0
Regular plurals	1 (11)	1 (1)	0
Irregular plurals	2 (22)	0	0

Table 2 Composition of training set.

Item	Number of tokens in training set	Cumulative total	Percentage of tokens in training set	Cumulative percentage
Irregular plurals	9	9	0.09	0.09
Possessives	39	48	0.39	0.48
Regular plurals	95	143	0.95	1.43
verbs	624	767	6	8
Sentence ending markers	1415	2182	14	22
Singular nouns	3014	5196	30	52
others	4803	9999	48	100

Table 3. Examples of coding scheme.

Item	Syntactic category			Type of noun		S present or not	
	noun	verb	other	singular	plural	S	No S
rats	YES	NO	NO	NO	YES	YES	NO
mice	YES	NO	NO	NO	YES	NO	YES
rat's	YES	NO	NO	YES	NO	YES	NO
chaser	YES	NO	NO	YES	NO	NO	YES
the	NO	NO	YES	NO	NO	NO	NO
chase	NO	YES	NO	NO	NO	NO	NO
Sentence ending marker	NO	NO	NO	NO	NO	NO	NO

of the input), regular plurals (0.95% of the input) and irregular plurals (0.09% of the input) formed such a low proportion of the input.

### Test Phase

After training, the network was presented with the following sequences:

- possessive followed by singular noun
- possessive followed by verb
- possessive followed by other
- possessive followed by sentence ending.
- Regular plural followed by singular noun
- regular plural followed by verb
- regular plural followed by other
- regular plural followed by sentence ending.
- Irregular plural followed by singular noun
- irregular plural followed by verb
- irregular plural followed by other
- irregular plural followed by sentence ending.

Thus, for example, in the test pattern *possessive followed by singular noun* the code for *possessive noun* was input and the target output was *singular noun*. However the network might not output *singular noun*. The actual output and the target output were compared and an error figure was calculated based on the difference between the two output weight values.

### Results

The error between the actual output and the target output was recorded after the network was presented with the test sequences. Many runs of the simulation were carried out but each produced almost identical results.

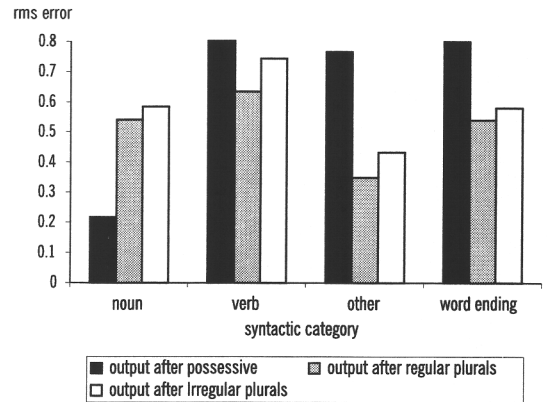


Figure 5. Error on producing nouns, verbs, other items and word endings after possessives, regular plurals and irregular plurals.

Figure 5 illustrates that at a descriptive level the error on producing a singular noun after a possessive was about half as high as the error on producing a singular noun after a plural of either type<sup>1</sup>. The network also learnt that the syntactic categories that make up other items and sentence-ending markers can follow plurals but not possessives. The network produced a high rate of error when the target output after a plural noun was a verb, despite the fact that in the input verbs followed regular plurals (25% of the time that regular plurals occurred) and irregular plurals (11% of the time that irregular plurals occurred). However, the training set contained very few verbs (6.24 % of the training set). Given that verbs were so underrepresented in the input it was unlikely that they would be predicted as the next item in a next word prediction task to any great extent.

### Discussion

This neural network was trained using naturalistic child directed speech. Gaining this ad-

<sup>1</sup> It was not possible to carry out a statistical test on the error rates shown in Figure 5 as the figures shown relate to the output of 1 test rather than to the output of several tests.



