Connectionism as a Model for Second Language Acquisition Research

Takahashi, Satomi
Institute of Languages and Cultures, Kyushu University

https://doi.org/10.15017/6796468

出版情報：言語科学. 32, pp.19-38, 1997-02-26. The Institute of Languages and Cultures, Kyushu University
バージョン：

Connectionism as a Model for Second Language Acquisition Research

Satomi Takahashi

Introduction

For the past 15 years, a new cognitive paradigm, which is variously known as *connectionism*, *parallel distributed processing*, or *neural networks*, has been influencing the area of cognitive science and signaling a new Kuhnian revolution (see Schneider, 1987). Let's call this new paradigm *connectionism* in this paper as it seems most generic. Precisely speaking, however, the term *new* to describe connectionism may not be appropriate because this paradigm originated mid-century. Connectionism was initially proposed as a model which describes the nature of human cognition using subsymbolic and neuron-like units. However, it became unfashionable in 1960s as the symbolic and rule-based tradition gradually dominated the small area of artificial intelligence and tried to replace behaviorism within psychology. This does not mean, however, that connectionism is a variant on behaviorism as some quarters of cognitive scientists claim (see Lachter & Bever, 1988). Although connectionism and behaviorism share some features as the paradigms emerging in the empiricist framework, they show essentially different orientations in their research goals. Connectionism emphasizes its interest in what is going on *inside* the brain, i.e., "the internal representations that are constructed between the inputs from and the outputs to the environment" (Gasser, 1990, p. 183). Behaviorism, on the other hand, stresses the observable phenomena *outside* the brain. Because of its *inside* orientation, connectionism could survive (in the area of neurobiology) even after behaviorism was replaced entirely by the symbolic paradigm representing a rationalist view of human cognition in 1970s.

As suggested in the term *symbolic,* the central tenet of the symbolic paradigm is the manipulation of symbols or rules. According to this paradigm, rules are specified "how symbols could be composed (syntax) and how they could be transformed" (Bechtel & Abrahamsen, 1991, p. 1) and govern entire human cognitive performance. The symbolic approach fully dominated the interdisciplinary fields of cognitive science, such as cognitive/developmental psychology, artificial intelligence, linguistics, anthropology, philosophy, and neuroscience, by the end of 1970s. At the same time, however, some quarters of cognitive scientists started to realize limitations of the symbolic paradigm. They argued that a serial system composed solely of symbolic representations and rules

Linguistic Science, No.32 (1997), pp.19-38

operating on them is more likely to be "brittle," searching for an alternative to this approach. Two alternatives were proposed. One was still a symbolic approach but more flexible than the earlier system, and the other was the connectionist approach which had been surviving during the era of the traditional symbolic approach. Most cognitive scientists chose the first alternative (for more details, see Bechtel & Abrahamsen, 1991). In 1980s, however, the cognitive scientists who selected the second alternative made a tremendous amount of effort to refine the original connectionist architectures. They combined the traditional neural model techniques with those from artificial intelligence and psychology (i.e., probabilistic feature models of categorization, semantic networks with spreading activation, and schema theory). They thereby established models stressing distributed representation and statistical explanation (see Rumelhart, McClelland, & the PDP Research Group, 1986; McClelland, Rumelhart, & the PDP Research Group, 1986; Bechtel & Abrahamsen, 1991). These models have been claimed to provide a powerful theoretical base to explicate the nature of human mental representation in a more successful and satisfactory manner than the traditional/nontraditional symbolic approach. In this sense, the connectionist models developed in 1980s can constitute a *new* paradigm and have a potential for causing a Kuhnian revolution as mentioned above.

As one of the subfields of cognitive science, second language acquisition (SLA) research has also been influenced by this new paradigm. Some of the researchers who are not satisfied with the Chomskyan symbolic tradition have been undertaking studies in this connectionist framework. They have so far provided some results which *seem* to be more persuasive and reasonable in explaining the nature of SLA than the proponents of the symbolic tradition.

In this paper, I will attempt to explore how the connectionist models have been and will be treated in the area of SLA research as well as language acquisition research, in general. In so doing, first, the connectionist architectures will briefly be examined to grasp the nature of the models. In the following section, I will present Rumelhart and McClelland's (1986) experiment undertaken in the connectionist framework. They specifically focused on acquisition of the past tense of verbs in L1 English. Furthermore, in an effort to grasp the extent to which language acquisition can be dealt with in this paradigm, I will discuss the critiques made by Pinker and Prince (1988) against Rumelhart and McClelland's experiment. The assumption here is that such critiques may also be applied to SLA research conducted in this new paradigm. Then, in the subsequent section, I will review some of the SLA research undertaken in this connectionist framework. This section will further be followed by the concluding comments on the future directions of this new research area.

Connectionist Architectures

The fundamental idea of the connectionist models is that there is a *network* which consists of neuron-like units. In most connectionist models, the basic constituents of these neuron-like units are *input units* and *output units* which have some degrees of *activation*, i.e., values indicating the unit's current level of activity. Specifically, input units receive activation from the external environment or from another part of the network. On the other hand, output units deliver the network's response to an input pattern, i.e., the pattern of activation across the n input units which is traditionally represented by a vector in n-dimensional space. Units in a network are connected to each other. Activated units *excite* or *inhibit* other units; or, to put it another way, *excitations* and *inhibitions* are spread among the units in a particular network. The crucial point here is the following: The initial activation supplied to input units should be viewed as specifying a problem, and the stability attained in the connected output units at the end of processing should be interpreted as the system's solution to the problem. Within the framework of connectionist models, then, *learning* takes place in the form of internalizing the particular patterns of activation with particular values of *weights*. These established weights are stored in memory and are evoked later for the knowledge representation. Accordingly, no symbol manipulations, such as the rule-governed transformations of symbols, are involved in the operations of connectionist models. Below I will present four types of connectionist architectures as the representative models in this newly emerging paradigm.

