DOCUMENT RESUME

ED 423 704 FL 025 522

AUTHOR	Jasdzewski, Gary
TITLE	How Useful Is Connectionism for SLA Research?
PUB DATE	1998-03-00
NOTE	17p.; Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (32nd, Seattle, WA, March 17-21, 1998).
PUB TYPE	Information Analyses (070) -- Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Artificial Intelligence; Cognitive Processes; Computer Oriented Programs; *Language Processing; *Language Research; *Learning Theories; Linguistic Theory; Neurolinguistics; Second Language Learning; *Second Languages
IDENTIFIERS	*Connectionism

ABSTRACT

The work of connectionist researchers is examined in order to understand better the implications for modeling second language learning processes. Connectionism is a biologically-oriented framework for understanding complex behavior, and provides a modeling tool (computer simulation) that behaves and learns without rules being explicitly wired into it. Its origins in cognitive science are traced to the 1950s, and its evolution within the field of artificial intelligence is reviewed briefly. It is noted that while there has been some discussion about parallel distributed processing (PDP) and its potential for understanding cognitive processes in the literature of second language acquisition (SLA), there has been little empirical work involving computer simulation of SLA. Several studies have addressed the utility of computer modeling for explaining some discrete SLA phenomena. Some arguments against connectionism are also found. It is concluded that if the more advanced PDP models can overcome some of the current problems and can allow predictions to be made about real second language learners, connectionism can be useful to SLA researchers. (Contains 26 references.) (MSE)

How useful is connectionism for SLA research?

Gary Jasdzewski

Purdue University

TESOL presentation, Seattle, March 18th, 1998

How useful is connectionism for SLA research?

For the last ten years researchers have attempted to apply connectionist or parallel distributed processing (PDP) models to second language acquisition phenomena. The goal of this research has been to see if these models can learn certain linguistic processes in a manner similar to actual second language learners without the use of explicit rules or symbols. If the models could accomplish this, it is argued, researchers would have a model that is closer to neurobiological reality than conventional symbol manipulation models like UG, and ultimately would allow us to predict and control the acquisition process to some extent. Early on, researchers in all of the cognitive sciences welcomed this line of thinking enthusiastically, for it seemed to offer an alternative to the existing symbolic paradigm. For example, in 1987 Sampson argued that PDP would lead to a paradigm shift as great as the one started by Chomsky's <u>Syntactic Structures</u>. Within the SLA field, Spolsky (1988) claimed that the "implications for second language learning theory are potentially immense" (393), and Schmidt 1988 wrote that we should expect some PDP inspired accounts of SLA in the future (63).

For people who are interested in the implications of neurocognitive research for second language theory and pedagogy, research of this kind is very interesting. For one thing, it shows that the SLA field is in touch with the intellectual mainstream since connectionism is a major development in the cognitive sciences and SLA is a cognitive science. Additionally, it is possible that SLA research involving PDP models could make contributions back to the cognitive science community, which would improve the status of the SLA field. Thus, this kind of work is pioneering and important. I believe, however, that this modeling tool could be-and in fact has been- useful for testing developmental hypotheses, but some caution is needed too. Now that the early excitement has abated, we can more objectively assess the status of this approach for the future. I think that it's safe to say that PDP modeling has contributed very little to our understanding of SLA. My purpose in this paper, therefore, is to examine the motivation of connectionist researchers so that we can better appreciate their efforts. Then I will briefly discuss what we have learned so far from this line of work.

What is connectionism?

Connectionism can be defined as a biologically oriented framework for understanding complex learning behavior. It is a modeling tool (computer simulation) that behaves and learns without rules being explicitly wired into it.

To understand the popularity of connectionism, it is necessary to look at the context surrounding the development of PDP in cognitive science. It has roots that extend back to the beginning of the cognitive revolution in the fifties and early sixties with the network models of Selfridge (1955) and Rosenblatt (1962) but it didn't catch on at that time for several reasons. First, computers were not nearly as powerful as they are now, and so it was difficult to implement complex networks on them. Additionally, there was the growing dissatisfaction with behaviorism and the rise of mentalism. In 1957, while the field of artificial intelligence was still very young, Chomsky published Syntactic Structures and Miller published his paper "The magical number seven, plus or minus two". These works were based on the belief that cognition was in essence the ability to manipulate symbols, that there were rules that the mind followed to reason. The basic assumption underlying most work in cognitive sciences at this time, including linguistics, is made explicit in the definition of cognition within artificial intelligence. According to the Physical Symbol Systems Hypothesis (PSSH), as advanced by Newell and Simon (1975):

> If a physical symbol system is any system in which suitably manipulable tokens can be assigned arbitrary meanings and, by means of careful programming, be relied on to behave in ways consistent with this projected semantic content, and if the essence of thought and intelligence is this ability to manipulate symbols, then any physical symbol system (such as the computer) can be organized to exhibit general intelligent action.