vantage, however, meant that the syntactic type of each token rather than individual tokens were used as input to the network. This means that syntactic type did not emerge during training as was the case for the models with smaller lexicons which were reported in Hayes et al (2002). However, this model offers an insight into how learning might take place when the frequencies of items in the input are more accurately represented. The syntactic category prediction task showed that the error on producing a singular noun after a plural, of either kind, was twice as high as the error on producing a singular noun after a possessive. This suggests that the network easily learned the sequence *possessive* [-ʒ] – *noun*. The network also learnt that the syntactic categories that make up other items and sentence-ending markers can follow plurals but not possessives. This learning seems to have occurred because these items appeared in consistent patterns in the input despite having very low frequencies in the messy context of child directed speech. Regular plurals and possessives were disambiguated in the input by the fact that the plural input unit was on in the case of a regular plural but off in the case of a possessive and from the patterns in which they occurred in the input. Thus it seems that the network was able to learn that the *noun* – *morpheme* [-s] pattern occurred in different patterns when it was plural to when it was singular. Some items follow one pattern (i.e. a second noun follows the *noun* [-s] *morpheme* pattern when it is singular but not when it is plural) while other items follow the reverse pattern (i.e. word ending markers and other items follow the pattern *noun* – [-s] *morpheme* when it is plural but not singular). That a neural network model with no explicit grammatical structure was able to learn these linguistic patterns is further support for the idea that there is sufficient evidence in the

input to constrain learning that a second noun is not included after a plural because the pattern *noun* – *morpheme* [-s] – *noun* is used to denote possession not plurality.

## 5. DEVELOPMENTAL MODELS

### *Introduction*

Connectionist models are also able to suggest mechanisms by which language learners might acquire particular linguistic functions by exposure to the linguistic input. The dual mechanism model does not need to propose learning mechanisms because they argue that linguistic constraints are innate. However, if a connectionist model without built in constraints is able to offer an explanation of how learning might proceed then the assumption that this knowledge is innate is weakened.

### *Probabilistic learning of language*

Seidenberg and McDonald (1999) posit a mechanism by which they argue language learners might acquire linguistic constraints (rules). Seidenberg and McDonald built upon the findings of sequential models, where grammatical learning emerges from exposure to language, to develop a probabilistic approach to language acquisition and processing. They posit that knowing a language is not equated with knowing grammatical rules. Instead, they argue that knowledge of language develops as children attempt to speak (production) and understand (comprehension) the speech they hear. To test these ideas, Allan and Seidenberg (1999) developed a connectionist model that was trained on two tasks. The first was to compute the semantics of a series of words (comprehension) the second was to compute a series of words (production) having been given semantic patterns. The network was then presented with a series of test

sentences and was required to identify whether these test sentences were grammatical or not (i.e. did the test sentences conform to the grammar that the network had been exposed to during training). The architecture of the model is shown in Figure 6.



Figure 6. Architecture of Allan & Seidenberg's 1999 model

The training set consisted of 20 examples of 10 types of sentences (i.e. sentences with different grammatical structures) from a vocabulary of 97 words. Each word was represented locally in the network. This meant that every individual word was encoded using a coding that was independent of the coding used for all other items. The semantics of each word were represented as the state of a space made up of 297 units. During training when the network was required to perform the comprehension task the units representing each word in the sentence were activated in sequence. The task of the network was to compute the correct semantic representation of each word in the sequence. On the production task the model was required to produce the correct localist code for each word in the sentence having been given the semantics of the words in that sentence. The network was trained by interleaving form to meaning and meaning to form tasks. Thus, the network was simply trained on exposure to examples and weights became adjusted towards structures to which the network had been exposed and weights became adjusted away from structures that had not been exemplified in the input. During the test phase, words making up a sentence were supplied to the network and the semantics of these words would be activated (via the form to meaning connec-

tions, see Figure 6.). This semantic pattern would then be translated back into words (via the meaning to form connections, see Figure 6.) and if the form of the translated sentences were unlike the patterns of words (sentences) used in the training set then a large error would be produced. However, if the translated sentences were similar in form to sentences in the training set then a lower error would be produced.

Allan and Seidenberg's (1999) model was successful at learning which structures were grammatical (i.e., similar to sentences seen previously) and led Haskell, Macdonald and Seidenberg, (2003) to conclude that for instance the treatment of plurals in compounds in English might be learnt in a similar manner. This stems from the series of experiments that have demonstrated that native English speakers include irregular plurals in compounds more frequently than they include regular plurals (e.g. Gordon, 1985). Haskell et al did not build a network to describe how the treatment of plurals in compounds is learnt but argued that weights would be adjusted towards plurals being omitted from compounds and away from plurals being included in compounds. Thus, plurals would become less likely to be included in compounds but could be in certain circumstances (if that item had been included in the plural form frequently in the input) and particularly early in training (before the weights settled down). The significance of this model is that it counters the argument that children cannot learn to include irregular plurals in compounds (which they did in Gordon's (1985) experiment) from the input they receive because they do not hear adults including plurals of either type in compounds. Haskell et al, applying this probabilistic learning viewpoint to compounding, argue that in the vast majority of instances children will have heard plurals used when plural seman-

tics are required and thus will have developed a language system based on this fact. They will experience far fewer examples of where plurals are omitted (i.e. in compounds) but gradually they will learn this exception to the general way that plurals are treated in English.

