Does Conceptualization Equal Explanation in SLA?

Cheryl Fantuzzi
University of California, Los Angeles

In the last issue of *IAL,* Shirai and Yap responded to my critique (December, 1992) of Shirai's (June, 1992) article on connectionism and language transfer. In their reply, Shirai and Yap characterized my paper as generally "questioning the merits of the connectionist paradigm," and stated that, in addition to critiquing Shirai's discussion of connectionism and transfer, my goal was to "present the weaknesses of connectionism in general" (p. 120). Dismissing my critique of Shirai's discussion of connectionism and transfer as a misinterpretation of Shirai's purpose, they focussed their reply on my "general criticisms" of connectionism and argued for the usefulness of the "connectionist framework" for developing "a general theory of second language acquisition." "What is needed on all sides," they wrote, "is a spirit of openness that is conducive to scientific inquiry (p. 125)." However, since it was never my aim to attack connectionist research, Shirai and Yap's defense of connectionism as a general "conceptual framework" for SLA was a moot argument. My critique was not aimed at "connectionism in general," but at Shirai's *particular claims* for connectionism, and especially his claim for a *connectionist explanation of language transfer*. The purpose of my article was to take a closer look at some of the issues involved in making such a claim, and at the models that Shirai used to support his argument. Although Shirai and Yap contend that my criticisms of Shirai's paper were based on a misunderstanding of his purpose, I do not believe that I "missed the point" of Shirai's argument; rather, I *disagree* that *vague and general statements* (connectionist or otherwise) offer elegant and unifying "theoretical explanations" of SLA phenomena. Since I too believe that our field stands to benefit from a clear discussion of the possibilities and limitations of connectionist research, I will

Issues in Applied Linguistics
ISSN 1050-4273

Vol. 4 No. 2 1993 295-314

continue this exchange by attempting to clarify what I perceive to be at issue in this discussion. I will also contrast Shirai and Yap's "defense of connectionism" with Seidenberg's (in press) own reply to McCloskey's (1991) critical comments on his model.

From my perspective, this exchange is not a debate about "connectionism versus symbolism" or about the general "merits" of connectionist research. It is about Shirai's claim for a connectionist explanation of language transfer, and it is also about theorizing in SLA. I disagree with Shirai and Yap's characterization of connectionism as a "rational epistemology" that may solve the proliferation of "too much empirical data" and "too many theories" in SLA, and with the notion that vague theoretical explanations of "messy" cognition are "elegant" and "the only possible result" for connectionist systems. I also disagree with Shirai and Yap's statement that "(a)t the *general conceptual level,* connectionism can *explain* a wide range of phenomena" in SLA (p.126, emphasis mine).

My previous paper pointed out some problems with Shirai's claim that connectionism could "effectively explain" transfer in SLA and raised the general issue of explanation versus implementation in connectionist modeling. One of my stated goals was to consider what a connectionist explanation of cognitive functioning meant, taking a closer and more critical look at existing models than Shirai had provided. This was not done to *attack* connectionism, but to bring some of the issues into better focus. Clark (1990), who I cited extensively in that section, argues that the connectionist "inversion" of traditional explanations (that is, explanation built bottom-up from a working model) has certain *advantages* over traditional approaches. That was not the issue for me. My point was that it is too soon to say whether connectionism can offer a truly explanatory account of SLA phenomena, and the models that Shirai cited could not handle the particular transfer phenomena that he outlined. I also took issue with his presentation of connectionism as a "paradigm shift" in cognitive science[1] and his claim that connectionist models give us a glimpse into the "black box" of language processing.[2] I argued that *Shirai's claims* for connectionism—explanation of transfer in SLA, paradigm shift, and neural plausibility—were not supported by *his discussion* of various models.[3]

I also pointed out that Shirai's reference to a generic "connectionist framework" was too vague, and that by ignoring how actual models worked he sidestepped important issues. Shirai and Yap say that I "wrongly assumed that Shirai was making some very concrete and specific (i.e., microstructural) claims regarding connectionism and transfer" (p.121), and that my critique was based on the mistaken assumption that his discussion was "at the level of instantiation/implementation." However, my critique was just the opposite: I stated that while it is still an open question whether connectionism can address issues in SLA, *possible answers could only be contained in specific models,* and Shirai's purported explanation of transfer was too broad and vague. The real source of our disagreement appears to lie in the role that we assign to connectionist models in connectionist theorizing. Shirai and Yap explicitly promote "speculative theorizing" that may perhaps later be formalized by computer simulation (pp. 127-128), while I argued that "(a)rm-chair speculating on the future capability of [connectionist] models, as Shirai does, certainly will not explain issues in SLA," and that "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 [connectionism] to SLA research" (p. 330, emphasis mine). I believe that discussion of connectionist implementations is integral to connectionist explanations, as I will elaborate below.