Such a system, if advanced enough, would be able to pass a Turing test; it should be able to respond to a person's questioning in such a way that that person could not tell if she were talking to a real person or a computer; it would be human in a sense. Note the characterization of thought and intelligence in these definitions: the essence of thought and intelligence is the ability to manipulate symbols. Elman, et al. call this the *first computer metaphor*.

Many people working in the cognitive sciences up into the eighties and many still today would claim that the PSSH is assumed in their field. That which cannot be represented symbolically might be relegated to the uninteresting pile, like performance has been in linguistics. Implicit in the idea is the distinction between the rule system and the instantiation of that rule system in some machine or body, the software and the hardware. This separation manifests itself in fields other than linguistics (such as philosophy and anthropology) in that the brain is often ignored in explaining cognition.

The artificial intelligence community was able to achieve a lot of success with the idea that thinking was all symbolic and rule based. Computers are particularly good at theorem proving and logical deduction, for example. We all know that they have infallible memories and they are fast. They are good at serially processing anything that can be formalized. They are now so good at chess that Deep Blue was able to defeat the highest rated chess player ever just this year, and the AI community considers this to be a landmark event in their field.

But despite success of this kind, artificial intelligence as a field has run up against a wall. It became increasingly clear in the 70's that there were certain types of cognitive processes which could not be modeled well via symbolic manipulation. Activities like pattern recognition, moving about in one's environment, speech and vision recognition, are very easy for even simple organisms to perform, but AI researchers have had tremendous difficulty getting machines to perform them because they are hard to represent symbolically and the computation time involved in performing activities of this type is enormous, making it impossible to have the machines perform them in real time.

This is why the first rule in AI is "The hard things are easy and the easy things hard." Presently, the most difficult problem for machine intelligence researchers is known as the "commonsense" or "background" or the "frame" problem . AI researchers now realize that commonsense knowledge plays a huge role in solving most tasks except those that are strictly defined (like games). Is this a problem of quantity or one of quality? This is what is being debated in the AI field. If it is the former, than larger and faster computers will solve the problem. This is the line that D. Lenat is taking in Texas. He believes that by entering in millions of facts into his computer's database, he can give his machine commonsense. Most commentators think the effort is futile. If the commonsense problem is one of quality, this would mean that commonsense knowledge is comprised of skills and capacities that are not representational. And if this is the case, then the PSSH is not applicable to much of human cognition.

So this is why AI critics can say that there hasn't been a significant advance in AI in 20 years. Deep Blue defeated Kasparov not because its IBM programmers had conceived of a theoretical innovation, a new way to represent knowledge. It succeeded because it had more processors and a larger data base; it just manipulated more symbols than its predecessors. And since we know that humans do not play chess in the same way that computers do, this landmark event in AI is relatively unimportant to the rest of the cognitive sciences.

Given the general dissatisfaction with the PSSH in the cognitive sciences, along with the development of faster computers and new learning algorithms (like back propagation), when connectionism reappeared in the 80's, there were a lot of people ready for an alternative to the symbolic manipulation paradigm. Connectionist models, although still computational (Elmann, et al., call it the *second computer metaphor*), differ from symbolic models in several ways. They are empiricist and analogical rather than rational and digital. They do not require someone to wire in the rules of cognition, rather, they seem to come up with them on their own. They are good at pattern recognition. They come closer to obeying the 100 step constraint required to perform functions in real time. They fill in missing parts of noisy signals much like living organisms. They appear to generalize from experience. They exhibit graceful degradation so that if part of the network is destroyed the network can continue to function (unlike UNIX or DOS!). They are built of simple units that are interconnected and memory seems to be not in one place but spread out. For all of these reasons, connectionist models are a lot like the brains of animals. It is no wonder that in every cognitive science, including SLA, researchers looked to apply this approach to their object of study.

<u>How is it related to SLA? What have we learned?</u>

So the question we may ask at this point is how has connectionism influenced the SLA field, and how has it contributed to our understanding of the SLA process? A survey of the literature reveals that while there has been a moderate amount of discussion about PDP and its potential for understanding cognitive processes, there is relatively little empirical work involving computer simulations (less than a handful), a situation which I find to be somewhat surprising, and which Carrol (1995) calls an "odd fact" considering the early optimism.