*Connectionist model investigating factors affecting the treatment of plurals in compounds*

Although they discuss how a connectionist model of the treatment of plural morphology in compounds using probabilistic constraints might be developed, Haskell et al (2003) do not build such a model. Instead they build a connectionist model to investigate whether the phonological structure of a word indicates whether this item is permissible before a second noun. This is to investigate their hypothesis that due to a phonetic constraint words that sound like regular plurals do not appear before a second noun. They hypothesised that adjectives (i.e. words that may occur before a second noun) have a particular phonetic structure (in particular they tend not to sound like regular plurals) that is not present in words from other syntactic categories. The network consisted of 26 input units that encoded phonetic features. The hidden layer had 20 units and the output layer had one unit, adjective or not. The frequency with which an item was presented to the network was representative of its frequency in the Brown Corpus produced by the Penn Treebank project (University of Pennsylvania, Philadelphia, PA). The training set was presented for 50 iterations. Over 3 test runs the network on average was able to correctly identify 75% of the adjectives it was trained on as being from this syntactic category and 84% of items as being from other categories i.e. not adjectives. On testing with novel input, the network classified 70 % of previously unseen

adjectives correctly and 79% of non-adjectives correctly. Thus, Haskell et al concluded that phonetics play a significant role in learning syntactic categories (Kelly (1992); Morgan (1996)).

## 5. CONCLUSIONS

The challenge for supervised neural net models of morpho-syntax has been to demonstrate that language learning that appears to entail a data base of rules and exceptions can be simulated without the need for these structures to be present. Several connectionist models of morpho-syntax have met this challenge. Both static and sequential models of the acquisition of morpho-syntax have been developed. That static connectionist models have been able to simulate the development of both verbal and nominal morphology has raised many questions of the dual mechanism model of morphological processing (Pinker, 1999). Sequential connectionist models have considerably aided our understanding of many areas of morpho-syntax including recently the treatment of plural nouns in compound words. The models discussed here and many other connectionist models have had a considerable effect on our understanding of language acquisition. If nothing else connectionist models have provided a test-bed for the learnability of linguistic properties previously assumed to be innate (Elman, Bates, Johnson, Karmiloff-Smith, Parisi and Plunkett, (1996), Christiansen and Chater, 1999).

## REFERENCES

- Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Amari, S.-I. (1967). A theory of adaptive pattern classifiers. *IEEE Transactions on Electronic Com-*