Shirai and Yap also emphasize that Shirai's discussion was not aimed at the "implementational level" but at the "conceptual level," but the sharp distinction that they draw between a conceptual and an implementational "level" is not clear. What seems to most clearly distinguish connectionist models from classical models is that the level of implementation *is* the level of explanation (Fodor & Pylyshyn, 1988). I agree that problems with particular implementations do not necessarily call the entire "connectionist conceptualization" into question (MacWhinney & Leinbach, 1990); however, the implementation and analysis of connectionist models are still fundamental to connectionist explanations. My critique was that Shirai's discussion of connectionism was *so general* and *disembodied from implementations* that it encompassed *all kinds of models*, connectionist, non-connectionist or hybrid. The "global theoretical framework" that Shirai and Yap present is really just cognitive science, and they have not made clear what *connectionism* can add to second language research.

A discussion of the applicability of connectionist models to SLA research is by no means an easy task, and Shirai and Yap must be commended for their attempt to bring connectionism to the general attention of SLA researchers. However, I felt that Shirai (1992) went beyond this goal and made much stronger claims, and this was the focus of my critique. Shirai stated that his purpose was to "comprehensively discuss the conditions under which L1 transfer tends to occur and to *explain* these conditions in terms of the connectionist framework" and to "argue that *the connectionist framework explains L1 transfer effectively*" (p. 91, emphasis mine). He stated that "this paper will argue [that] the connectionist approach may provide *new and more sophisticated interpretations of language transfer* as well as *new insights into the role of contrastive analysis in predicting language transfer"* (p. 93, emphasis mine) and that it will "attempt to *explain the mechanisms* of language transfer using the connectionist framework" (p. 97, emphasis mine). Shirai suggested that connectionism may help to clear up "*the confusion created when universals in acquisition were over-emphasized*" (p.112, emphasis mine), and that "the connectionist framework, as presented in this paper, may contribute further to the *specification of L1 transfer:* which factors condition transfer and the role transfer plays in second language processing and acquisition" (p. 113, emphasis mine). My critique was that Shirai made some strong claims for connectionism without adequately backing them up.

Shirai and Yap continue to make strong claims in their reply.[4] In a discussion of the role of connectionism in theory construction in SLA, they say that connectionism can explain a wide range of phenomena at a general conceptual level. They view connectionism as a "rational epistemology" which provides a small number of theoretical constructs that can "integrate" and "make sense of" a wide range of data. A rational epistemology, they say, allows theories to "emerge as inventions, products of cognition rather than empirical observation" (p. 126), and "vague statements at the general conceptual level" may later be "formalized/quantified" through network simulations (p. 127), but at this early stage of theory construction what is needed are "general conceptual statements." It seems odd, however, to think of connectionism as a rational epistemology, considering the importance of simulations for connectionist modelers and, as I will discuss below, I think that Shirai and Yap have the role of connectionist modelling in connectionist theory-construction backwards. I do not agree that

vague statements at the "general conceptual level" offer a unifying or elegant "explanation" of SLA data, nor that vagueness is the best that we can expect from connectionist research, at this stage or any other. That connectionism may provide us with "vague explanations" hardly seems to be a sound argument in *defense* of its usefulness in developing theories of SLA, and that is certainly not how Seidenberg (in press) defends his own model (to be discussed below).

VAGUE THEORIES VERSUS THEORIES OF VAGUE PHENOMENA

In my previous paper, I briefly discussed McCloskey's (1991) argument that connectionism provides us with vague statements about cognitive functioning rather than explicit theories, and that connectionist models might be best viewed as "animal models" that may help to develop theories of human cognition. Shirai and Yap respond that "vagueness" is all that we can expect from connectionist models such as Seidenberg and McClelland's (1989), since some phenomena, such as sound-spelling correspondences, cannot be precisely predicted: "For such systems that cannot be handled by rules...the only possible result is something vague" (p. 122). They say that once we realize that human cognition is essentially "vague and messy," we may need to change our notion of explanation: "To always expect precision may be misguided... theoretical explanations can be 'vague' (in the sense that they make general statements rather than precise descriptions/ explanation) if they offer attractive advantages such as elegance, consistency and 'making senseness'" (p. 123). Vagueness is consistent with Shirai's aim to provide a "global framework" that simply includes everything, since vagueness is compatible with everything. However, as I discussed in my previous article, the job of *theoretical explanation* is not that easy, and perhaps especially difficult for a connectionist modeler. As Boden (1988) suggests,

> task analysis is needed for theories of learning, whether connectionist or not. To explain how systems learn to do *x*, we must understand what counts as *x*-ing and what it is necessary to be able to

> do in order to be able to do *x*. There is no painless road to the explanation of learning. (p. 224)