More common in the literature than simulations are discussions about the potential of PDP models to explain some SLA phenomena. Sokolik (1990), for example, describes how PDP models contribute to an understanding of the Adult Language Learner Paradox, which holds that adults should be better L2 learners than children given their more developed cognitive abilities, but they aren't. According to Sokolik, when viewed from the perspective of PDP, there is no paradox because the adult brain is less plastic than the child's brain in the sense that there is a reduction in the amount of modifiability of the connections between neurons. In a similar way, once a network is trained, it becomes more difficult to modify it (But Long (1988) presents evidence that adults learn faster earlier). In another paper, Shirai (1992) explores transfer in light of PDP and claims that the interconnectedness of the units lend themselves to a PDP explanation, and this is supported by an empirical study by Gasser (1990).

The literature also contains arguments against connectionism. Fantuzzi (1992) and Carrol (1995) do especially good jobs of developing these arguments, in that they do not only cite the arguments of thinkers from outside the field, especially the early criticisms of connectionism from Pinker and Prince, and Fodor and Pylyshyn, but also because they apply it to SLA. Fantuzzi comes to the conclusion that we are in no position at this time to say that PDP has or will eliminate higher level explanations, and I agree. Here is a summary of issues that have been raised that remain unresolved:

First there is the charge that connectionism is in essence behaviorism or associationism and so the arguments against associationism should apply to connectionism as well. They are certainly similar on the surface: stimulus-response, etc. Jerry Fodor still maintains that PDP cannot account for the rationality of thought, and is quoted as saying that "people will give it up for the reasons they gave up associationism" (cited in Baumgartner & Payr, p. 94). Even proponents of PDP like Elman, et al., question the ability of networks to handle higher level cognition. How, for instance, does implicit knowledge become explicit? How does a network build a theory? Where does awareness emerge? Some supporters of PDP will say that awareness that awareness is an emergent property of our billions of neurons, and this is obviously true. But our networks, which have perhaps 10^3 connections compared to 10^11 in our brains, have a long way to go before they become conscious! This may seem esoteric but it is a real problem for SLA. L2 learners are strongly guided by conscious strategies. (The reply to this from Nick Ellis, an SLA researcher who is using connectionist models, is that PDP is concerned with what goes on in the black box; they are concerned with representations. But in what sense? With several hidden layers programmers can't tell what represents what, and even if they could point to some representation, so what? How would this *explain* anything?)

Related to this is the issue of ecological plausibility. Real L2 learners don't learn past tense of Monday, plurals on Tuesday, etc. People are active with agendas, models are passive. People are social, models not. Even a good simulation of a cognitive process will neglect this social dimension.

PDP can also be attacked for one of its touted strengths--its neural plausibility. They are like brains in the ways mentioned previously, but the learning algorithm of back propagation is not. Back prop is a way for a network to produce accurate outputs, to learn. If there is a mismatch between the desired output and actual output, the network tries it again until a better match is made. But Dreyfuss (cited in Baumgartner & Payr, p. 78) points out, back prop can't count as a theory of neuroscience because "everyone knows the brain's not wired that way." He does admit that some undiscovered algorithms might work, and if they are discovered, then we may be onto something. Additionally, it has been pointed out that PDP moels don't really meet the 100 step constraint needed to accomplish the computations that our brains must make in real time (Baumgartner & Payr, p. 108).

These problems may not be insuperable, but at this time they are real, and so we will have to wait for better modeling. If humans don't learn in the same way that PDP models do, then it seems the whole point of doing them is lost. As Carrol (1995) puts it, "Computer modeling experiments which misrepresent the nature of the learning problem or of the linguistic input will teach us nothing" (p. 204).

What have we learned from models that pertain to SLA? Not much. Gasser (1990) trained a network to generate L1 and L2 sentences. He reports that his model was able to generate sentences without any rules being wired in, and that in multilingual contexts the model exhibited strong transfer effects. When the network couldn't find the appropriate L2 item it borrowed a similar one from the L1. Also, the L2 patterns were easier to learn if the words and word order of the two languages were similar.

Another frequently discussed study was conducted by Sokolik and Smith (1992), who investigated the extent to which a PDP model could assign the correct gender to French nouns that it had not seen before (apparently difficult problem for L2 learners). Using back propagation, the machine was able to correctly classify a high percentage of nouns it had not seen before. It is amazing that these ignorant devices can do this kind of thing! How useful these studies are I'll leave up to you to decide.

