EXCHANGE

Connectionism: Explanation or Implementation?

Cheryl Fantuzzi
University of California, Los Angeles

Although purely empiricist, or environment-based, theories of language acquisition suffered some serious setbacks with the rise of generative grammar, they have very recently come into vogue again with a new brand of cognitive modeling known as connectionism. Connectionism represents the strongest form of empiricism: radical connectionists typically argue that all learning is based on the processing of input, and that there is no need to posit any a priori internal structure to the processing system at all. What linguists describe as rule-governed behavior, radical connectionists say is only a description of the emergent behavior of the processor. Learning is simply a matter of strengthening and weakening neural connections in response to the statistical frequency of patterns in the input, and the abstract symbols and rules that are so crucial to current linguistic theory have no place at all in a connectionist system. There have been varied responses to these strong claims. Some have been wildly enthusiastic about the new approach (e.g., Sampson, 1987) while others have severely criticized many of its claims (e.g., the papers in Pinker & Mehler, 1989). There is also an extensive middle ground between the two extremes however. While "pure" Parallel Distributed Processing (PDP) models seem to work best with problems involving motor control or the earliest stages of visual processing, connectionists working with more complex cognitive processes such as language or problem solving have often incorporated symbols into their connectionist architectures (e.g., the papers in Hinton, 1991).

There has recently been some interest in the applicability of connectionist models to Second Language Acquisition (SLA) theory and research (Schmidt, 1988; Gasser, 1990; Sokolik, 1990). In the last issue of *IAL,* Yas Shirai (1992) added his voice

Issues in Applied Linguistics ISSN 1050-4273

Vol. 3 No. 2 1992 319-340

to this growing literature by arguing that the connectionist framework effectively explains L1 transfer in SLA. He also argued that since connectionist models present a more neurally plausible model of the lower-levels (i.e., neural level) of language processing, they may provide second language researchers with a new opportunity to look inside "the black box" of language behavior. In this special issue on neurobiology and language, it seems particularly appropriate to take a closer look at just how neurally plausible connectionist models are, and if connectionism adequately explains all of the transfer phenomena that Shirai claims that it does. While Shirai and I are in agreement on the need for a general cognitive model of SLA which integrates research in several fields (Fantuzzi, 1989, 1990; Hatch, Shirai & Fantuzzi, 1990), I disagree that connectionism can yet explain the high-level transfer phenomena that Shirai outlines in his article.

More than simply reply to Shirai's claims for a connectionist explanation of language transfer, I will also look more closely at what a connectionist explanation of a cognitive function entails. McCloskey (1991), for instance, argues that "connectionist networks should not be viewed as theories of human cognitive functions, or as simulations of theories, or even as demonstrations of specific theoretical points" (p. 387). An important question, then, is precisely what role connectionism plays in the development of cognitive theories and in the explanation of linguistic phenomena, a question which is of course frequently raised in the connectionist literature itself (e.g., Smolensky, 1988; Fodor & Pylyshyn, 1988; Pinker & Prince, 1988; Minsky & Papert, 1988; McClelland, 1988). This requires a closer and more critical look at some existing connectionist models of language functions than Shirai has given.

PRACTICING CONNECTIONISM WITHOUT A CONNECTIONIST MODEL

Shirai largely relies on one particular implementation of a connectionist model to support his argument that the connectionist framework can provide an explanation for language transfer. This model is Gasser's (1988) localized model of bilingual sentence production. However, he also liberally peppers his discussion with references to other models (including Gasser, 1990) and to a general "connectionist framework". This reference to a generic framework, however, ignores the important differences between the many different types of models, and evades deep discussion of their actual capabilities. Gasser (1990), for example, explicitly points out that connectionist models can not yet model "stages" of acquisition, or environmental factors or monitoring, and it is unclear how they could. Yet Shirai lists discourse/pragmatic knowledge, sociolinguistic context, learning environment, level of proficiency, markedness, age, attention, and monitoring all as conditions on transfer that the connectionist framework can "effectively explain."

As a representative sampling of Shirai's arguments, in this section I will discuss and critique his proposals for modeling the effect of the learning environment, level of proficiency, and sociolinguistic context on L1 transfer. Section 2 will focus on the general issue of the role that connectionist models play in developing theories of cognition. In section 3, I will take a closer look at the claim of neural plausibility, and discuss how connectionism has been used to model age-related conditions on learning. Section 4 concludes that Shirai has not demonstrated how connectionism may provide SLA researchers with new and more sophisticated interpretations of language transfer or new insights into the role of Contrastive Analysis (CA) in predicting language transfer. In my opinion, the precise role that connectionist models of cognition might play in SLA research, beyond Gasser's (1990) first noteworthy attempt, has yet to be articulated.