Two-Layer Perceptron

There are two types of networks in the connectionist models in terms of the directionality of activation propagation: feedforward networks and interactive networks. The two-layer perceptron is one of the feedforward networks and the best-known variety of the basic architecture. It consists of two layers of units, input and output units. Those two units are directly connected to each other; and once the weights are properly set, this type of a network can "respond to each of a variety of input *patterns* with its own distinctive output *pattern*" (Bechtel & Abrahamsen, 1991, p. 312, italics mine). Because of the nature of direct connection between input and output patterns, the two-layer perceptron is sometimes called pattern associator.

This perceptron network was originally proposed and developed by Rosenblatt (1962) in his pursuit of problems related to pattern recognition in networks. According to Rosenblatt, the input and output units take binary (i.e., *on* or *off*) activations; and problem solving is accomplished with linearly separable functions. To be more specific, if its net input (i.e., the sum of all of the inputs to a unit *u*) exceeds a threshold (usually zero), the activation is set to 1; otherwise its activation is set to zero.

Resenblatt later modified his perceptron network in a way that the weights between units can be continuous rather than binary. As a result, the networks can change their responses, if necessary, in the course of training. Specifically, if the weights on connections feeding into a particular unit is judged to be an incorrect response, the initial weights are to be changed. In the framework of the two-layer perceptron network, a finite number of repetitions of the training procedures (which are usually called *epochs*) eventually enable the network to learn to respond correctly.

There is one serious problem with this network, which was pointed out by Minsky and Papert (1969). According to Minsky and Papert, the two-layer perceptron cannot work out when neither of the following two constraints is met: (1) linear independence of the input patterns; and (2) linear separability of the input-output assignments. The representative case which violates the above constraints is *exclusive OR* (XOR) ("A XOR B" is stated as true if A is true and B is not, or B is true and A is not). Minsky and Papert argue that the two-layer perceptron cannot deal with the case of exclusive OR though it can successfully manipulate the *logical AND* and *logical OR* (see Figure 1). One solution for this problem is to include an additional layer of units, which are referred to as *hidden units*, between the input and output units. As a result of this additive operation, multi-layer networks are realized, in which the units in all layers except the input and output layers are the hidden units. The back-propagation network presented next is one type of such networks.

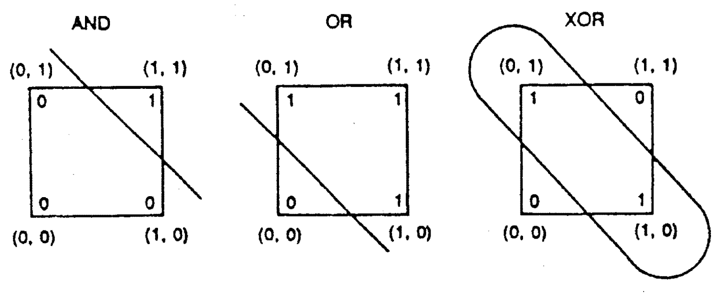

Figure 1. Geometric representations of logical AND, logical OR, and exclusive OR. (taken from Plunkett, 1988, p. 6)

Back-Propagation Network

The back-propagation network, which is also known as the Generalized Delta-Rule, is a learning algorithm proposed by Rumelhart, Hinton, and Williams (1986a, 1986b). As referred to above, this network is one of the multi-layer networks and contains hidden units. The hidden units do not have access to the environment and cannot be controlled by a network designer. To use Gasser's (1990) term, "they develop their significance as the network learns to associate inputs and outputs" (p. 182). For this multi-layer network, a generalization of the delta rule is applied when the network engages in learning. The delta rule is a learning rule that "utilizes the discrepancy between the desired and actual output of each output unit to change the weights feeding into it" (Bechtel & Abrahamsen, 1991, p. 306). To put it another way, any errors identified as a result of the discrepancy between the expected and observed output of each output unit are propagated back through the network layer by layer. Then, at each layer, weight adjustment is carried out using the following equation: $\Delta\ weight_{jk} = lrate\, delta_j\, a_k$.

Though the back-propagation network is more powerful than the above two-layer perceptron network in the training/learning phase of the simulation, there are some drawbacks. The first weakness is that learning speed with this network is extremely slow (Bechtel & Abrahamsen, 1991). The second weakness is concerned with the problem inherent in a network with a back-propagation feature: There is no evidence that back-propagation is occurring in any known biological processes (ibid.). That is, it cannot conclusively be claimed that information is passed backwards through the nervous system as seen in the back-propagation of the connectionist network.