- puters, 16, 299–307.
- Brown, R. (1973). *A first language: the early stages*. Cambridge, Ma: Harvard University Press.
- Chandler, S. (1993). Are rules and modules necessary for explaining language? *Journal of Psycholinguistic Research*, 22, 593–606.
- Christiansen, M., Allen, J., & Seidenberg, M.S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221–268.
- Christiansen, M.H., & Chater, N. (1999). Connectionist natural language processing: The state of the art. *Cognitive Science*, 23, 417–437.
- Chomsky, N. (1959). Review of Skinner (1957). *Language*, 35, 26–58.
- Christiansen, M.H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157–205
- Cottrell, G., & Plunkett, K. (1991). Using a recurrent net to learn the past tense. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*. Pittsburgh, PA: Cognitive Science Society Incorporated.
- Creutz, M. and Lagus, K. (2002). Unsupervised discovery of morphemes. *Proceedings of the Workshop on Morphological and Phonological Learning of ACL-02*, 21–30, Philadelphia, Pennsylvania, July 11.
- Daugherty, K. G., & Seidenberg, M. S. (1994). Beyond rules and exceptions. In S. D. Lima, R. L. Corrigan & G. K. Iverson (Eds.), *The reality of linguistic rules*. Amsterdam: John Benjamins.
- Daugherty, K.G., & Hare, M. (1993). What is a rule? The past tense by some other name might be called a connectionist net. In M. Mozer, P Smolensky,
- D. Touretzky, J. Elman & A Weigand (Eds.), *Proceedings of the 1993 connectionist models summer school*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review*, 104, 801–838.
- Dell, G. S., Burger, L. K., and Svec, W. R. (1997), Language Production and Serial Order: A Functional Analysis and a Model, *Psychological Review*, 104, 123–147.
- Dell, G. S., Chang, F., and Griffin, Z. M. (1999), Connectionist Models of Language Production: Lexical Access and Grammatical Encoding, *Cognitive Science*, 23, 517–542.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness*. Cambridge, MA: MIT Press.
- Everitt, B.S. (1993). *Cluster Analysis*. London: Edward Arnold.
- Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: 'A critique'. *Cognition*, 28, 3–71.
- Gordon, P. (1985). Level-ordering in lexical development. *Cognition*, 21, 73–93.
- Grossberg, S. (1969). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks, *Journal of Statistical Physics*, 6, 319–350.
- Hahn, U., & Nakisa, R.C. (2000). German Inflection: Single or Dual Route? *Cognitive Psychology*, 41, 313–360.
- Hanson, S.J., & J. Kegl. (1987). PARSNIP: a connectionist network that learns natural language grammar from exposure to natural language sentences. *Proceedings of the 9th Annual Conference of the Cognitive Science Society*.
- Hare, M., Elman, J.L., & Daugherty, K.G. (1995) Default generalization in connectionist networks. *Language and Cognitive Processes*, 10, 601–630
- Harm, M., & Seidenberg, M.S. (2001). Are There Orthographic Impairments In Phonological Dyslexia? *Cognitive Neuropsychology*, 18, 71–92.
- Harm, M., & Seidenberg, M.S. (1999). Reading acquisition, phonology, and dyslexia: Insights from a connectionist model. *Psychological Review*, 106, 491–528.
- Harm, M., McCandliss, B.D., & Seidenberg, M.S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, 7, 155–182.
- Haskell, T.R., MacDonald, M.C., & Seidenberg, M.S. (2003). I. Language learning and innateness: Some implications of compounds research. *Cognitive Psychology*, 47, 119–163.
- Hayes, J.A., Murphy, V.A., Davey, N., Smith, P.M. & Peters, L. (2002). The /s/ morpheme and the compounding phenomenon in English. In W. D. Gray & C.D. Schunn (Eds.) *Proceedings of the 24<sup>th</sup> Annual Conference of the Cognitive*

- science Society. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall.
- Hecht-Nielsen, R. (1990). *Neurocomputing*, Reading, MA: Addison-Wesley
- Honkela, T. & Vepsäläinen, A.M. (1991). *Interpreting Imprecise Expressions: Experiments with Kohonen's Self-Organizing Maps and Associative Memory*. In T. Kohonen, K. Mäkisara, O. Simula and J. Kangas (eds.). Proceedings of ICANN-91 (Artificial Neural Networks), North-Holland, vol. I.
- Honkela, T., Pulkki V. & Kohonen T. (1995). *Contextual Relations of Words in Grimm Tales Analyzed by Self-Organizing Map*. In F. Fogelman-Soulie' and P. Gallinari (eds.). Proceedings of International Conference on Artificial Neural Networks, ICANN-95., Paris: EC2 et Cie.
- Honkela, T. (1997). *Learning to Understand – General Aspects of Using Self-Organizing Maps in Natural Language Processing*. In D. Dubois (ed.). Computing Anticipatory Systems. Woodbury, NY: American Institute of Physics.
- Honkela, T. (2000). *Self-Organizing Maps in Symbol Processing*. In S. Wermter & Ron Sun (eds.). Hybrid Neural Systems. Heidelberg:Springer.
- Howells, T. (1988). VITAL - A connectionist parser. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*.
- Jacoby, L. L. (1983). Remembering the data, Analysing interactive processes in reading. *Journal of Verbal Learning and verbal Behavior*. 22, 485–508.
- Joanisse, M.F., & Seidenberg, M.S. (1998). Specific Language Impairment in children: An impairment in grammar or processing?. *Trends in Cognitive Science*. 22, 40–246.
- Jordan, M. (1986). *Serial order: A parallel distributed approach* (Tech. Rep. No. 8604). San Diego, CA: University of California, San Diego, Institute for Cognitive Science.
- Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*. 99, 349–364.
- Kohonen, T. (1989). *Self-Organization and Associative Memory*. Berlin: Springer-Verlag.
- Kohonen, T. (1972). Correlation matrix memories. *IEEE Transactions on Computers*, 21, 353–359.
- McClelland, J.L., and Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- MacDonald, M.C., & Christiansen, M.H. (in press). Reassessing working memory: A comment on Just & Carpenter and Waters & Caplan. *Manuscript accepted by Psychological Review*.
- MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualisations: revising the verb learning model. *Cognition*, 40, 121–157.
- Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language* 21, 339–366.
- Marchman, V., Plunkett, K., & Goodman, J. (1997). Overregularization in English plural and past tense inflectional morphology. *Journal of Child Language*. 24, 767–779.
- Marcus, G. F. (1995a). Children's overregularisation of English Plurals: a quantitative analysis. *Journal of Child Language*, 22, 447–459.
- Marcus, G. F., Brinkmann, U., Clahsen, H., Weise, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29, 189–256.
- Oja, E. (1982). A Simplified Neuron Model as a Principal Component Analyzer. *Journal of Mathematical Biology*, 15, 267–273.
- Pinker, S., & Prince, A. (1988). On language and connectionism: An analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 73–193.
- Pinker, S. (1991). Rules of language. *Science*, 253, 530–535.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: William Morrow and Company.
- Pinker, S. (1999). *Words and Rules*. London: Weidenfeld & Nicholson.
- Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, 23, 463–490.
- Plunkett, K., & Marchman V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 43–102.
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48, 21–69.