Learning Environment

Shirai argues that connectionism explains transfer in the classroom setting in that an acquisition-poor learning environment tends to result in a "grammar-translation" approach, which "necessitates that the learner 'connect' L1 to the L2." In naturalistic settings, learners who have little knowledge of L2 and must communicate also "have to make L1-L2 connections between lexical concepts. As a result of this process, L1-L2 connections become stronger and harder to eliminate later" (p. 105). This discussion of learning environment is typical of how Shirai treats most of the conditions on transfer: there is a strong connection between L1 lexicon/L1 concepts and L2 lexicon/L2 concepts, and if L1 concepts are activated while speaking L2, L2 performance will be influenced by L1. The connectionist model implements this via a highly interconnected network of spreading activation. The details of the implementation, however, are not given, and many other questions remain unanswered: if transfer is simply the consequence of L1-L2 connections between lexical concepts, when does transfer not occur? Is the strengthening of connections merely a matter of stimulus-response? If these strong connections are hard to eliminate, how is the L2 ever learned? While it is an open question whether connectionism can address these issues or not, possible answers could only be contained in a specific model. As will be discussed further below, the model that Shirai uses to illustrate his argument was not designed to handle this particular problem.

Shirai himself points out in another section that L2 learners in both classroom and naturalistic settings are strongly guided by conscious strategies, but conscious following of rules is difficult for connectionist models to handle. Since grammar-translation, as much of L2 learning, appears to involve conscious strategies and rule-application, it is not clear just what aspects of transfer connectionism explains. Again, the assertion that transfer occurs because L1 is somehow connected to L2 and activated with language input is abstract enough to be modeled in many ways. A specific model adds the crucial details of how information may be represented and processed, but Shirai's purported explanation for how these conditions might be modeled is very broad and vague.

Level of Proficiency

As another condition on transfer, Shirai notes that a learner of lower proficiency must fall back on L1 syntax. When there is little syntactic knowledge of L2 and the speaker has to say something, she will simply plug L2 words into L1 structures, a process known as "relexification." Transfer thus occurs because the learner must activate her knowledge of L1 in order to produce L2. Here, Shirai explicitly invokes Gasser's (1988) model to explain the phenomenon. In Gasser's model, syntactic information in L1 is partly determined by the lexicon (as in traditional linguistic theory) and by a "sequencing component." The sequencing component is not directly involved in transfer, but is simply necessary for this network to produce sentences at all. Therefore, relexification occurs in the model because general linguistic structures exist and are accessed for both L1 and L2, and when the L2 is less developed, L1 structures are accessed. Proposing general syntactic structure in linguistic representation, however, is not unique to Gasser's model but is, of course, found in many non-connectionist models as well. Again, we have not been shown how connectionism provides a superior account of language transfer.

Sociolinguistic Context

Although he gives us only a partial description of a particular localist model which generates a few simple sentences and cannot model all of the transfer phenomena he describes, Shirai's claim is that the model could be easily augmented to handle it. An example is his discussion of how "sociolinguistic context" aids transfer: "when the learner is speaking with someone from the same culture, the hearer-role (represented as a node) is specified as such. Whether the learner likes it or not (i.e., is goal-driven or not), the hearer-role is activated, which leads to a spreading activation of the nodes connected to it. Thus, the model would be able to show the kinds of adjustment which are called 'accommodation'" (p. 109). An example of accommodation is the observation that native Chinese speakers show more L1 influence when speaking in Thai with other native Chinese speakers than with a Thai speaker. Shirai is not suggesting that the "hearer-role" node alone explains accommodation, but there must be many more nodes connected to that one which must also be specified in the model.

Since all of the connections in Gasser's model have been set by hand, to augment it to show language accommodation would require specifying what the appropriate "sociolinguistic context" for the transfer is, all of the relevant social or cultural information that the hearer-role node is connected to, and all of the connection weights as well, including of course when the hearer-role's effect on transfer would be overridden since the hearer-role must always be activated in a bilingual. This, however, is not a trivial task, considering that a *single sentence* in the model required 292 nodes and 1374 connections!

While Gasser's model might be augmented to incorporate more complex information, it is still not clear how much the transfer phenomena would be explained by it, since the hand-set connections really instantiate the programmer's assumptions of how information should be represented in order to perform a certain task. Consider the example of code-switching that Shirai uses to illustrate the model's general processing style. Gasser cites a case in which a Japanese speaker inserts an English word, "spoil", into an otherwise Japanese sentence, presumably because the "connections" from this concept to linguistic expression are stronger to English than to Japanese (*ano okanemoti wa ozyoosan o spoil sita,* 'That rich man spoiled his daughter,' Gasser, 1988, p. 7). In order to model this behavior, Gasser manually set the connection "weights" so that the English word 'spoil' was stronger than the equivalent Japanese expression, and was thus naturally chosen. Gasser's point in presenting this example was not to demonstrate transfer (which is after all only a general side effect of the processing style), but, rather, how a single concept ('spoil') may map into more than one linguistic structure in the model. These linguistic structures were also specified beforehand by Gasser, through use of such high-level linguistic constructs as "transitive clause," "direct object," and "verb".

Although learning is an important component of transfer, Gasser's model does not actually learn its own representations. However, Shirai speculates on how the connections to 'spoil' may have been acquired: "if a Japanese-speaking learner of English keeps saying 'spoil' instead of 'amayakasu', the association will be stronger, and it will be easier for him to say the word when speaking English. This process would constitute learning the word "spoil" for the learner" (p. 97). As in the earlier examples, though, the argumentation is circular, and Shirai has in effect just restated the problem to be explained. How did the learner acquire the word "spoil" in the first place? How are associations formed, strengthened, or weakened? How are words or sentences actually represented? How may this be implemented in a real brain?