McCloskey (1991) argues that rather than explicit theories of cognitive functions, many of the theoretical proposals made by connectionists are just general statements such as "representations are distributed and similar words have similar representations." Even though the details of a *network's* functioning may be explicit, this does not necessarily provide an explicit theory of *cognitive* functioning because many times the model's designer cannot say what knowledge is represented in the network, just how it is encoded or processed, which aspects of the network's function are crucial or irrelevant to its performance, and so on. In the more radical models, the modeler "grows" the network by relying on sophisticated learning algorithms for building complex nonlinear systems that are difficult to analyze. McCloskey argues that mimicking a cognitive function does not mean that the modeler has an explicit theory of that function, just as a gardener who grows a plant from a seed does not necessarily have a theory of plant physiology. This is not true for symbolic models, since a traditional cognitive modeler builds the model from an existing theory. The traditional modeler "must build in each of the crucial features of an independently specified theory. If the theory is not explicitly formulated, the simulation cannot be built" (p. 391).

Shirai and Yap's argument is that since most of cognition is "vague and messy," we may have to settle for vagueness in our "descriptions/explanations" of cognition (the slashed term is theirs). However, they confuse vague theories with theories of vague phenomena. While "messy" phenomena may not be captured well by categorical rules, it does not mean that our theories of these phenomena must be vague. It is possible, at the very least, to have a symbolic theory of semi-regular patterns or of metaphor implemented in a connectionist architecture; symbolic theories can easily be implemented in constraint-satisfaction networks (Pinker, 1987; Fodor & Pylyshyn, 1988). Athough Shirai and Yap suggest that proposing "soft laws" to capture irregular patterns is a "paradigm shift" away from generative linguistics, Marcus, Brinkmann, Clahsen, Wiese, Woest and Pinker (1993) point out that many generative linguists have proposed these sort of "soft laws" to capture the semi-regular patterns of the irregular past tense in English (Jackendoff, 1975; Aronoff, 1976; Lieber, 1980;

Perlmutter, 1988; Spencer, 1990), and "soft laws" are perfectly compatible with symbolic theories.[5] Vague theoretical explanations are certainly not the only possible *nor the most desirable* result for connectionist systems.

Shirai and Yap equate McCloskey's (1991) argument that connectionist proposals are vague with my argument that Shirai's statements are vague, but this is not the same argument at all. Shirai's argument is vague because although he champions a connectionist alternative to symbolic modelling, he does not offer a substantive discussion of any issue or of any model. He purposely aims his discussion at a vague "conceptual" level, distinct from connectionist implementations. McCloskey (1991), on the other hand, provides a coherent argument for why connectionist proposals are vague statements and not explicit theories of cognitive functioning.

VAGUE PROPOSALS VERSUS EXPLICIT THEORIES

McCloskey (1991) gives these guidelines of what a theory of cognitive functioning should include: 1) the theory should organize data in such a way as to allow generalizations to be stated (e.g., because sound-spelling correspondences are accomplished in a certain way, certain variables will affect performance and others will not); 2) the theory should support clear credit-blame assignment (which factors are responsible for correct predictions and which for incorrect predictions); 3) the theory should provide a basis for discerning its differences and similarities from other theories in the field. McCloskey argues that because a connectionist model is simply grown from a learning algorithm and is very difficult to analyze, its designer cannot say just which aspects of the model's structure and functioning are responsible for its performance or are irrelevant to it. They can only make very general statements such as representations are distributed and processing is accomplished by the spread of activation throughout the network. McCloskey uses Seidenberg and McClelland's (1989) model of word recognition and naming to illustrate the difficulties in claiming that the fully distributed models provide a theory of cognitive functions: Seidenberg and McClelland cannot say what idiosyncracies and regularities are captured by the network and how the network

represents the information, which details of the simulation are relevant or irrelevant to its behavior, or just how it differs from other models, since the models are too complex to analyze.

Interestingly, Seidenberg (in press) accepts the criteria named by McCloskey as the criteria needed for a theory of *descriptive adequacy* (Chomsky, 1965), but argues that symbolic models and connectionist models are in the same boat when it comes to descriptive adequacy: both simulate cognitive behavior but neither one explains it. To McCloskey's criteria for descriptive adequacy, Seidenberg adds other criteria, which allows for a theory of *explanatory adequacy*: 4) the theory must explain phenomena in terms of independently-motivated principles; 5) the theory shows how phenomena previously thought to be unrelated actually derive from a common underlying source. In other words, rather than merely implementing and simulating existing domain-specific theories in computer models, an explanatory theory provides a small and independent set of explanatory principles that can explain a wide range of data, and Seidenberg believes that connectionism can contribute to such a theory. [6]