- Plunkett, K., & Marchman, V. (1996). Learning from a connectionist model of the English past tense. *Cognition*, 61, 299–308.
- Plunkett, K., & Nakisa, R. C. (1997). A connectionist model of the Arabic plural system. *Language and Cognitive Processes*, 12, 807–836.
- Prasada, S., Pinker, S., & Snyder, W. (1990). Some evidence that irregular forms are retrieved from memory but regular forms are rule generated. *Paper presented at the 31<sup>st</sup> Annual meeting of the Psychonomic Society, New Orleans.*
- Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
- Saffran, J. R. (2001). Words in a sea of sounds: The output of statistical learning. *Cognition*, 81, 149–169
- Saffran, J.R., Aslin, R.N., & Newport, E.L., (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926–1928.
- Scholtes, J. C. (1992a). Neural data oriented parsing. In *Proceedings of 3rd Twente Workshop on Language Technology*, Twente, Netherlands.
- Scholtes, J. C. (1993). *Neural Networks in Natural Language Processing and Information Retrieval*. PhD thesis, Universiteit van Amsterdam: Amsterdam, Netherlands.
- Seidenberg, M.S. (1999). Visual word recognition. *MIT Encyclopedia of Cognitive Science*. Cambridge, MA: MIT Press.
- Seidenberg, M.S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275, 1599–1604
- Seidenberg, M.S., & Elman, J.L. (1999). Do infants learn grammar with algebra or statistics? *Science*, 284, 433.
- Seidenberg, M.S., & MacDonald, M.C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569–588.
- Sejnowski, T. J. and Rosenberg, C. R. (1986). *NETalk: A parallel network that learns to read aloud*. Technical Report JHU/EECS–86/1, John Hopkins University Department of Electrical Engineering and Computer Science.
- Tabor, W., Juliano, C., & Tanenhaus, M. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211–271.

## KIELEN OMAKSUMINEN KONNEKTIONISTIEN MUKAAN: MORFOLOGIAN PROSESSOINNIN MALLIT

Jenny Hayes, Department of Psychology, University of Hertfordshire, U.K.

Kielen omaksuminen vaikuttaa perustuvan sääntöjen ja poikkeustapausten oppimiseen. Morfosyntaksin prosessointia mallintavia, ohjattuun oppimiseen perustuvien neuroverkkomallien haaste on ollut osoittaa, että kielen omaksumista voidaan simuloida ilman tällaisia rakenteita. Tässä katsausartikkelissa tarkastellaan morfosyntaksin konnektionistisia malleja, joiden avulla on pyritty vastaamaan tähän haasteeseen. Artikkelissa kuvataan aluksi konnektionististen mallien toimintaperiaatteita ja sen jälkeen selitetään kuinka staattiset ja sekventiaaliset kielen omaksumisen malleja on kehitetty. Mallien kykyä simuloida sekä verbien että substantiivien omaksumista tarkastellaan suhteessa siihen, miten nämä tulokset kyseenalaistavat morfologisen prosessoinnin ”kaksoismekanismi”-mallit (Pinker, 1999). Sekventiaalisten konnektionististen mallien osuutta yhdyssanoissa esiintyvien monikollisten substantiivien käsittelyn selittämisessä tarkastellaan yksityiskohtaisesti. Lopuksi esitetään katsaus morfosyntaksin oppimisen kehityksellisiin malleihin.

**Avainsanat:** konnektionistinen mallintaminen, morfosyntaksi, kielen omaksuminen, ohjattuun oppimiseen perustuvat neuroverkkomallit