While Shirai uses Gasser's model to broadly illustrate connectionism, Gasser (1988) himself says that his "localized" model is unlike most connectionist approaches in that he started out with the sorts of representations that one would find in a symbolic model, such as linguistic "schemas", and nodes for NPs, direct objects and accusative case. He then implemented them in a connectionist network to see if simple bilingual sentences could be generated with this new processing style. Certain desirable properties, such as automatic generalization and cross-linguistic transfer, come with this type of processing for free. "Pure" connectionist systems, which use fully distributed representations, do not start with such high-level constructs, but develop their own representations. However, these networks cannot perform very complex linguistic tasks, and, as noted above, connectionists working with language often use the more traditional approaches or theories implemented in a connectionist processing style—in other words, hybrid models.

To sum up the discussion so far, I have argued that Shirai has presented neither a clear account of the particular language transfer phenomena to be explained nor a model that would be able to implement such a theory if it existed. As Boden (1990) points out, any computer model, whether symbolic or connectionist, cannot embody a psychological theory without having a theory to begin with. Connectionists have had more success with modeling low-level aspects of vision because neural theories of low-level visual processing exist. The very early stages of visual processing are known to involve massively parallel neural computation, which is suitably implemented within connectionist networks. There is no connectionist theory of language as yet, although traditional linguistic theories of competence and performance work well with more traditional forms of computing in Artificial Intelligence (AI). The next section considers how connectionist models might help to develop such a theory.

CONNECTIONISM: IMPLEMENTATION OR EXPLANATION?

An ongoing debate in the AI literature has been whether a computer simulation of human intelligence may constitute an explanation of it (see, for example, the exchange between Searle, 1990, and Churchland & Churchland, 1990). Some critics of conventional AI as a model of human cognition see connectionism as a more neurally plausible glimpse into the "black box" (e.g., Churchland, 1986; Dreyfus & Dreyfus, 1985), and Shirai is clearly a proponent of this position. However, as we saw above, Shirai merely points to a vague connectionist framework to support this point of view. In this section, I will raise some general problems with viewing connectionist modeling as an explanation of cognition, although it could play a role in developing a theory. My own view is that connectionist and symbolic models are both useful for studying different aspects of cognitive processing.

This, of course, runs counter to Shirai's enthusiastic presentation of connectionism as a potential "paradigm shift" in psychology and linguistics, a proposition that I feel is not only fundamentally misguided, but also unproductive for the field. Shirai fails to mention, for instance, that Clark (1989) considers such polarization to be "an extreme danger" to the cooperative efforts of cognitive science. While Shirai notes that the mechanisms by which a connectionist system might handle conscious rule-learning are "unclear" (p. 108), he relegates Clark's own proposal for a hybrid model to a footnote (but see Fantuzzi, 1991), and says that he argues strongly for the more radical approach "because it offers a new perspective" (p. 114).[1] However, even a radically new perspective does not constitute a paradigm shift and, at the very least, connectionism may be integrated with traditional symbol systems, if not viewed as simply implementing them. The mere existence of hybrid models invalidates the claim that these are separate paradigms, and the vast majority of connectionists see themselves as building upon rather than supplanting previous work (see Boden, 1990, for an excellent review and a discussion of this issue). At any rate, there is just no evidence as yet that connectionist models can provide a better account of linguistic phenomena than symbolic models do.

An argument is typically made that parallel distributed processing is more brain-like than traditional AI computing because serial computations in the brain would take too long, and highly interconnected and redundant connectionist processing units are more like real neurons in that they display graceful degradation rather than complete disruption if one part of the system fails (e.g., Feldman & Ballard, 1982; Sokolik, 1990). However, these and many similar arguments have been countered as being mere details of implementation, and not a principled distinction between connectionist and symbolic models of cognitive processes (Fodor & Pylyshyn, 1988). Many point out that localist connectionist systems are not fundamentally different from symbol systems, since both store patterns representing symbols in the network (e.g., Fodor & Pylyshyn, 1988; Bechtel, 1987). The model on which Shirai bases most of his discussion is of course such a model. Whether or not there are also explicit rules represented in the system is not necessarily a central issue for linguists (Fodor & Pylyshyn, 1988; Stabler, 1983).

The more radical fully distributed connectionist models *do* offer a clearer alternative to symbolic AI since they do not distinguish representations (symbols) from the physical functioning of the system itself (Cummins & Schwartz, 1987). However, an important question is how well these systems can handle complex phenomena such as language. While simple pattern association may be one part of language learning, it is hard to see how sentence structure and abstract symbols such as NP can be completely absent from it, and the current state of connectionist research gives us no reason to discard them. Pinker and Prince (1988), for example, convincingly argue that children's double-marking errors *(wents, thoughted)* are not due to pattern blending, but to the inflection of the wrong stem, although the notion of "stem" or "affix" is unrepresentable in a connectionist system. Kim, Pinker, Prince and Prasada (1991) demonstrate that a native speaker's representation of past tense formation includes the intuition that denominalized verbs take the regular past by default *(The football player grandstanded to the crowd; The baseball player flied out to left field).* This also is unrepresentable in a "pure" connectionist system.[2]

Because fully distributed connectionist systems cannot handle structured knowledge very well, connectionists working with complex problems of language often incorporate symbols into their connectionist architectures (Hinton, 1991). Cummins and Schwartz refer to this type of model as "conservative connectionism," and Pinker and Prince (1988) refer to it as "revisionist-symbol-processing connectionism," as opposed to the radical "eliminativist" position that Churchland (1986) or Shirai advocate. It has even been suggested that this dual-mode of processing may reflect the different types of processing that people seem to do (Schneider and Schiffrin, 1977; Schneider, 1987, 1988; Clark, 1989). When symbols are distributed over many units, one gets the same performance benefits that one has with a connectionist system: robustness, redundancy, resistance to noise or damage, automatic generalization and so on (Fodor & Pylyshyn, 1988), including the sort of transfer effects that are seen in Gasser's (1988) localist model.