Boltzmann Machines

The Boltzmann Machine was proposed by Hinton and Sejnowski (1983, 1986) and is also known as the Constraint Satisfaction Network. Unlike the above two feedforward networks, this connectionist model is one of the interactive networks, in which units are bidirectionally connected to one another and activations change dynamically across a large number of cycles. Because of its interactive nature without relying on a layered organization, the Boltzmann Machine is also called a *non-layered network.* The basic assumption underlying the Boltzmann procedure is similar to that of the delta rule and generalized delta rule. Namely, changes in weights are induced by discrepancies between desired and actual outputs. Bechtel and Abrahamsen (1991) summarize the way of changing weights in the Boltzmann architecture as follows:

"If two connected units are jointly active more frequently when the network is running with the desired output clamped than when it is running free, the joint activity needs to be increased by increasing the weight of the connection. Conversely, if they are jointly active more frequently when running free than when the desired output is clamped, the joint activity needs to be decreased by decreasing the weight on that connection" (p. 99).

In short, once an input pattern is presented, this evokes multiple cycles of processing; and throughout these cycles of processing, the activations of units interact with each other dynamically, eventually reaching a stable condition. According to Hinton and Sejnowski (1986), this stable condition could be described in the following way: A temperature parameter (as used in physics to find a very low energy state of a metal) is typically lowered through the *simulated annealing* procedure taken by analogy with a physical cooling schedules adopted to avoid the formation of crystals. It should be noted here that the Harmony Theory proposed and developed by Smolensky (1986) claims an interactive network similar to the Boltzmann Machine.

As is always the case with the connectionist architectures, there is a problem with this network as well. The problem here is the network's tendency to land in local minima, i.e., the states which are stable but are not the best condition for the network's ultimate solution. In order to avoid this tendency, Hinton and Sejnowski (1986) argue that the network should have a system which *slowly decreases* the temperature parameter.

Recurrent Network

The recurrent network is another variation of the feedforward (multi-layer) network architecture, which was proposed by Elman (1988) (see also Plunkett, 1988). This network is presented here because of its unique nature that cannot be found in other connectionist models. In this network, the information already processed can be used to influence the processing of the current input. Specifically, after a given element is processed, the network copies the pattern of this particular element in the hidden units onto the special units. While the network is processing the second element, the hidden units for this new element receive input both from their regular input units and from the special units on which the information of the first element was copied. In this way, inputs are sequentially received; and the network can change its response depending on what information was received at previous stages in the sequence. This architecture is very innovative in that it can "develop a capacity to predict or control future events" (Plunkett, 1988, p. 9), which is essential for human information processing.

Connectionist Simulation and Reactions from the Symbolic Tradition

In this section, I will review Rumelhart and McClelland (1986) on the learning of past tense of verbs to grasp the extent to which language acquisition is explainable in the connectionist framework. This review will further be followed by the examination of some reactions from the symbolic tradition against Rumelhart and McClelland's study.

Learning the Past Tense of Verbs

Referring to the three stages of the acquisition of past tense forms by English-speaking children claimed by Brown (1973), Rumelhart and McClelland (1986) attempted to demonstrate that their connectionist simulation can also produce the similar sequence of the acquisition of those past tense forms. Those three stages are: Stage One, in which children learn the past tense of a few specific verbs (e.g., *came, got, went*); Stage Two, in which children learn a general rule for the regular past tense but very often overgeneralize this rule, incorrectly applying it to irregular verbs which were previously produced in their correct forms (e.g., *comed, camed*); and Stage Three, in which children generally produce correct forms for both regular and irregular verbs and try to learn the exceptions to the rule. Those three stages are claimed to manifest the U-shaped developmental sequence for acquisition of irregular verbs by (English-speaking) children.

In their simulation, Rumelhart and McClelland employed a two-layer pattern association network in which units take binary propagated activation (see the previous section). Specifically, the context-sensitive phonological features called *Wickelfeatures* were adopted to represent a verb stem at the input units. The past tense form of the same verb was represented at the output units, again in terms of the Wickelfeatures (see Figure 2). The basic assumption of the Wickelfeatures encoding/decoding in their simulation was the following: Such procedures can capture the phonological similarity among verb stems for which a similar past tense form is required, enabling the network to make a generalization about the relationship between a verb stem and its past tense form. Then, they provided the network with a series of training by, first, supplying high-frequency verbs (regular and irregular) and then medium- and low-frequency verbs (regular and irregular) during the later stage of training.

The following findings were yielded: (1) the high-frequency regular verbs were hardly affected throughout the simulation; and (2) the irregular verbs overall followed the three-stage acquisition sequence, as hypothesized. More specifically about the second finding, the network first generalized the proper past tense forms from the stems for both

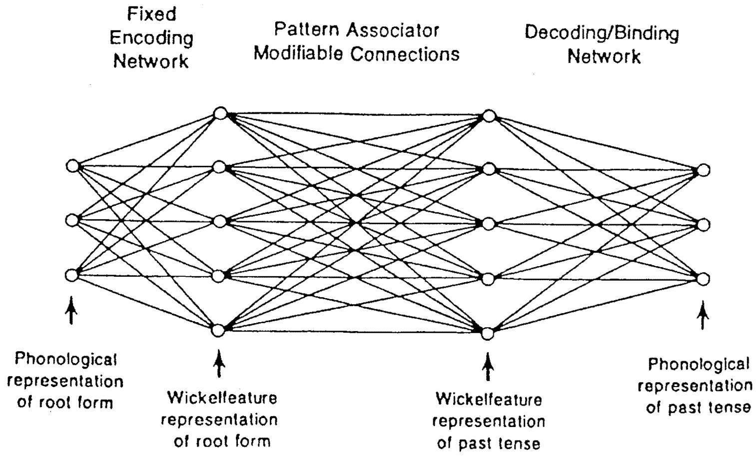

Figure 2. Basic structure of the model of Rumelhart and McClelland (1986) (taken from Rumelhart & McClelland, 1986, p. 222)

regular and irregular verbs successfully. This is equivalent to the learning phase of Stage One above. The network then started to show the poorer performance on the irregular verbs. Rumelhart and McClelland equated this learning phase to Stage Two above. However, the irregular verbs began to improve again and gradually increased accuracy in their past-tense formation at the later stage. According to Rumelhart and McClelland, those irregular verbs were relearned as *exceptional*, as in Stage Three of the acquisition of past tense forms of irregular verbs by children.