We are familiar with such principles and constraints from current linguistic theory. Seidenberg argues that connectionism may also provide explanatory principles. However, he states that "given the present state of our understanding, *these principles are largely concerned with the properties of artificial neural networks"* (p. 8, emphasis mine), and "it isn't by any means clear yet whether connectionism provides an adequate set of principles" as it relies on the "analysis of [connectionist] systems—for example, *determining what kinds of problems can and cannot be solved by neural networks of a given size and type*—[which] proceeds slowly" (p. 22, footnote 4, emphasis mine). Seidenberg illustrates his argument with connectionist models of aphasia, which, to my mind, it is not that different from McCloskey's suggestion that connectionist models may provide "animal models" of cognitive functioning that may lead to theories of human cognition. While McCloskey notes that connectionist models allow manipulations that human subjects do not and may be simpler and easier to analyze, and therefore help to develop theories of the human system, Seidenberg says that it is hoped that the theoretical principles gleaned from the properties of connectionist systems will eventually "evolve into the relevant neurophysiological ones."

Whether McCloskey's and Seidenberg's views can be reconciled is not the central issue here. The point that I wish to emphasize is Seidenberg's argument that connectionist principles derive from the *computational* properties of connectionist models, which he likens to the constraints on the very general principle of *move-alpha* in linguistic theory. These constraints are an essential part of the theory. No one would seriously characterize Government and Binding (GB) Theory simply as *move-alpha* (i.e., *move anything anywhere),* but Shirai and Yap's characterization of connectionist "theory" as "a small number of theoretical constructs such as nodes, activation, connections and hidden units" is just that general. Theoretical constructs such as NP, lexical root or X-bar structure are only a small part of linguistic theory. The strength of the GB/Minimalist framework is that it makes precise predictions that can be tested, not that it is vaguely "compatible" with everything. If we build a connectionist machine that mimics a behavior but we do not know how it has done so, have we explained it? Don't we already *know* that the brain is a connectionist "black box?" The question of how much artificial neural networks are like *real brains* cannot be answered by speculation.

As I stated in my previous paper, what connectionism may contribute to a theory of cognition is still an open question, but the *analysis of connectionist implementations* is certainly not of marginal relevance to this question: it is central to it. The claim that a connectionist theory/framework offers explanations/conceptualizations at the level of description/explanation is simply *too vague* to be useful for constructing theories of SLA. Although Shirai and Yap characterize connectionist theory-building as "speculative theorizing" at a "conceptual level," which may then be "formalized" through network simulations, I believe that they have the relationship backwards. Connectionist *implementations* are clearly integral to *theory construction for connectionists*. Symbolic models may start out with a theory, but the theory must be *precisely* specified in order to be implemented. Whichever way one chooses to approach *explanation*, it requires much more than the general *conceptualization* of models, and this was the crux of my critique of Shirai's paper.

MORE MODELS, BUT NOT MORE EXPLANATION

Shirai and Yap mention some more connectionist models in their article but, like Shirai's (1992) cursory description of connectionism and transfer, the discussion offers little substance. In response to Shirai's claim that connectionism could *effectively explain* such high-level language phenomena as discourse/pragmatic knowledge, sociolinguistic context, learning environment, level of proficiency, markedness, age, attention and monitoring *as conditions on transfer,* I pointed out Gasser's (1990) comment that connectionist models could not yet model such things as "stages" in learning, environmental factors and monitoring. Again, my point was not that connectionist models will *never* handle all of these phenomena *in principle,* but that *Shirai's particular claims for connectionist explanation of language transfer* were premature. However, by shifting the focus away from my critique of Shirai's discussion of transfer to a general "debate" about connectionism, Shirai and Yap avoid addressing the criticisms that I raised about Shirai's claims and simply introduce a new topic: a connectionist model of "stage-like" acquisition. Shirai and Yap reply that Elman's (1991) model captures "stage-like" incremental learning, noting that it is "not clear" that stages exist in human learning.

In Seidenberg's terms, Elman's model appears to be a model of descriptive adequacy, attempting to simulate the observation that children do not learn their language all at once and that they begin with a limited memory capacity. The model will probably exhibit the problems and limitations of any descriptive model, but Shirai and Yap do not concern themselves with problems or possible solutions. "The most important finding" in their view is that the simulation demonstrated "the importance of simple input at the early stages of development. . . . If children have a learning capacity comparable to a connectionist network, which is very likely, they can learn complex sentences successfully if given simple input at the beginning" (p. 124). The question, however, is not if children have a learning capacity *comparable* to a connectionist model, but *how comparable*? How much like human beings are connectionist models? What do they tell us about human learning?

Shirai and Yap provide *no analysis of the input, of the model, or of human behavior*. They simply note that the model "simulated environmental change by manipulating the input" (p.