One problem with claiming that the fully distributed models can explain a certain phenomenon (a problem that one doesn't have with symbolic models) is that it is difficult to see exactly how the distributed models arrive at their solution. Hinton (1991) points out that when there are several hidden layers, as there are in many of the learning networks, it is hard to say what each hidden unit represents. Although the system arrives at a solution to its task, even its designer does not know how it has done so, and an "existence proof" that a model can perform in a certain way is not a good "explanation." As Klein (1990) notes:

> Connectionists make models tick, but do not make us understand as yet what makes them tick. Turning now to SLA more specifically, we do not just want a network which, when fed with sufficient input in the form of sentences, provides us with the appropriate regular and irregular

> morphology. We want to know the principles according to which the human mind breaks down the sound stream into smaller parts, assigns structure and meaning to these, retreats from false generalizations, and the like. (p. 226)

Clark (1990) uses the text/phoneme conversion model, NETtalk (Sejnowski & Rosenberg, 1986), to illustrate this problem. While the model does not encode traditional phonological rules, it is given a rich prior analysis of its domain through the choice of input and output representation, hidden unit architecture and learning rule. Although the network was 95% successful in its task of converting text to speech after 50,000 trials of "supervised" learning, even its designers don't know how it actually performed the task, and Clark discusses various strategies that have been used to investigate how it was done, and to try to discover the principles by which the system arrives at its solution.

McCloskey (1991) makes essentially the same kinds of observations about Seidenberg and McClelland's (1989) connectionist model of word recognition and naming. He argues that theoretical proposals for cognitive functions, tied to specific descriptions of particular networks, are too vague to be explicit theories of cognitive functions, although they are valuable tools for developing theories. For example, while the performance of Seidenberg and McClelland's model matches the performance of human subjects, the authors cannot specify the things that a theory should specify: what regularities and idiosyncrasies of the orthographic and phonological representation of words are encoded by the network (i.e., how do people encode the different representations of the letter *a* in various contexts); how the acquired knowledge is actually represented in the network; and how the propagation of activation throughout the network results in the appropriate representation being chosen in the appropriate context (e.g., the appropriate *a* in *hat, hate, have).* While Seidenberg and McClelland have provided an explicit computer simulation of a cognitive behavior, McCloskey argues that the underlying theory of human cognition remains vague: just general statements to the effect that representations are distributed and similar words are represented similarly. The problem is that the dynamics of complex nonlinear connectionist systems are difficult to analyze, and thus to understand. Unless one is satisfied with an "existence proof" that something can be modeled, the models still do little by way of explaining the behavior. They again provide a "black box" simulation of a cognitive behavior rather than a theory of it.

Another important difference between connectionism and traditional AI, then, is the way that each constructs theory:

> The connectionist, by whatever means, achieves her high-level understanding of a cognitive task by reflecting on, and tinkering with, a network which has *learnt* to perform the task in question. Unlike the classical Marr-inspired theorist, she does not begin with a well worked out (sentential, symbolic) competence theory and then give it algorithmic flesh. Instead she begins with a level-0.5 theory, trains a network, and then seeks to grasp the high level principles it has come to embody. (Clark, 1990, p. 303, emphasis his)

Thus, having built a distributed network which successfully completes a certain task, the next problem is to seek to understand the principles that caused the behavior, the same task facing a neuroscientist who studies a real neural network. Clark uses the term level 0.5 theory (as opposed to level 0, which is no theory at all and level 3, which would be a high-level competence theory) to refer to the fact that some amount of theory must be used to set up the program in the first place. Differing amounts of a priori theory may be applied to set up a program, but they all make some initial assumptions, even if it is only which features will be represented on which units and how many units or hidden layers are needed to do the task.

One important issue may be the "psychological reality" of these assumptions. Pinker and Prince (1988) and Lachter and Bever (1988), for example, presented a sharp critique of the assumptions underlying Rumelhart and McClelland's (1986) past tense acquisition model. Clark (1990) points out that even when the initial assumptions are minimal, they may be psychologically unrealistic; for example, the amount of units used and the best form of the solution must be specified beforehand. Another question about the psychological reality of connectionist systems that linguists often raise is the assumption that there is an explicit "teacher" which looks at the output and incrementally corrects it, a quite implausible suggestion for first language acquisition.

Fodor and Pylyshyn (1988) argue forcefully that connectionism may at best provide an account of the "abstract neural" structures in which symbolic "cognitive structures" are implemented, each thus representing a distinct level of cognitive modeling. But then again, the tasks that these distributed models can perform, such as transcribing text to speech, adding phonological past tense endings to verbs, or recognizing words from non-words, are not the sort of complex linguistic problems which usually occupy linguists. Furthermore, the behavior of the models doesn't always match human behavior (for discussion see Pinker & Prince, 1988). Clearly, what connectionist models may someday teach us about how humans process language is still very much an open question.

