Connectionist Models: Implications in Second Language Acquisition

Farid Ghaemi
Islamic Azad University, Karaj Branch, Department of Humanities, Iran.
ghaemi@kiau.ac.ir

Laleh Fakhraee Faruji
Islamic Azad University, Science and Research Branch, Department of Literature and Foreign Languages, Tehran, Iran
fakhraeelaleh@yahoo.com

Abstract

In language acquisition, 'Emergentists' claim that simple learning mechanisms, of the kind attested elsewhere in cognition, are sufficient to bring about the emergence of complex language representations (Gregg, 2003). Connectionist model is one of the models among others proposed by emergentists. This paper attempts to clarify the basic assumptions of this model, its advantages, and the criticisms leveled at it.

Keywords: Associationism, Connectionism, Emergentism, Information Processing (IP)

1. Introduction

'Emergentism' is the name that has recently been given to a general approach to cognition that stresses the interaction between organism and environment and that denies the existence of pre-determined, domain specific faculties or capacities (Gregg, 2003). Emergentism offers itself as an alternative to modular, 'special nativist' theories of the mind, such as theories of Universal Grammar (UG). One of the models among others proposed by emergentists is called 'connectionism'. Richards & Schmidt (2002) defined connectionism as a theory in cognitive science that assumes that the individual components of human cognition are highly interactive and that knowledge of events, concepts and language is represented diffusely in the cognitive system (p. 108). According to them this theory has been applied to models of speech processing, lexical organization, and first and second language learning.

Connectionism provides mathematical models and computer simulations that try to capture both the essence of information processing and thought processes. The basic assumptions of the theory are (Richards & Schmidt, 2002, p. 108):

1. Information processing takes place through the interactions of a large number of simple units, organized into networks and operating in parallel.
2. Learning takes place through the strengthening and weakening of the interconnections in a particular network in response to examples encountered in the input.
3. The result of learning is often a network of simple units that acts as though it "knows" abstract rules, although the rules themselves exist only in the form of association strengths distributed across the entire network.

Connectionist approaches to learning have much in common with Information Processing (IP) perspectives (Saville-Troike, 2006, P. 80), but they focus on the increasing strength of associations between stimuli and responses rather than on the inferred abstraction of "rules" or on restructuring. Indeed, from a connectionist perspective learning essentially is change in the strength of these connections. Some version of this idea has been present in psychology at least since the 1940s and 1950s, but Connectionism has received widespread attention as a model for first and second language acquisition only since the 1980s.

Ahlsén (2006, p. 10) called connectionism as associationism, assumes in which higher functions are dependent on the connections between different centers in the cortex. Linguistic ability is seen as the relationship between images and words. He believed that aphasia results from broken connections between the centers that are needed for linguistic function. Connectionist

models challenges the traditional separation of item memory and abstract, symbolic rule systems, and provides an empirically testable alternative to nativist approaches to learning (Williams, 2005).

Gass & Selinker (2008) provided a premise and a metaphor for connectionism as (p. 494):

- *Premise*: Emphasizes environment influence and exposure to language towards a gradual buildup of language.
- *Metaphor:* The builder needs the assistance of others to help construct the house; he cannot do it entirely on his own.

An early connectionist model was a network trained by Rumelhart and McClelland (1986, as cited in Jordan, 2004, p. 243) to predict the past tense of English verbs. The network showed the same tendency to overgeneralize as children, but there is still no agreement about the ability of neural networks to learn grammar.

2. Elements of connectionist models

Connectionist models consist of a large number of simple processing units with multiple connections linking them. Activation flows along the connections, just as electrical impulses transmit information through neurons in the brain. The ease with which activation spreads from one unit to another is determined by the strength of the connections along which it travels. The stronger the connection to a unit, the more readily that unit becomes activated. A connection's strength depends upon how frequently it is used. Thus, over time, connections to a frequent word will become strong, ensuring that the word is activated more rapidly than other less common ones (Field, 2004, p. 73).

According to Field (2004, p. 73) strength of the connectionist approach is that, besides modeling processes such as word recognition, it can also model learning. In a computer simulation, each connection receives a number or weight to indicate its relative strength. At the outset, these can be set at 0; but, as connections are used, their weights are adjusted by means of a complex formula. If a connection is not used, its weight declines to a negative value, indicating an inhibitory relationship. The effectiveness of this learning process has been increased by a feedback mechanism known as back propagation which provides the program with a kind of memory. It compares what a connectionist network outputs from a particular stimulus with what it should output.