123), but what is the nature of structures given to the network and how similar are they to the types of structured input that children get? Is the language that the network learns a possible human language? How successful a language learner is the network and how well does its behavior match actual human behavior? What does the model actually *predict* about human learning, other than that children do not learn their language all at once? and so on. Just the fact that connectionist models *exist* tells us very little. There is of course no clear evidence that "simple" input facilitates language acquisition in children, and suggestive evidence that it does not. The relation of input to language learning is a thorny issue in language acquisition research that can not be so easily answered or brushed aside.

Shirai and Yap also say that the model suggests that children "probably create a prototype based on simple input and generalize it to more complex/varied situations" (p. 124). Again, there is no discussion of *how* this is so. The notion of "prototype" is compatible with both symbolic and connectionist models, but there is the deeper question of whether and how prototype can explain linguistic phenomena, and which phenomena. When Rumelhart and McClelland's (1986) past tense acquisition model simulated a U-shaped learning curve, it also raised many questions that stimulated further discussion and research, such as: How psychologically real is the model (Lachter & Bever, 1988)? How well does it match the quantitative data of human development (Marcus, Pinker, Ullman, Hollander, Rosen & Xu, 1992)? Again, my point is that connectionist models need to be critically evaluated in order to understand *what* they do and *why* if they are to be of benefit to language acquisition research. Speculation about the possible capabilities of connectionist models on a general "conceptual" level does not explain SLA phenomena.

THE "SYMBOLIC" SIDE OF THE COIN

Let's approach this question from another angle, that of the symbolic "camp." Part of the work of an explanatory theory is to predict and explain what does *not* occur. Pinker and Prince (1988) raised many interesting problems for Rumelhart and McClelland's (1986) past tense model, which stimulated further research in this

area in both the symbolic and the connectionist circles. For example, why doesn't a person always apply the more frequent irregular past form to a new verb such as **Clinton landslid to victory,* on analogy with *The land slid* or *I hand-wrote the letter* (not *handwrited)?* Many people prefer *Clinton landslided to victory in the elections,* suggesting that they have a default rule that applies regular past tense to denominal verbs (Kim, Pinker, Prince and Prasada, 1992; Fantuzzi, 1993). The theoretical questions involve whether there are distinct lexical entries for words and whether people represent linguistic symbols such as 'noun' and 'verb' and have a categorical rule of 'affix' attachment. Also, are there distinct psychological mechanisms for representing rule-governed and rote-learned items? Some have suggested, for example Bybee (1991), that speakers might not represent regular past tense in English as a categorical rule but as a "schema" as in some connectionist models.[7]

Marcus, Brinkmann, Clahsen, Wiese, Woest and Pinker (1993) and Clahsen, Rothweiler, Woest and Marcus (1992) and others have pointed out that to test between the two hypotheses one needs a phenomenon that, unlike the regular past tense inflection in English, is both a default rule and infrequent in the input. Denominal verbs in English and German noun pluralization appear to provide us with a way to test the theories. My own experiment with native and non-native speakers of English (Fantuzzi, 1993), replicating Kim, Pinker, Prince and Prasada's (1992) studies with adult and child native speakers of English, found that both groups tended to regularize denominal verbs but not metaphorical verbs *(Federal agents ringed the compound /Gunshots rang throughout the night),* despite the infrequency of denominal verbs that are homophonous with irregular verbs in English. Even though second language learners should be doubly disposed to use frequent irregular forms due to explicit ESL instruction, they also tended to regularize denominal rather than semantically extended verbs, which points to the psychological representation of such constructs as noun and verb roots and a productive rule of affix attachment.

Another example is noun pluralization in German. Although it is highly irregular, Marcus et al (1993) list many contexts which suggest that *-s* is a "default" plural affix, although *-n* is more frequent: *-s* is the only affix that can appear in any morphophonological environment and *-s* occurs on names that are homophonous with nouns *(Manns),* onomatopoeic nouns

(Kuckucks), quoted nouns, nouns based on other grammatical categories such as conjunctions or verb phrases, truncations and acronyms, etc. (see Marcus et al. for discussion). Clahsen et al. (1992) provide empirical evidence for a correlation in children's speech between plural overregularization and its omission in compounds, similar to Gordon's (1985) finding that English-speaking children will accept *mice-eater* and *rat-eater* but not **rats-eater*. These studies of course provide support for a "dual route" model whereby irregular forms are stored in the lexicon and regular forms are created from a productive rule of affixation, as opposed to a single undifferentiated pattern-associating network.