McCloskey (1991) suggests that connectionist models may be viewed as "animal models" of human functions. He argues that an animal model is not itself a theory or a simulation of a theory, but an object of study which may lead to theories of human systems. In the same way, artificial neural networks may also be easier to study and analyze than actual brains, and thus may one day help to develop a theory of human processing, although there is no connectionist theory of cognition as of yet. As Clark says, explanation in connectionism requires, at the minimum, "reflecting on, and tinkering with, a network which has *learnt* to perform the task in question" and then seeking the principles it has come to embody. Arm-chair speculating on the future capability of models, as Shirai does, certainly will not explain issues in SLA. A clearer discussion of theory, explanation and of the underlying assumptions and actual capabilities of existing models must be present in any discussion of the applicability of these models to SLA research.

This section has pointed out some problems with viewing any connectionist model as an "explanation" of linguistic phenomena and takes issue with Shirai's presentation of connectionism as a potential paradigm shift in cognitive modeling. The next section considers how neurobiologically plausible the models are.

CONNECTIONIST MODELS AND NEURAL PLAUSIBILITY

As Rumelhart and McClelland (1986, Part V) point out, connectionist models are considered to be neurally plausible to varying degrees. While certain models of psychological processes, such as Gasser's (1988) sentence production model, are "neurally inspired," one could say that this inspiration is minimal

(Schumann, 1990a). Some connectionist models are much more biologically-oriented, such as Munro's (1986) model of the development of ocular dominance in the visual cortex, and this model will be discussed briefly below.

The Neural Plausibility of Connectionist Algorithms

As noted above, some connectionist learning algorithms have been criticized as being psychologically implausible, since they rely heavily on constant feedback from an external "teacher" who knows what the correct answer should be. One widely-used learning algorithm, known as back-prop, is criticized as being neurally implausible as well, because real neurons do not feed error information back down the neurons so that they can re-adjust their connections (Thorpe & Imbert, 1989).

Shirai describes memory and learning in very general terms: as the strength of connections between "nodes" in a network and the "activation" and "firing" of patterns of nodes. Transfer is the selection of a pattern of nodes that are more strongly associated with an input representation. He describes learning in the connectionist model in this way:

> Essentially, the more often a particular node at the ends of connections are activated and/or fired, the stronger the connections become; consequently stronger connections become more easily activated, and this greater ease of activation causes more learning. (p. 96)

This corresponds to the very simplest learning algorithm, known as the Hebb rule, which is indeed an abstraction of actual neuronal processing (Hebb, 1949). McClelland, Rumelhart, and Hinton (1986) describe Hebbian learning in this way: "When unit A and unit B are simultaneously excited, increase the strength of the connection between them" (p. 36).

This rule may be adjusted to cover both positive and negative activation values (excitation and inhibition). However, McClelland and colleagues go on to say that because the Hebb rule has some limitations, most connectionists do not generally use it in this form for more complex computations but have devised more sophisticated learning algorithms, such as the "delta rule" (which Sokolik 1990, discussed below, uses) and "back-prop", (which Gasser 1990 uses).

While there does seem to be neurobiological support for Hebbian learning, it is unknown how much this very simple type of associative learning, which is observed in simple slugs responding to electric shocks, is involved in higher cognitive functions. Nevertheless, Shirai describes complex human learning in such simple terms. For example, as slugs "learn" to associate shocks with light, Shirai talks about people learning to associate L1 concepts/words with L2 words through the simple strengthening of connections. While we certainly cannot say that some human learning is *not* due to this type of conditioning, it may be a leap of faith to attribute complex language learning or language transfer to simple associations between simultaneous inputs - especially in light of the enormous difficulties connectionist systems have in representing complex linguistic knowledge and, of course, the many convincing arguments from generative linguistics to the contrary. Shirai himself brings up the point that Chomsky (1957, 1959) effectively defeated the behaviorist paradigm, but offers little compelling evidence that a neo-behaviorist revolution is in the making.

According to Shirai, once L1 connections are "formed and solidified as a system," subsequent alteration of connections may become difficult (p. 107). He relates this to the notion of unitization (Kennedy, 1988) at the information-processing level: once knowledge becomes automatized and "solidified" as a unit it is difficult to alter later. However, how connectionism explains age-related transfer at either the so-called neural (connectionist) or psychological level is again quite vague. Although he suggests that "frequent" or "salient" or even innate connections may become "too strong to alter later in life" (p. 103), we are still faced with the problem of how connections are formed, how later learning occurs, how a crucial balance is maintained between the malleability and rigidity of connections, and how real neurons function. Shirai refers to Munro's model of the development of the visual cortex as a possible connectionist explanation of "age-related" conditions on transfer, but offers no discussion of it.