Reactions from the Symbolic Tradition

Opponents of the connectionist models contend that the symbolic approach can successfully explain all the linguistic phenomena and that the connectionist approach cannot do so. In fact, they cast doubt upon the validity of Rumelhart and McClelland's (1986) simulation and argue for a superior role played by the symbolic approach in explicating human linguistic capabilities. Among them are Pinker and Prince (1988), Fodor and Pylyshyn (1988), and Lachter and Bever (1988). In what follows, I will present the five major critiques made by Pinker and Prince (1988) on Rumelhart and McClelland's study. Their critiques ought to be considered as those representing the major criticisms by the symbolic tradition against the connectionist models.

The first major critique made by Pinker and Prince is that the simulation conducted by Rumelhart and McClelland is not an accurate model of children's stages of regularization of the past tense morphemes: "there is no question that it is a valuable demonstration of some of the surprising things that PDP models are capable of, but our concern is whether it is an accurate model of children" (p. 81). In fact, there is no evidence that the input to the network approximates to the input to children's past-tense learning mechanism (Bechtel & Abrahamsen, 1991). Rumelhart and McClelland admitted that their procedure departs from the actual process of acquisition of past tense verbs by children to some extent. Specifically, they acknowledged that a verb stem is paired with its past tense form during training, whereas a child hears just one form of a verb at a time. However, the critique of Pinker and Prince above can also be directed to the studies done in the symbolic paradigm: Have researchers from the symbolic tradition successfully been presenting an *accurate* account of child language acquisition, on the whole? Proponents of the symbolic approach, including Pinker and Prince, must be aware that what they are doing with the manipulation of human-made abstract symbols or rules does not provide an accurate model of language acquisition, either. There is no evidence that the knowledge representation explained with such abstract symbols approximates to that made by children (or humans, in general) both psychologically and physiologically (see Sokolik, 1990; Slobin, 1971). Pinker and Prince made a good point; but generally, a connectionist simulation is very exploratory and probabilistic in its nature in an effort to obtain a condition similar to that manifested in actual data elicited from human subjects. In this sense, we may still be able to argue that connectionist simulations, including that of Rumelhart and McClelland (1986), are in the right direction of research into human mind.

In contrast, the second major critique made by Pinker and Prince is very persuasive. They argued that the formation of the proper past tense requires semantic, syntactic, and morphological information as well as phonological knowledge represented by the Wickelphones and Wickelfeatures. Specifically, the three forms of the regular past tense can be determined successfully based on phonological principles represented by the Wicklephones and Wicklefeatures; and, therefore, the prediction of forms by the model is relatively easy and straightforward. As for the irregular past tense, however, their forms are varied and, at best, can be predicted based on the family resemblance structures shared by particular groups of irregular verbs. For the irregular past tense, then, phonological principles alone may not be sufficient for the model to successfully simulate the past-tense formation on the whole. As Pinker and Prince claim, if the principles of semantic, syntactic, and morphological levels are available and incorporated into the connectionist simulation, a more valid and reliable simulation may be realized.

In order to incorporate semantic, syntactic, and morphological principles into a connnectionist simulation, however, a more elaborate architecture than the one used by Rumelhart and McClelland should be established. The necessity of creating such a refined architecture constitutes the third major critique by Pinker and Prince. According to them, the architecture employed by Rumelhart and McClelland (i.e., a two-layer network) is too simple to realize complicated human cognitive processes. As a matter of fact, a two-layer (perceptron) network is the simplest kind of architecture among the four networks presented in the previous section. In view of this, another simulation of learning the past tense of verbs using a more advanced and elaborate network, such as the back-propagation network or the recurrent network, should be undertaken to see whether similar findings are obtained.

Pinker and Prince' fourth major critique is concerned with the three stages for learning past tense forms. Rumelhart and McClelland contended that their simulation successfully yielded the U-shaped acquisition function, in which correct past tense forms during Stage One were often replaced by overgeneralized forms during Stage Two. Pinker and Prince argue that the obtained result was attributable to the nature of the input, rather than any intrinsic characteristic of learning in connectionist networks. According to them, the inputs, i.e., stems of regular and irregular verbs, were discontinuous between Stages One and Two. In fact, we can recall here the following: Rumelhart and McClelland first supplied high-frequency regular and irregular verbs at the outset of training; and later (at the beginning of Stage Two), they *abruptly* added medium-frequency regular and irregular verbs to the original training set of the high-frequency verbs. Based on this fact, we may be persuaded that the discontinuous supply of inputs ultimately created the U-shaped acquisition curve in the connectionist simulation performed by Rumelhart and McClelland. It should be noted here that Rumelhart and McClelland themselves acknowledged the artificiality in supplying inputs in their simulation. Nevertheless, they tried to justify their procedure by arguing that "the actual transition in a child's vocabulary of verbs would appear quite abrupt on a time-scale of years so that our assumptions about abruptness of onset may not be too far off the mark" (1986, p. 241).