2.1. Backpropogation

By dint of many repeated presentations of the input, some connections within the network become strengthened while others become weakened. In this way, the network can gradually be 'trained' to produce correct responses through a process of error reduction (Field, 2004, p. 75).

Using back propagation, a connectionist program has managed to simulate the acquisition of a set of regular and irregular English Past Simple verb forms. It succeeded in discriminating between cases where an -ed inflection was appropriate (walk-walked) and those where a new form had to be learnt (write-wrote). It also, in the process, manifested the kind of U-shaped development observed both in first and second language acquisition, where a speaker acquires a correct irregular form, then later replaces it with a regular form that has been over-generalised (writed). Simulations such as this are sometimes cited in support of an empiricist view of language acquisition. They suggest that linguistic patterns can be identified through exposure to multiple examples, with no need to presuppose a genetically transmitted mechanism which drives the acquisition process (Field, 2004, p. 75).

2.2. Parallel Distributed Processing (PDP)

Network performance in PDP is determined by several main factors (MacDonald & Seidenberg, 2006, P. 587): (1) the architecture of the system (e.g., the configuration of units and connections); (2) the characteristics of the input and output representations; (3) characteristics of the patterns used in training the model; and (4) characteristics of the learning algorithm. In other words,

the model's performance depends on its initial state, what it experiences, and how it learns from those experiences.

MacDonald & Seidenberg (2006) focused on three properties that inform the probabilistic constraints approach to comprehension as follow (p. 588):

First, the networks incorporate a theory of statistical learning. The main idea is that one way that people learn (there may be others) is by gathering information about the frequencies and distributions of environmental events. This type of learning is thought to be general rather than language specific. The reading models in particular developed the idea that lexical knowledge consists of statistical relations between orthographic, phonological, and semantic codes. Learning then involves acquiring this statistical knowledge over time. Subsequent research on statistical learning in infants and adults has provided strong evidence consistent with this view. This learning mechanism provides a way to derive regularities from relatively noisy data, a property that is likely to be highly relevant to the child's experience in learning language.

Second, the models provide a basis for understanding why particular types of statistics are relevant and not others. Given the properties of these representations, other aspects of the model architecture (e.g., number of units or layers patterns of connectivity between layers), and a connectionist learning algorithm, the model will pick up on particular statistical regularities implicit in the examples on which the model is trained. Thus, motivating the various elements of a model and how it is trained is very important, but the model itself determines which statistics are computed.

Third, the framework provides a powerful processing mechanism, the exploitation of multiple simultaneous probabilistic constraints. Information in a network is encoded by the weights. The weights determine ("constrain") the output that is computed in performing a task.

Processing involves computing the output that satisfies these constraints. This output changes depending on what is presented as input (e.g., the current word being processed in a network that comprehends sentences word by word). This type of processing, known as constraint satisfaction, has several interesting properties. One is that the network's output is determined by all of the weights. Such models illustrate how a large number of constraints can be utilized simultaneously without imposing excessive demands on memory or attention. Constraint satisfaction is passive – activation spreads through a network modulated by the weights on connections – rather than a resource-limited active search process. Another important property is that the constraints combine in a nonlinear manner. Bits of information that are not very informative in isolation become highly informative when taken with other bits of information. Much of the power and efficiency of the language comprehension system arises from this property. Languages exhibit much partial regularity. Different types of information are correlated, but weakly. The comprehender cannot wholly rely on any one type of information, but combinations of these partial cues are highly reliable. This concept may seem paradoxical at first. If individual cues are unreliable, wouldn't combinations of these cues be even more unreliable? MacDonald & Seidenberg (2006, p. 587) answers: no, not if cue combination is nonlinear. The informativeness of each cue varies as a function of other cues.

3. Connectionism and SLA

Even though connectionist approaches have been around for a number of years, it is only recently that research within a second language context has begun to take place (Gass & Selinker, 2008, p. 220). Gasser (1990) examines the implications of connectionist models of cognition for second language theory. Connectionism offers a challenge to the symbolic models which dominate cognitive science. In connectionist models all knowledge is embodied in a network of simple processing units joined by connections which are strengthened or weakened in response to regularities in input patterns. These models avoid the brittleness of symbolic approaches, and they exhibit rule-like behavior without explicit rules.