Munro's Model of a Critical Period for Visual Processing

Munro (1986) presents a mathematical model of a specific neural system whose circuitry is relatively well known, the visual cortex. He argues that the degree of plasticity in single neurons may reflect sensitive periods in learning, although sensitive periods do not necessarily reflect decreasing plasticity in neurons. Munro does not propose that existing neural connections may become "solidified" or difficult to modify, but that uncommitted neurons will naturally change their state more rapidly and easily than already committed neurons. Further, he suggests that this type of plasticity may hold only for the earliest stages of cognitive processing, which in the domain of language acquisition might be phoneme recognition, and that higher cognitive processes may not show a sensitive period at all. This is of course an empirical question. Whether Munro's framework might be applied to cognitive systems more generally, and to issues in language acquisition in particular, remains an open question. However, the connection that Shirai attempts to make between Munro's model and issues of transfer in SLA is pitched at much too general a level, which is, simply, that a reduction in the modifiability in neural connections might be one factor leading to a sensitive period for some language functions. We then need to ask which functions, which neurons, why, and how.

Sokolik's PDP Model for a Critical Period in SLA

Munro's approach to modeling a critical period for the visual cortex may be compared with an explicit PDP model of age-dependent conditions on language acquisition that has been proposed in the SLA literature. Sokolik (1990) notes that a protein known as Nerve Growth Factor (NGF) is thought to be linked to rate of learning. If children have a higher amount of this protein than adults, and we assume that NGF affects the ability to learn languages more quickly, then we have a principled physiological basis for setting the "learning rate" in a connectionist algorithm higher for children than for adults. Set at a higher value, a PDP model will learn more quickly, which, Sokolik suggests, may offer an explanation for why children acquire second languages "more readily" than adult language learners do.

Sokolik presents an example in which the learning parameter for the acquisition of a certain feature is set higher in the child PDP model than it is in the adult PDP model. The child model, therefore, attains near mastery of the feature more quickly than the adult model. However, there are at least three problems with this scenario, other than the psychological reality of the learning algorithm itself, as discussed above. Sokolik herself mentions that the significance of NGF to learning rate is speculative, and ignores other factors that may be involved in variable learning for adults, such as psychological and sociological factors (Schumann, 1990b). A second problem is that much empirical research has suggested that adult language learners may at first be quicker at acquiring a second language than children, but that children overtake them in the long run (Long, 1988). A third problem is that, even at the level of the single neuron, assuming that NGF is a factor in learning rate is simply an ad hoc explanation of critical period effects. Munro contrasts his own position with the popular idea that the neurotransmitter norenephrine might act as a global modulator of neuronal plasticity and thus account for lower learning rates. He points out that one may always simply add an explicit learning rate into one's learning rule in order to obtain certain pragmatic results, or factor in global modulators which affect the learning rate, but this is unnecessary. As we saw above, his solution for a critical period is simply that uncommitted neurons form their connections more easily. But again, the translatability of his particular model to language acquisition issues is not straightforward.

What the above discussion makes clear is that the focus of Sokolik's and Munro's PDP models is not on how proficiency changes as a result of changes in the form of the mental representations, but as a result of a change in a learning rule, or in the weights and connectivity of the processing units themselves. A connectionist system is, as Bialystok (1990) points out, the quintessential processing model. But since the models only apply to on-line processing (e.g., learning) and do not apply over time (i.e., to development), they perform quite different tasks than competence models, which are concerned in detail with changes in structured mental representation. Therefore, Bialystok argues, PDP models may be construed as models of processes rather than of the mental representations which are the focus of competence theories. The two approaches, representing different sides of the competence-performance distinction, may eventually co-exist as complementary explanations for different problems.

The important point is not to maintain a strict dichotomy between performance and competence, of course, but to realize that different aspects of cognitive modeling may be reconciled into one whole. This directly relates to the issue of "levels" of explanation mentioned in the last section. The descriptions of cognitive behavior at the level of neural processes, of connectionist networks, or of competence theories may be viewed as different levels of abstraction (Fodor & Pylyshyn, 1988). Clearly, we are not yet at the point where we can say that connectionism, itself an abstraction of neural processing, will ever be able eliminate the higher levels.

CONCLUSION: CA REVISITED

It appears that the general thrust of Shirai's article is to revive a Contrastive Analysis (CA) approach to transfer by buttressing its theoretical framework with connectionism. This topic alone is a broad one, and by focussing on it Shirai may have been able to cover at least one area of his article in more depth.

A radical connectionist approach, which Shirai clearly favors, is incompatible with parameter setting (White, 1985, Flynn, 1987; Hilles, 1986) or with a learner's own internal contribution to learning (e.g., the natural order hypothesis in the acquisition of morphemes). Indeed, Shirai suggests that "natural order" phenomena "can be explained by 'naturalness factors such as perceptual saliency, frequency and invariance of forms, as well as by the 'L1' factor." As he puts it:

> In connectionist terms, such a claim can be interpreted as follows: the naturalness factor makes it easy for a particular form to be connected to a particular meaning/function. It will be easy to identify and easy to match; there will be many opportunities to strengthen connections. This will result in the Natural Order. (p. 100)

However, this explanation completely sidesteps the problem of what is meant by "salient" or "easy" to map, why some frequent items are not learned first, how this mapping is done, etc.—that is, all of the issues that are of interest to SLA researchers. While the simple mapping strategy that Shirai describes is intuitively plausible, it is notoriously difficult to establish causal relations between these "naturalness" factors and language acquisition, as the wide literature on morpheme acquisition shows. Also, as Larsen-Freeman and Long (1991) point out, the claim that morphological development shows much commonality across unrelated languages, pointing to some internally driven organization of input, has simply been too well documented to be ignored.