As the fifth major critique, Pinker and Prince claim that the performance of connectionist models must inherently depend on ways of encoding inputs which are borrowed from other theories, usually symbolic theories. As a matter of fact, the Wickelfeatures are obtained as a result of linguistic (i.e., symbolic) analyses. As the usual connectionist response, Bechtel and Abrahamsen (1991) provide the following: "much processing remains to be done once an encoding scheme is decided upon, and the connectionist contribution is in offering a nonrule-based means of accounting for this processing" (p. 183). By responding this way, proponents of the connectionist approach appear to succeed in checking the above criticism. But does this mean that connectionist networks cannot satisfactorily handle ways of encoding inputs within their own territories? We may have to wait for more elaborate connectionist architectures of the next generation to provide relevant answers to those questions.

Connectionism and Second Language Acquisition Research

The limitations of the symbolic paradigm in language acquisition research are apparent in view of its inability to fully explicate irregularities (exceptions) manifested in human verbal behaviors observed in a real context of language use (Ney and Pearson, 1990). In fact, Chomskyan linguists supporting the Government and Binding Theory have been making their efforts to revise the rules whenever they encounter linguistic phenomena which cannot be accounted for with their previously established rules. A good example here is the Binding Theory. For this abstract theory, the notion of a governing category plays an essential role in determining whether a given structure follows the binding principles (i.e., Binding Principles A (for anaphors) and B (for pronominals)). However, once the governing category previously defined was found not to explicate the grammaticality of certain types of structures, the linguists continuously modified that definition; and there is still a possibility of an existence of a language which is not explicable with its latest definition (see Chomsky, 1981, 1986). Besides, there exist the parameterized definitions of a governing category; namely, the definitions of a governing category differ from one language to another (see Manzini & Wexler, 1987; see also Takahashi, 1991). The more crucial problem here with the Chomskyan approach is their belief that the abstract rules they created represent human mental operations working in language acquisition and processing. As Sokolik (1990) claims (and as pointed out earlier), "at the behavioral level, there is no direct evidence that these types of symbols are indeed manipulated by rule-like operations in the brains of language learners" (p. 686).

In spite of the limitations and problems of the Chomskyan linguistics as described above, we cannot deny that SLA research has greatly been influenced by the Chomskyan symbolic approach. As a matter of fact, research into the role of Universal Grammar in SLA has revolved around the following Chomskyan claims: (1) learners acquires a system of highly abstract rules; and (2) language acquisition should be considered as the process of inducing such rules (Schmidt, 1988). While UG research in SLA provides us with the significant findings as to the differences and similarities between first and second language acquisition processes, they constitute only a small part of the findings contributing to SLA research. In fact, there are many other issues which should duly be addressed in the area of SLA. Those include: individual differences or variations arising out of different learning contexts, different cognitive styles, different affective factors, different aptitudes, and so forth (see Skehan, 1989; Tarone, 1988; and others). It is then obvious that those issues are explorable only in paradigms which stress "variability, flexibility, and subtlety in its behavior which was not being adequately captured in traditional rule-based systems" (Bechtel & Abrahamsen, 1991, p. 227). Non-UG SLA researchers are thus now realizing that the connectionist approach is among such paradigms which can help them pursue their research goals.

Several SLA studies have been undertaken within the connectionist framework during the recent years. Sokolik and Smith (1989, 1992) examined whether the information inherently manifested in the structure of individual French nouns can induce the correct gender assignment without recourse to other types of information. They utilized the pattern associator network (i.e., a two-layer network) and found out that the gender assignment was carried out by relying solely on information inherent in the structure of the nouns themselves without referring to explicit rules per se (cf. Carroll, 1995).

Sokolik (1990) further attempted a study using the pattern associator network. In her 1990 study, she assessed the Adult Language Learning Paradox (i.e., the disparity between child and adult second language learning rates). Sokolik first linked the size of the learning rate variable in humans to the availability of some nerve growth factor (NGF). Then, a hypothetical learning situation was set out, in which a three-year-old English-speaking child and 25-year-old English-speaking parent in France will learn Feature X_f of French. The results indicated the remarkable difference in learning that particular feature of French between the child and adult: The child learned the target feature much faster than the adult.

Some connectionist simulations were intended to explore the precise effect of input frequency and the L1-L2 interplay in processes of SLA. Among them, Blackwell and Broeder (1992, cited in Broeder & Plunkett, 1994) examined the extent to which their connectionist architecture could simulate the acquisition of L2 Dutch pronouns by Arabic and Turkish speakers. Blackwell and Broeder trained a three-layered feedforward, back-propagation network. The network first learned the L1 Turkish/Arabic pnonominal form-function systems; and then the data for relative frequency of the different pronominal form-functions in Dutch were included. Their simulation indicated that adding new pronominal distinctions (= the Arabic nets learning Dutch) required less language exposure than compressing pronominal distinctions (= the Turkish nets learning Dutch). Furthermore, the order of acquisition of L2 Dutch pronouns in the neural network corresponded with the order found in the L2 Dutch data elicited from Turkish and Arabic speakers.