Pinker and Prince (1988) proposed that verbs undergoing *irregular* past tense alternations in English need not be generated by a specific rule (for example, a rule such as Lowering Ablaut in the *sing/sang* alternation), but may be captured by a pattern associator such as Rumelhart and McClelland's (1986). This does not preclude the existence of "hard laws" as well, since the human mind is also able to override prototypes based on clusters of similar exemplars in memory (as in the *landslided* example). Pinker and Prince pointed out that the phonological connectionist model would still need to be embedded in a larger model with components for morphological, syntactic and semantic representations. A model of past tense acquisition must not only associate phonological past forms with their stems, but must also be able to represent the different argument structures of lexical items *(The ball flew out of his hand/The batter flied out to center field),* to apply a default regular rule to "denominal" verbs, to choose the correct morphological form for certain syntactic structures: *If I won a million dollars tomorrow . . .* and so on. They noted that the phonological model might thus be embedded in a collection of networks that reproduces the traditional account of linguistic modularity. Far from being a "paradigm shift", this suggests how connectionist models might be integrated with more traditional approaches.

Shirai and Yap, however, continue to insist that connectionism constitutes a *paradigm shift,* and even cite Pinker and Prince's "new approach" to morphology—rules for regular past and associative and rote memory for irregulars—as evidence. Shirai and Yap should be more careful with this term, since it is vague and has strong connotations. For Shirai and Yap, a paradigm shift appears to be just a "major change" in thinking, but research *generally*

progresses by changes in thinking. A paradigm shift, in contrast, is often understood to be a radical overthrow and supplantation of a previous paradigm (Kuhn, 1970). If Shirai does not associate himself with "radical eliminativism," he might avoid using the term "paradigm shift."[8] Even the commentators he cites explicitly warn against polarization of the two approaches (Schneider, 1988, p. 52; Clark, 1989, p. 83).

I pointed out in my previous article that connectionists working with complex problems of language often incorporate symbols into their architectures (e.g., Hinton, 1991). This could very well be the case with some of the models that Shirai and Yap cite as well, since what they mean by "connectionism" is not defined. We might settle this particular "debate" if we both agree that language is a complex and multi-faceted phenomenon and not handled completely by either "soft" or "hard" laws![9] My own position is similar to Pinker's, who has often commented that his criticisms are not against connectionism *per se,* but against *currently unsubstantiated* claims that language can be represented in a single pattern associating network without *any* representation of traditional linguistic symbols or operations at all (see Marcus et al, 1993).

CONCLUSION

As Seidenberg (in press) points out, the relationship between explanatory theories and computational models is "one of the hoariest issues in cognitive psychology" (p. 3), whether the computer simulation is connectionist or not. While most connectionists view connectionism as potentially contributing to a theory of general principles of cognition, they usually also admit, as Seidenberg does, that "we are not very far down the long road to the creation of wholly explanatory theories within this framework" (p. 14). While the analysis of connectionist systems is difficult and proceeds slowly, careful analyses of specific implementations, such as Pinker and Prince's critique of Rumelhart and McClelland's model, are valuable for stimulating further research. Shirai and Yap, however, argue for a more "qualitative" or conceptual approach, and state that "(b)ased on this speculative theorizing, we can then start actual network simulations to see whether our qualitative theoretical statements can actually be formalized/

quantified (p. 128)." There is a lot of work to be done between the vague idea and the formalization. Instead of arm-chair speculation about the future capability of models, it seems to me that SLA researchers who are truly interested in the applicability of connectionism to issues in SLA need to do some empirical research at this point. What Shirai and Yap offer is neither explanation nor theory, but only speculation at a very general level.

Much of the success of Chomsky's attack on behaviorism had to do with the specificity and testability of the theory of generative grammar. If connectionism is to have an impact on SLA research, I believe that it will need to do more than offer a vague and general framework for "making sense" of varied phenomena. Just as critical analysis of Chomsky's ideas fueled the rise of generative linguistics, it seems to me that the implementation and critical analysis of connectionist models, and not vague conceptualizations, are what is needed for its continued development. Most emphatically, I believe that general conceptualizations are *not* explanations of linguistic behavior, and that second language theorists must *not* be satisfied with vagueness in their explanations and theories. I do not agree that vagueness offers "elegance, consistency and 'making senseness'" or that vagueness has anything to do with connectionist theory-building at all.

ACKNOWLEGEMENTS

I thank John Schumann for many stimulating discussions of these issues and the editors of IAL for making this exchange of ideas possible.

NOTES

[1] The term "paradigm shift" is indeed vague, and thus part of our disagreement may be a confusion of terms. My critique concerned the strong use of the term, which suggests a radical change in research paradigms so that the new "paradigm" completely supplants the old, such as, for example, when generative grammar replaced behaviorist approaches to explaining linguistic behavior. Shirai (1992) in fact prominently presented connectionism as a "paradigm shift" comparable to the "Chomskyan Revolution" (p. 92). Many commentators, including myself, view the symbolism/connectionism dichotomy as creating an unnecessary polarization between the two approaches. Boden (1988), for example, discusses why symbolism and connectionism may not be separate paradigms so much as "feuding

cousins" within the same "family" of computational modelling. If Shirai and Yap do not view connectionism as a radical paradigm shift, then there is no disagreement between us. I stated in my article that I think both connectionist and symbolic models are useful for studying cognitive processing and that I view a polarization between the two approaches as divisive and unhelpful. I also stated that, in my opinion, connectionist models will probably never completely replace higher-level explanations, because I see a place for different levels of analysis, as Fodor and Pylyshyn (1988) argue.