In a similar vein, Shirai argues that connectionism explains "interlingual" mapping between L1 and L2 because "when a new pattern is encountered which is similar to another existing pattern in the learner's representation, the new pattern would activate the existing pattern" (p. 111). How similar must they be to be activated? What defines similarity? Why are some items not transferred but learned? What happens in the network if patterns are not similar?

Gasser (1990), of course, has applied the "connectionist framework" to an actual test of the CA hypothesis. One thing that he found was that the learning performance of the model was more complex than "traditional" CA would predict, namely that all differences between L1-L2 should be equally difficult to learn. The independent variables of L1/ L2 that Gasser modeled (word order and lexical form) in fact showed an effect of interaction. While he admits that the conditions of the model were a gross oversimplification of an actual language learner's task, Gasser notes that:

> (W)hile these results should be regarded very tentatively, they point to a possible line of connectionist research, one in which networks test out particular hypotheses about transfer and suggest what types of data are needed to flesh out the transfer picture. The main conclusion to be drawn from these simulations is that, even with this extremely simple model of the transfer process, it was impossible to predict precisely how the network would behave. Thus simulations have an important role to play. (p. 196)

While connectionist networks may indeed provide new ways of testing our hypotheses about language processing and learning, simulations serve to help develop and refine our theories of language, not to eliminate them. The connectionist approach as Shirai describes it does not provide new and more sophisticated interpretations of language transfer or new insights into the role of Contrastive Analysis (CA) in predicting language transfer. The role of CA, and the connectionist explanation of transfer, are treated with the same brevity and superficiality as is his discussion of the connectionist framework itself. While Shirai has given us an informative overview of the conditions thought to influence transfer in SLA, its tie-in to connectionism may have benefitted from a more narrow focus, perhaps a closer look at Gasser's (1990) connectionist model of transfer and a more detailed discussion of its implications for a CA position than Gasser himself provides.

As has been made abundantly clear throughout this paper, I believe that Shirai's claim for a connectionist explanation of transfer is greatly overstated. Second language researchers who are to start research projects in the connectionist framework will need to know more precisely how models work and how they may be applied to particular problems. I am of the opinion that connectionist models will probably never replace higher-level explanations in cognitive modeling, although they may help to develop theories at the level of implementation.

ACKNOWLEDGEMENTS

I would like to thank Bruce Burns, Patti Mannix and Charlene Polio for their very helpful comments on an early draft of this paper. I must of course claim responsibility for the final version.

NOTES

[1] Indeed, his token references to some potential compromises belie the obvious prominence given to the "alternative view" throughout the paper. He appears to treat the distinctions between various types of connectionist models as some irrelevant detail.

[2] Although attempts have been made to improve the original model which was the focus of Pinker & Prince's extensive criticisms, these particular problems have not been addressed; indeed, that would require adding symbols to the network.

REFERENCES

Bechtel, W. (1987). Connectionism and the philosophy of mind: An overview. *Southern Journal of Philosophy*, *26* (Suppl.), 17-41.

Bialystok, E. (1990). The competence of processing: classifying theories of second language acquisition. *TESOL Quarterly*, *24*:, 635-648.

Boden, M. (1990). *Computer models of mind.* Cambridge, UK: Cambridge University Press.

Boden, M. (1991). Horses of a different color? In W. Ramsey, S. Stich & D. Rumelhart (Eds.), *Philosophy and connectionist theory.* Hillsdale, NJ: Erlbaum.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1959). A review of B.F. Skinner's *Verbal Behavior*. *Language*, *35*, 26-58.

Churchland, P. S. (1986). *Neurophilosphy: Toward unified understanding of the mind/brain.* Cambridge, MA: MIT Press.

Churchland, P. M. & Churchland, P. S. (1990). Could a machine think? *Scientific American* , *262*, 32-37.

Clark, A. (1989). *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. Cambridge, MA: MIT Press.

Clark, A. (1990). Connectionism, competence and explanation. In M.A. Boden (Ed.), *The philosophy of artificial intelligence*. Oxford, UK: Oxford University Press.

Clark, A. (1991). In defense of explicit rules. In W. Ramsey, S. Stich & D. Rumelhart (Eds.), *Philosophy and connectionist theory*. Hillsdale, NJ: Erlbaum.

Cummins, R. & Schwartz, G. (1987). Radical connectionism. *Southern Journal of Philosophy* , *26* (Suppl.), 43-61.

Dreyfus, H. & Dreyfus, S. (1985). *Mind over machine*. New York: MacMillan/The Free Press.

Fantuzzi, C. (1989). *The relevance of connectionism to theories of second language acquisition*. [Master's thesis]. University of California,Los Angeles.

Fantuzzi, C. (1990). *Connectionism and second language acquisition*. [PhD. Qualifying Paper]. University of California, Los Angeles, Applied Linguistics Dept.

Fantuzzi, C. (1991). Review of *Microcognition* by Andy Clark (1989). *Issues in Applied Linguistics* , *1*, 129-133.

Feldman, J. A. (1986). Neural representation and conceptual knowledge.[Technical Report TR189]. Rochester: University of Rochester.