In SLA the assumption is that transfer from L1 to L2 occurs because strong associations already established in L1 interfere with establishment of the L2 network. Because frequency is the

primary determinant of connection strength, it might be predicted that the most common patterns in L1 would be the most likely to cause interference in L2, but research on transfer from linguistic perspectives does not support this conclusion in any strong sense; L1-L2 relationships are not that simple. Proponents of connectionist approaches to language acquisition note that while frequency is "an all-pervasive causal factor" (Ellis 2002, as cited in Saville-Troike, 2006, P. 81), it interacts with other determinants, including how noticeable the language patterns are in the input learners receive, and whether the patterns are regular or occur with many variations and exceptions.

According to Saville-Troike (2006) the best-known connectionist approach within SLA is Parallel Distributed Processing, or PDP (P. 80). Based on this viewpoint, processing takes place in a network of nodes (or "units") in the brain that is connected by pathways. As learners are exposed to repeated patterns of units in input, they extract regularities in the patterns; probabilistic associations are formed and strengthened. These associations between nodes are called connection strengths or patterns of activation. The strength of the associations changes with the frequency of input and nature of feedback.

The claim that such learning is not dependent on either a store of innate knowledge (such as Universal Grammar) or rule-formation is supported by computer simulations. For example, Rumelhart and McClelland (1986, as cited in Saville-Troike, 2006, P. 80) demonstrated that a computer that is programmed with a "pattern associator network" can learn to associate English verb bases with their appropriate past tense forms without any *a priori* "rules," and that it does so with much the same learning curve as that exhibited by children learning English L1 (Saville-Troike, 2006, P. 80).

The model provides an account for both regular and irregular tense inflections, including transfer to unfamiliar verbs, and for the "Ushaped" developmental curve which is often cited in linguistic models and in other cognitive approaches as evidence for rule-based learning. Assumptions about processing from a connectionist/PDP viewpoint differ from traditional IP accounts in other important ways. For example (McClelland, Rumelhart, and Hinton 1986; Robinson 1995, as cited in Saville-Troike, 2006, P. 80):

  i.   Attention is not viewed as a central mechanism that directs information between separate memory stores, which IP claims are available for controlled processing versus automatic processing. Rather, attention is a mechanism that is distributed throughout the processing system in local patterns (p. 81).

  ii. Information processing is not serial in nature: i.e. it is not a "pipeline . . . in which information is conveyed in a fixed serial order from one storage structure to the next" (Robinson 1995 as cited in Saville-Troike, 2006, P. 81). Instead, processing is parallel: many connections are activated at the same time.

  iii. Knowledge is not stored in memory or retrieved as patterns, but as "connection strengths" between units which account for the patterns being recreated.

It is obvious that parallel processing is being applied when tasks simultaneously tap entirely different resources such as talking on a cell phone while riding a bicycle, but it also less obviously occurs within integrated tasks such as simply talking or reading, when encoding/decoding of phonology, syntactic structure, meaning, and pragmatic intent occur
simultaneously (Saville-Troike, 2006, P. 81). Many connections in the brain must be activated all at once to account for successful production and interpretation of language, and not processed in sequence (i.e. one after the other).

4. Advantages of connectionist models

Connectianist models appear to offer a number of advantages (Feldman & Ballard, 1982). Ellis (2003, as cited in Doughty & Long, 2003, p. 85) believed that a detailed transition theory is needed for describing the process of language acquisition. If language is not informationally produced in its own module, then we must eventually show how generic learning mechanisms can result in complex and highly specific language representations. We need dynamic models of the

acquisition of these representations and the emergence of structure. And we need processing models where the interpretation of particular utterances is the result of the mutual satisfaction of all of the available constraints. For these reasons, emergentists look to connectionism, since it provides a set of computational tools for exploring the conditions under which emergent properties arise (Ellis, 2003, as cited in Doughty & Long, 2003, p. 85).

According to Ellis (2003, as cited in Doughty & Long, 2003) connectionism has various advantages for this purpose: neural inspiration; distributed representation and control; data-driven processing with prototypical representations emerging rather than being innately pre-specified; graceful degradation; emphasis on acquisition rather than static description; slow, incremental, non-linear, content- and structure-sensitive learning; blurring of the representation/learning distinction; graded, distributed, and non-static representations; generalization and transfer as natural products of learning; and, since the models must actually run, less scope for hand-waving (Ellis, 2003, as cited in Doughty & Long, 2003 p. 85).