[2] In defense of my criticism of Shirai's claims for the neural plausibility of connectionist models, Shirai and Yap state that at least connectionism strives for neural plausibility while traditional approaches to cognitive modelling "disregard" it and consider it "unimportant." This is incorrect, and simply trivializes the issues about cognitive architecture. As Feldman and Ballard (1982) put it, "The distributed nature of information processing in the brain is not a new discovery. The traditional view (which we shared) is that conventional computers and languages were Turing universal and *could be made to simulate* any parallelism (or analog values) which might be required" (p. 206, emphasis mine). They go on to say that "Most cognitive scientists believe that the brain appears to be massively parallel and that such structures can compute special functions very well. But massively parallel structures do not seem usable for general purpose computing and there is not nearly as much knowledge of *how to construct and analyze such models*. The common belief (which may well be right) is that there are *one or more intermediate levels of computational organization layered on the neuronal structure and that theories of intelligent behavior should be described in terms of these higher-level languages* ...We have not yet seen a reduction (interpreter if you will) of any higher fomalism which has plausible resource requirements, and this is a problem well worth pursuing" (p. 210, emphasis mine). Fodor and Pylyshyn (1988) present an in-depth discussion of the issue of levels of explanation in cognitive theory and the need for a "symbolic" level of representation. They certainly do not consider neural plausibility "unimportant" to theories of cognition. Indeed, they state that "understanding both psychological principles *and* the way they are neurophysiologically implemented is much better (and, indeed, more empirically secure) than only understanding one or the other. That is not at issue. *The question is whether there is anything to be gained by designing "brain style" models that are uncommitted about how the models map onto brains...the degree of relationship between facts at different levels of organization of a system is an empirical matter"* (p.62, emphasis mine).

[3] Although Shirai and Yap state that I "claimed" that "language involves higher-level functions which cannot be handled by connectionism" (p. 121), what I actually said was that "I disagree that connectionism can *as yet* explain *the high-level transfer phenomena that Shirai outlines in his article* " (p. 320, italics added). For example, Shirai used Munro's model of visual development to argue for a connectionist explanation of age-related effects on language transfer, and I pointed out that *if* this model could be applied to language acquisition it might correspond to phoneme recognition and not the higher-level transfer phenomena that Shirai outlined, since Munro explicitly said that his model involved only the earliest stages of processing. I also noted that Shirai might have paid more attention to Gasser's (1990) connectionist model of transfer in his discussion, which is a very simple model of language transfer and, again, cannot handle the sorts of high-level transfer phenomena that Shirai outlined. My point was not that Shirai's very general claims

were "incompatible" with Gasser's model, but that SLA researchers could benefit from a more in-depth discussion of an *actual model of language transfer* and of the theoretical issues that it raises. As a second language researcher who is interested in the applicablity of the models to SLA research, I would like to see more than vague speculation about non-existent models, and this was the source of my criticism of Shirai's discussion of transfer.

[4] Their explicit fusion of the terms *explanation/conceptualization* and *theory/framework* continue to make the claims vague, however.

[5] Fodor and Pylyshyn (1988: 57-58) state that "the notion that 'soft' constraints can vary continuously (as degree of activation does), are incompatible with Classical rule-based symbolic systems is another example of the failure to keep the psychological (or symbol-processing) and the implementational level separate. One can have a Classical rule system in which the decision concerning which rule will fire resides in the functional architecture and depends on varying magnitudes." Fodor and Pylyshyn argue that connectionism may be viewed as a theory of how (Classsical) cognitive systems can be implemented in "abstract neural" architecture.

[6] Even if we agreed, for argument's sake, that connectionism can provide an explanatory account of cognition, this does not necessarily mean that it will replace symbolic theories, as there are many competing theories in existence. Seidenberg notes that "the major differences between the approaches is that whereas Chomsky's principle claims concern types of knowledge representations and constraints thought specific to language, connectionists have focused on general mechanisms thought to apply across domains. In a complicated world, of course, *both could be correct*" (p. 22, footnote 3, emphasis mine).