Frequency effects were also investigated by Ellis and Schmidt (in press) in their study of acquisition of L2 regular/irregular plural affixes. Unlike the previous research, they employed the artificial language (e.g., "garth" for the concept of "car"). Ellis and Schmidt also used a standard three-layered, back-propagation network. Their architecture first learned 20 singular stems, which then served as the starting point for the training on plurals. Half of the 20 experimental items had a regular plural marker (bu-) and the remaining items contained idiosyncratic plural markers. Furthermore, in this phase of training, frequency was crossed with regularity, with half of each set (regular/irregular) being presented five times more often. Ellis and Schmidt found: (1) the regular/high frequency items were learned faster than the irregular/low frequency items; (2) there was much less regularity effect for high frequency items than for low frequency items; (3) the frequency effect was larger for the irregular items; and (4) the larger frequency effect for irregular items was maximal in the mid stage of learning (i.e., a lesser effect at early and later stages). The similar findings were obtained when the same experimental design was applied to human subjects by means of the reaction-time measures. One could argue here that the simulation here is too artificial in that: (1) human learners hear *both* singular and plural forms at each stage of actual development (as admitted by Ellis & Schmidt); (2) phonological factors operating on the morpheme acquisition are completely disregarded, which cannot be observed in real language learning processes; and (3) the target language is an "artificial language," which may not sufficiently be equivalent to a natural language. In spite of those criticisms, however, we cannot deny the fact that their "connectionist model, as an implementation of associative learning provided with the same language evidence, can simulate the human acquisition data" (Ellis & Schmidt, in press, p. 18).

In contrast to the above studies which mostly investigated the units at the level of morphology, Gasser (1990) focused on a connectionist simulation at the level of syntax. Gasser employed the back-propagation network (auto-associator) on a simple English sentence consisting of a subject and a verb (e.g., "John sings") and examined the transfer from L1 to L2. He specifically focused on the relationship between the word order and words. He trained simple L1 forms, first, and then introduced L2 forms which were either quite similar or quite different from the L1 forms. Gasser's connectionist simulation yielded the following findings: (1) the L1 patterns clearly suffered interference from the L2 pattern; (2) the network had difficulty learning L2 patterns when the word order differed from the L1 patterns than when it was the same; (3) as more and more L2 patterns were learned, the word order difference had less and less effects on mastery of the L2 patterns (and interference with the L1 pattern); and (4) the word order difference was more significant when there was a relationship between the L1 and L2 words. This fourth finding suggested that a connectionist experiment may be able to handle the issue of perceived language distance influencing transfer claimed by Kellerman (1978) (see also Broeder & Plunkett, 1994, for similar observation).

A connectionist simulation at the level of sociolinguistics was done by Henning (1989). The primary goal of Henning's study was to simulate the variable behavior that occurs in the American address forms of an academic community by using a (hetero-associative) back-propagation network. She first collected data of address forms in a questionnaire format for the purpose of creating the target representation of a particular item at the output layer. For the numerical representation of an input for a particular item, Henning sequentially strung together 13 bits from the seven categories (age, status, sex, situation, marital status, name known, friend) to form the input vector for that item. On the whole, the simulation produced Henning's hypothesized performance; namely, the model succeeded in representing the variable system of address terms without recourse to rule-like flowcharts. Because of its nature of a pilot study, most of Henning's findings seem to be tentative; however, her application of a connectionist simulation beyond a sentence level should be noteworthy.

As Hatch, Shirai, and Fantuzzi (1990) claim, the outcome of the above SLA research undertaken in the connectionist framework clearly indicates the following two points: (1) the computer can acquire or learn small parts of the language, such as elements of phonology, morphology, and syntax; and (2) it can produce errors similar to those of human learners while learning. At the same time, however, we must admit that all the simulations presented above did not systematically take into account psychological and sociological factors which might affect learning rates and outcome (Sokolik, 1990). Related to this issue, Gasser (1990) further claims that "factors involved in interaction between language users are even more remote from current models, since they would seem to require a relatively complete characterization of the separate cognitive systems" (p. 190). Taken together, the model's incomplete simulation of human cognitive processes is attributable to the current level of connectionist architectures; namely, they are too simple to realize complicated human mental operations requiring the involvement of multi-factors. Recall here Pinker and Prince' (1988) second and third critiques against Rumelhart and McClelland (1986); they also pointed out the similar drawback of connectionist models. All these suggest that the development of more advanced and refined connectionist architectures should seriously be considered.

We should now note that the connectionist simulations done by the SLA researchers above well represent the SLA in a natural setting (or L2 implicit learning), rather than in a classroom setting. The point here is how the connectionist models can simulate the second (or foreign) language acquisition primarily done through classroom instructions. We must also note that L2 learners normally rely on *explicit rules* known as *grammar*, in particular, at their initial stage of learning. This is also related closely to the issue of *consciousness* in second language learning; namely, learners internalize the target language system when they consciously engage in learning tasks (Schmidt, 1990; see also Schmidt 1988). Are there any connectionist models which can satisfactorily handle those issues? As far as I know, Smolensky (1988) provides one of the models which may relevantly cope with the issues addressed above.[1]