Feldman, J. A. & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science* , *6*, 205-254.

Flynn, S. (1987). *Parameter -setting models of L2 language acquisition: experimental studies in anaphora*. Dordrecht: Reidel.

Fodor J. A. & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* , *28*:, 3-71. Reproduced in Pinker & Mehler (Eds.) (1989).

Gasser, M. (1988). *A connectionist model of sentence generation in a first and second language*. [Technical Report UCLA-AI-88-13]. Los Angeles: University of Los Angeles, Computer Science Dept.

Gasser, M. (1990). Connectionism and universals of second language acquisition. *SSLA* , *12*:, 179-199.

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

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley. Reproduced in J.A. Anderson & E. Rosenfeld (Eds.), (1988). *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.

Hilles, S. (1986). Interlanguage and the pro-drop parameter. *Second Language Research* , *2*, 33-52.

Hinton, G. E. (Ed.) (1991). *Connectionist symbol processing*. Cambridge, MA: MIT Press.

Hinton, G., McClelland, J. L. & Rumelhart, D.E. (1986). Distributed representations. In J. McClelland & D. Rumelhart (Eds.).

Kennedy, B. L. (1988). Adult versus child L2 acquisition: An information-processing approach. *Language Learning* , *38*, 477-495.

Kim, J. J., Pinker, S., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science*, *15*, 173- 218.

Klein. W. (1990). A theory of language acquisition is not so easy. *SSLA*, *12*, 219-231.

Lachter J. & Bever, T. (1988). The relationship between linguistic structure and associative theories of language learning: A constructive criticism of some connectionist learning models. *Cognition*, *28*, 195-247. Reproduced in Pinker & Mehler (Eds.), (1989).

Larsen-Freeman, D. & Long, M. (1991). *An introduction to second language acquisition research.* London, UK: Longman.

Long, M. (1988). Maturational constraints on language development. *University of Hawaii Working Papers in ESL*, 7(1), 1-53.

McClelland, J. L. (1988). Connectionist models and psychological evidence. *Journal of Memory and Language*, *27*, 107-123.

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

McClelland, J. L., Rumelhart, D. E. & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland, (Eds.).

McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science*, *2*, 387-394.

Minsky M. & Papert, S. (1988). *Perceptrons: An introduction to computational geometry* (Expanded ed.). Cambridge, MA: MIT Press.

Munro, P. W. (1986). State-dependent factors in neural plasticity: A partial account of the critical period. In J. L. McClelland & D. E. Rumelhart, (Eds).

Pinker, S. & Mehler, J. (Eds.). (1989). *Connections and symbols.* Cambridge, MA: MIT Press.

Pinker, S. & Prince, A. (1988). On language and connectionism. *Cognition*, *28*, 73-193. Reprinted in Pinker & Mehler (Eds.), (1989).

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

Sampson, G. (1987). Parallel distributed processing. Review article, *Language*, *63*, 871-876.

Schmidt, R. (1988). The potential of PDP for SLA theory and research. *University of Hawaii Working Papers in ESL*, 7: 1, 55-66.

Schneider, W. (1987). Connectionism: Is it a paradigm shift for psychology? *Behavior Research Methods, Instruments & Computers*, *19*:(2), 73- 83.

Schneider, W. (1988). Commentary to Smolensky. *Behavior and Brain Sciences*, *11*, 52.

Schneider, W. & Shiffrin, R. (1977). Controlled and automatic processing. 1: Detection, search and attention, *Psychological Review*, *84*, 1-64.

Schumann, J. (1990a). Extending the scope of the acculturation/pidginization model to include cognition. *TESOL Quarterly*, *24*, 667-684.

Schumann, J. (1990b). The role of the amygdala in mediating affect and cognition in second language acquisition. In *Georgetown University Round Table on Language and Linguistics 1990.* (pp. 169-176). Washington, DC: Georgetown University Press.

Searle, J. R. (1990). Is the brain's mind a computer program? *Scientific American*, *262*(1), 26-31.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review* , *96*, 523-568.
Sejnowski, T. J. & Rosenberg, C. R. (1986). NETtalk: A parallel network that learns to read aloud. [Technical Report JHU/EEC-86/01] John Hopkins University. Reproduced in J.A. Anderson & E. Rosenfeld (Eds.), (1988). *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.
Shirai, Y. (1992). Conditions on transfer: A connectionist approach. *Issues in Applied Linguistics*, *3*, 91-120.
Smolensky, P. (1988). On the proper treatment of connectionism. *The 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.
Stabler, E. P. (1983). How are grammars represented? *The Behavioral and Brain Sciences*, *6*: 391-421.
Thorpe, S. J. & Imbert, M. (1989). Biological constraints on connectionist modeling. I R. Pfeifer, Z. Schreter, F. Fogelman-Soulie, & L. Steels (eds.). *Connectionism in perspective*. Amsterdam: Elsevier Science Publishers B.V.
White, L. (1985). The acquisition of parameterized grammar: Subjacency in second language acquisition. *Second Language Research*, *1*, 1-17.

Cheryl Fantuzzi is a doctoral student in applied linguistics at UCLA, specializing in second language acquisition. Her special interests include research in cognitive science and developmental psycholinguistics.