[7] By categorical rule, I do not mean that speakers *always attach* the regular past tense morpheme to novel denominal verbs, only that speakers represent a categorical rule of affix attachment as well as linguistic categories of noun and verb. It is accepted that some extraneous factors may come into play in grammaticality judgements, not the least of which is the tendency for many people to be "prescriptively correct" and say *Clinton landslid to victory*. The connectionist model predicts that denominal verbs that are homophonous with irregular verbs will *never* be regularized, or at least not more often than merely semantically extended verbs. This prediction was not supported by my data. (See Kim et al., 1991, for discussion.)

[8] As I noted in my previous article, Shirai's passing mention of hybrid models in the last footnote of his article sharply contrasts with the prominent place he gives to a paradigm shift. While "paradigm shift" is a vague term, it does have strong connotations of a radical overthrow of the previous paradigm, and the general tone of Shirai's paper was skewed toward that interpretation. Although Shirai and Yap focus on my "claim" that Shirai is a radical connectionist, my critique was not actually against Shirai as a radical connectionist, or even against radical connectionism; it was about Shirai's *general, unsubstantiated claims* about connectionism and transfer. Shirai's claims are so vague that they may well be "compatible" with everything. I stated that "many critics of conventional AI as a model of human cognition see connectionism as a more neurally plausible glimpse into the "black box", and Shirai is clearly a proponent of this position. However, ... [he] *merely points to a vague connectionist framework to support this point of view*

(p. 325). This was the actual point I was making, and amply supported throughout my paper.

[9] Shirai and Yap make reference to a symbolic/connectionist debate, and characterize this exchange as part of that debate, but the issues of that "debate" are not clear. I hope that I have clarified in this reply that I do *not* consider this exchange to be part of a general connectionist/symbolic debate, and that my critique concerned Shirai's claim for a "connectionist explanation" of language transfer and *not* connectionist research in general. Hopefully, though, this exchange has also touched on some general issues that may provide more insight into both the "possibilities and limitations" of connectionist research for theories of SLA.

REFERENCES

Aronoff, M. (1976). *Word formation and generative grammar.* New York, NY: Cambridge University Press.

Boden, M. (1988). *Computer models of mind.* Cambridge: Cambridge University Press.

Bybee, J. (1991). Natural morphology: The organization of paradigms and language acquisition. In T. Huebner and C. Ferguson (Eds.), *Crosscurrents in Second Language Acquisition and Linguistic Theories.* Amsterdam: Benjamins.

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: Oxford University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Clahsen, H., Rothweiler, M., Woest, A. & Marcus, G. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition , 45*, 225-255.

Elman, J. L. (1991). *Incremental Learning, or the Importance of Starting Small.* (CRL Technical Report 9101). San Diego: University of California, Center for Research in Language.

Fantuzzi, C. (1992). Connectionism: Explanation or implementation? *Issues in Applied Linguistics, 3*(2), 319-340.

Fantuzzi, C. (1993). *The Acquisition of Verbal Morphology in a Second Language: Testing Competing Hypotheses.* Paper presented at the annual meeting of American Association of Applied Linguisics, Atlanta, Georgia.

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

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.

Gordon, P. (1985). Level-ordering in lexical development. *Cognition, 21*, 73-93.

Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. *Language, 51*, 639-671.

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

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions.* Chicago, IL: Chicago University Press.

Lachter J. and 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.

Lieber, R. (1980). *On the Organization of the Lexicon.* Unpublished doctoral dissertation, Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA.

MacWhinney, B. & Leinbach, H. (1990). Implementaions are not conceptualizations: Revising the verb-learning model. *Cognition, 40*, 121-157.

Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R. Woest, R. & Pinker, S. (1993). *German Inflection: The Exception that Proves the Rule.* Occasional Paper #47, Massachusetts Institute of Technology, Cambridge, MA.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J. and Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development , 57* (4, Serial No. 228).

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

Perlmutter, D. (1988). The split morphology hypothesis: Evidence from Yiddish. In M. Hammond and M. Noonan (Eds.), *Theoretical Morphology*. New York, NY: Academic Press.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition.* Hillsdale, NJ: Erlbaum.

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

Rumelhart, D.& McClelland, J. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, and The PDP Research Group (Eds.), *Parallel Distributred Processing: Explorations in the micro-structure of cognition,* Vol. 2. Cambridge, MA: Bradford Books/MIT Press.

Seidenberg, M. (in press). Connectionist models and cognitive theory. To appear in *Psychological Science.*

Schneider, W. (1988). Structure and controlling subsymbolic processing. *Behavioral and Brain Sciences, 11*, 51-52.

Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Shirai, Y. (1992). Conditions on transfer: A connectionist approach. *Issues in Applied Linguistics, 3*, 91-120.

Shirai, Y. & Yap, F. (1993). In defense of connectionism. *Issues in Applied Linguistics, 4*, 119-133.

Spencer, A. (1990). *Morphological Theory.* Cambridge, MA: Blackwell.

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.