The basic assumption of Smolensky's (1988) model is that the networks may develop the capacity to interpret and produce *external* symbols. To put it another way, we can teach a network to "use" a system in which information, including rules (such as grammar), can be encoded symbolically. Within the framework of "Proper Treatment of Connectionism (PTC)," Smolensky proposes two virtual machines: the *conscious rule interpreter* and the *intuitive processor*. With regard to the conscious rule interpreter, Smolensky first set up what he called the *cultural knowledge*. Three properties of this special knowledge are presented below:

a. *Public access:* The knowledge is accessible to many people.
b. *Reliability:* Different people (or the same person at different times) can reliably check whether conclusions have been validly reached.
c. *Formality, bootstrapping, universality*: The inferential operations require very little experience with the domain to which the symbols refer. (p. 4)

According to Smolensky, the goal at the cultural level is to express knowledge in a form that can be carried out in a reliable manner by different people; and thus the top-level conscious processor of individual people (i.e., the conscious rule interpreter) is essential. In case of SLA, learners at the beginning proficiency level are much more likely to rely on this conscious rule interpreter to internalize unfamiliar rules or symbols (i.e., their target language grammar and other systems).

On the other hand, the intuitive processor is closely related to the notion of *expert systems*; namely, this processor is completely dependent on ample experience. Smolensky argues that the above cultural knowledge is a completely irrelevant precondition to the intuitive processor; and the intuitive processing does not require any conscious rule application at all. Practically all skilled performance is caused by the intuitive processor. In case of SLA, learners with advanced- or near-native-level proficiency more likely rely on this processor though they may simultaneously adopt the conscious rule interpreter.

Traditionally, in the symbolic paradigm, both conscious rule application and intuition are described at the *conceptual* level (as conscious and unconscious rule interpretations). On the other hand, in the traditional subsymbolic (connectionist) paradigm, conscious rule application can be formalized at the conceptual level but intuition must be formalized at the *subconceptual* level. Smolensky casts doubt on this treatment of the conceptual and subconceptual levels in the traditional subsymbolic paradigm. According to Smolensky, two formalisms should be continuous rather than dichotomous, so that "the hybrid system (would) evolve with experience, reflecting the development of intuition and the subsequent remission of conscious rule application" (p. 12, parentheses mine). Accordingly, Smolensky strongly claims that both intuition and conscious rule interpretations must be treated at the *subconceptual* level in the subsymbolic paradigm. From this perspective, then, explicit rules or external symbols, such as target language grammar, are expected to be handled in connectionist simulations at the subconceptual level. If this kind of a simulation is possible, the issues addressed above will be resolved to a great extent.

Future Directions of Connectionism

The above overview of the connectionist models suggests that the connectionist paradigm will provide a strong theoretical base for explicating the nature of human cognition in place of the symbolic paradigm. Although there are some controversies over the form of knowledge representation among connectionists (e.g., localist view vs. distributed view),[2] no one can deny a great role of the connectionist models played in future research into human cognition. Toward that end, however, connectionist researchers must overcome many problems that remain to be solved. Two most crucial problems are raised here.

The first problem is that, as repeatedly pointed out above, the current connectionist networks are still primitive to simulate the complex systems of human cognition. Hence, more elaborate architectures should be developed so that constituting units can dynamically interact with various environments. One way to do this may be to increase the number of hidden units than necessary as suggested by Bechtel and Abrahamsen (1991). The further refinement of the recurrent network is also promising in this regard.

The second crucial problem for connectionist researchers is the issue of *innateness* or *poverty of the stimulus* (Hatch, et al., 1990; Gasser, 1990; Sokolik, 1990; and others). Many opponents of the connectionist models have been criticizing this subsymbolic approach in the following way: The initial state of a computer model is *tabula rasa,* whereas a child certainly has something at his birth which is genetically programmed and interacts with a language in the environment, though this does not have to be Universal Grammar (see Sokolik, 1990). In order to treat this issue, cognitive scientists should first clarify and define what the initial state is. At the same time, the following point raised by Gasser (1990) should also be verified: "the network, without the help of symbolic predispositions, can produce structures which it has never been exposed to and, at the same time, recognize as anomalous other structures which it has not been exposed to" (p. 189).

In the area of SLA, the influence of the connectionist models has already been noted. Other than the empirical research in this area, the connectionist models are now affecting the theory construction in SLA. For instance, Schumann (1990) attempts to extend the scope of his Acculturation/Pidginization Model by including some cognitive components; and a connectionist model is one of them. Furthermore, Hatch, Shirai, and Fantuzzi (1990) strongly claim the combination of a "neurally plausible theory of learning" and a "pedagogically plausible theory of teaching" (p. 712). This suggests the need for an integrated theory of SLA based on the connectionist models.

As a matter of fact, the connectionist models have a potential for attaining a unified theory of SLA due to its relatively strong explanatory power. Ellis and Schmidt (in press) raise five strengths of the connectionist models as follows: (1) they are neurally inspired; (2) they incorporate distributed representation as well as control of information; (3) they enables prototypical representations to emerge as a natural outcome of the learning process rather than as the result of the operation of pre-specified innate faculty; (4) they show graceful degradation; and (5) they are essentially models of learning and acquisition rather than static descriptions. In view of these strengths, it is now greatly expected that the connectionist models may eventually be able to deal with phenomena observed in the whole areas of language acquisition (see Spolsky, 1989). Toward a complete general theory of SLA, more extensive refinement and greater development of the current connectionist architectures are surely to be expected.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Richard Schmidt, University of Hawaii at Manoa, for his valuable comments on an earlier version of this paper. Responsibility for any errors in the paper, however, lies solely with the author.