More recently, Ellis and Schmidt (1997) have used a connectionist simulation in a way that I think is interesting and novel. They conduct a pair of what they call 'laboratory' studies of the acquisition of L2 morphological abilities with an artificial langauge they created. They used an artificial language so that they could control the input as much as possible. After charting the learning progress of seven human subjects as they tried to pluralize made up words, they trained a network to do the same thing. Interestingly, the network produced the same learning curve as the human beings. The authors conclude that this aspect of learning a second language reflects associative learning processes. While it doesn't prove that L2 learners learn to pluralize without rules, it doesn't contradict the idea that it is an associative process. Regardless, what I find interesting about this study is the use of a network as a corroborating line of evidence for lab work involving human subjects. But overall, the amount of empirical work is not overwhelming, and the results do not seem particularly useful just yet. As Klein (1990) so eloquently put it: "It is one thing to build a functioning clock and another a theory of time."

Conclusions

In conclusion, it looks like the early optimism of SLA researchers for PDP approaches was appropriate given the shortcomings of the physical symbol systems hypothesis for the cognitive sciences. Clearly the first computer metaphor can only describe a part of human cognition. Even so, the second computer metaphor has contributed little to our understanding of SLA. Perhaps its greatest value is that it is providing a challenge to nativist accounts of language acquisition which rely on the first computer metaphor. If in the future more advanced PDP models can overcome some of the current problems, and if these models can allow us to make predictions about real L2 learners, then connectionism will be useful to SLA researchers.

References

Baumgartner, P., & Payr, S. (Eds.). (1995). Speaking minds. Princeton, NJ: Princeton UP.

Carrol, S. (1995). The hidden dangers of computer modeling: Remarks on Sokolik and Smith's connectionist learning model of French gender. Second Language Research, 11(3), 193-205.

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

Ellis, N. (1996a). Sequencing in SLA: Phonological memory, chunking, and points of order. Studies in Second Language Acquisition, 18, 91-126.

Ellis, N. (1996b). Analyzing language sequence in the sequence of language acquisition: Some comments of Major and Ioup. Studies in Second Language Acquisition, 18, 361-368.

Ellis, N, & Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. Studies in Second Language Acquisition, 19, 145-171.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking Innateness: A connectionist perspective on development. Cambridge, MA: A Bradford Book.

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

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

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

Hatch, E., Shirai, Y., & Fantuzzi, C. (1990). The need for an integrated theory: Connecting modules. TESOL Quarterly, 24(4), 697-716.

Klein, W. (1990). A theory of language acquisition is not so easy. Studies in Second Language Acquisition, 12, 219-231.

Lee, C., & Gasser, M. (1991). Learning morphonemic processes without underlying representations and rules. Language Research, 27(2), 303-317.

McClelland, J., Rumelhart, D., & the PDP Research Group (Eds.) (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models. Cambridge, Mass.: MIT Press.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

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

Rosenblatt, F. (1962). Principles of Neurodynamics. New York: Spartan.

Sampson, G. (1987). Review article: Parallel distributed processing. Language, 63, 871-886.

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

Selfridge, O. (1955). Pattern recognition in modern computers. Proceedings of the Western Joint Computer Conference.

Shirai, Y. (1992). Conditions of transfer: A connectionist approach. Issues in Applied Linguistics, 3(1), 91-120.

Shirai, Y., & Yap, F. (1993). In defense of connectionism. Issues in Applied Linguistics, 4(1), 119-133.

Sokolik, M. E. (1990). Learning without rules: PDP and a resolution of the adult language learning paradox. TESOL Quarterly, 24, 685-96.

Sokolik, M.E. & Smith, M. (1992). Assignment of gender to French nouns in primary and secondary language: A connectionist model. Second Language Research, 8(1), 39-58.

Spolsky, B. (1988). Bridging the gap: A general theory of second language learning. TESOL Quarterly, 22(3), 377-396.

Yap, F., & Shirai, Y. (1994). Exchange: On the nature of connectionist explanations. Issues in Applied Linguistics, 5(1), 173-194.

FL025522

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: HOW USEFUL IS CONNECTIONISM FOR SLA RESEARCH?

Author(s): GARY JASDZEWSKI

Was this a TESOL presentation? ✓ yes ___ no If not, was it another conference presentation? Specify: ___________

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: [signature]	Printed Name/Position/Title: GARY JASDZEWSKI	
Organization/Address: PURDUE UNIVERSITY W. LAFAYETTE, IN 47906	Telephone: 618 687-4681	FAX:
	E-Mail Address: garyjaz@omni.cc.purdue.edu gary@siu.edu	Date: 8-23-98

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on
Languages & Linguistics
1118 22nd Street NW
Washington, D.C. 20037

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com