NOTES

1. Norman (1986) also provides a system which involves conscious awareness of the learning process, which is called "deliberate conscious control (DCC)." This system manifests a serial processing, which simultaneously works with a parallel processing.

2. The localist view claims that units represent particular concepts; and the distributed view contends that each unit represents many concepts while concepts are distributed over many units (see Gasser, 1990, for more details).

REFERENCES

Bechtel, W. & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks.* Cambridge, MA: Basil Blackwell.

Blackwell, A., & Broeder, P. (1992). Interference and facilitation in SLA: A connectionist perspective. (seminar on Parallel Distributed Processing and Natural Language Processing, San Diego, UCSD, May 1992). (cited in Broeder, P., & Plunkett, K. (1994)).

Broeder, P., & Plunkett, K. (1994). Connectionism and second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 421-453). San Diego, CA: Academic Press.

Brown, R. (1973). *A first language.* Cambridge, MA: Harvard University Press.

Carroll, S. E. (1995). The hidden dangers of computer modelling: Remarks on Sokolik and Smith's connectionist learning model of French gender. *Second Language Research, 11,* 193-205.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. (1986). *Knowledge of language, its nature, origin, and use.* New York: Praeger.

Ellis, N. C., & Schmidt, R. W. (in press). Morphology and longer-distance dependencies: Laboratory research illuminating the **A** in SLA. To appear in *Studies in Second Language Acquisition, 19,* 2.

Elman, J. L. (1988). Finding structure in time. *Cognitive Science, 14,* 179-211.

Fodor, J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3-71.

Gasser, M. (1990). Connectionism and universals of second langauge acquisition. *Studies in Second Language Acquisition, 12,* 179-199.

Hatch, E., Shirai, Y., & Fantuzzi, C. (1990). The need for an integrated theory: Connecting Modules. *TESOL Quarterly, 24,* 697-716.

Henning, M. J. (1989). *Variable rules of forms of address: A connectionist model.* Manuscript. University of Hawaii at Manoa.

Hinton, G. E. & Sejnowski, T. J. (1983). Optimal perceptual inference. In *Proceedings of the Institute of Electronic and Electrical Engineers Computer Society Conference on Computer Vision and Pattern Recognition.* Washington, DC: IEEE, 448-453.

Hinton, G. E. & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1.* Cambridge, MA: MIT Press.

Kellerman, E. (1978). Giving learners a break: Native speaker intuitions as a source of predictions about transferability. *Working Papers on Bilingualism, 15,* 59-92.

Lachter, J. & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning: A contrastive critique of some connectionist learning models. *Cognition, 28,* 195-247.

Manzini, M. R. & Wexler, K. (1987). Parameters, binding theory, and learnability. *Linguistic Inquiry, 18,* 3, 413-444.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol, 2, Psychological and biological models.* Cambridge, MA: MIT Press.

Minsky, M. A. & Papert, S. (1969). *Perceptrons.* Cambridge, MA: MIT Press.

Ney, J. W. & Pearson, B. A. (1990). Connectionism as a model of language learning: Parallels in foreign language teaching. *The Modern Language Journal, 74,* iv, 474-482.

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1.* Cambridge, MA: MIT Press.

Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73-193.

Plunkett, K. (1988). Parallel distributed processing. *Psyke and Logos, 2,* 307-336.

Rosenblatt, F. (1962). *The principles of neurodynamics.* New York: Spartan.

Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E., Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol 2.* Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol 1.* Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986b). Learning representations by back-propagating errors. *Nature, 323*, 533-536.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1, Foundations.* Cambridge, MA: MIT Press.

Schmidt, R. W. (1988). The potential of parallel distributed processing for S.L.A. theory and research. *University of Hawaii Working Papers in ESL, 7*, 1, 55-66.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schneider, W. (1987). Connectionism: Is it a paradigm shift for psychology? *Behavior Research Methods, Instruments, and Computers, 19*, 73-83.

Schumann, J. H. (1990). Extending the scope of the Acculturation/Pidginization Model to include cognition. *TESOL Quarterly, 24*, 667-684.

Skehan, P. (1989). *Individual differences in second language learning.* London: Edward Arnold.

Slobin, D. I. (1971). On the learning of morphological rules. In D. I. Slobin (Ed.), *The ontogenesis of grammar.* New York: Academic Press.

Smolensky, P. (1986). Information processing in dynamic systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1.* Cambridge, MA: MIT Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1-74.

Sokolik, M.E. (1990). Learning without rules: PDP and a resolution of the adult language learning paradox. *TESOL Quarterly, 24*, 685-696.

Sokolik, M. E. & Smith, M. E. (1992). Assignment of gender to French nouns in primary and secondary language: A connectionist model. *Second Language Research, 8*, 1. (Also, Sokolik, M. E. & Smith, M.E., (1989). French gender recognition: A network model and implications for second language acquisition. Paper presented at the Ninth Second Language Research Forum, Los Angeles, CA, 1989.)

Spolsky, B. (1989). *Conditions for second language learning.* Oxford: Oxford University Press.

Takahashi, S. (1991). *Analyzing the governing category and proper antecedent of the Japanese reflexive zibun-zisin.* Manuscript. University of Hawaii at Manoa.

Tarone, E. (1988). *Variation in interlanguage.* London: Edward Arnold.