Connectionist studies are important in that they directly show how language learning takes place through gradual strengthening of the associations between co-occurring elements of language, and how learning the distributional characteristics of the language input results in the emergence of rule like, but not rule-governed, regularities. Given that connectionist models have been used to understand various aspects of child language acquisition, the successful application of connectionism to SLA suggests that similar mechanisms operate in children and adults, and that language acquisition, in its essence, is the distributional analysis of form-function mappings in a neural network that attempts to satisfy simultaneously the constraints of all other constructions that are represented therein (Ellis, 2003, as cited in Doughty & Long, 2003 p. 91).

Traditional symbolic models (taking the classical approach to categorization) work on the all-or-none basis and can satisfy *hard constraints* only. That is, even if one condition is not met, a rule does not apply.

A connectionist system can satisfy soft constraints, meaning that it finds the best solution to a situation where multiple constraints compete by meeting as many of them as possible, even if none of the conditions are met completely. There is one more general point to be made. Being motivated by the recognition that the brain is a neural network, connectionism equates mental representations with patterns of neural activity, which what connectionism shares with enactivism, both present a clear contrast to traditional cognitivism in this respect (Zalewski, 2010, p. 95).

5. Criticisms of connectionist models

According to Gregg (1996) nonmodular approaches in L2 acquisition research has been largely motivated by a misguided concern for communicative competence, there has been little in the way of explicit nonmodular theorizing. Instead, there has been a more or less implicit denial of modularity: Learning is seen as a general process irrespective of object. That is, it is often assumed by L2 acquisition theorists that such processes as hypothesis testing, generalization, analogy, automatization, and so forth, apply equally to any learning task-linguistic or otherwise (McLaughlin, 1987, as cited in Gregg, 1996, p. 58 [6]).

One counterargument of strong deterministic role of frequency of input in language learning is that some of the most frequent words in English (including the most frequent, *the*) are relatively late to appear, and among the last (if ever) to be mastered. Still, whatever one's theoretical perspective, the effects of frequency on SLA clearly merit more attention than they have typically received since repetition drills went out of fashion in language teaching. Researchers from several approaches to SLA which focus on learning processes are taking a renewed look at how frequency influences learning (Saville-Troike, 2006, P. 81 [10]).

Connectionists take a very different antimodular approach to language learning which is away from rules, structures, and so on, and instead see learning as the relative strengthening (or spreading activation) of associations, or connections, between interconnected units or nodes. The rules or principles proposed by linguists are simply epiphenomena, in this view; they have no non metaphoric existence whatever, and thus play no role in language learning. The fact-if indeed it is a

fact-that one can successfully model the acquisition of English past-tense forms with a connectionist computer program does not say anything as to whether the human mind goes about acquiring in that way; nor is a spreading activation model necessarily incompatible with a more traditional concept of mental structure (J. A. Fodor & Pylyshyn, 1988, as cited in Gregg, 1996, p. 59 [6]).

Further, as Carroll and Meisel (1990 as cited in Gregg, 1996, p. 59 [6]) pointed out, connectionist accounts fail to deal with the fact that humans have knowledge that transcends the input; but that, of course, is the very heart of the logical problem. Something like spreading activation could possibly be involved in the establishing of connections between, say, irregular verbs and their past-tense endings; but we cannot appeal to a *lack* of activation for our knowledge that one sentence (e.g., *She may have been being misled)* is a possible sentence of English whereas another (e.g., *She may been have being misled)* is impossible. After all, one have probably never heard either type of sentence, and thus they are equally unlikely to have been activated. Given the serious shortcomings of the connectionist model, it is hard to see any good reason for taking it up in the stead of a structured system of mental representations, as assumed by almost any other view of the mind.

It is true; of course, that the mind is instantiated in the brain, and that any mental activity whatever must have a neurological basis. But it by no means follows from this that the neurological level is an appropriate, let alone the only proper, level at which to account for mental activity (Gregg, 1996, p. 59).

What otherwise widely different theories such as O'Grady's "general nativism" and connectionism have in common is a reductionist animus, an attempt to account for language acquisition phenomena as nothing more than special cases of other kinds of phenomena(Gregg, 1996, p. 59)

6. Concluding remarks

Connectionism now offers a radical alternative to innativism view. The old questions about innateness, the mind and the body, and what it means to know are being asked again, and a new set of answers is being proposed.

Not everyone agrees that connectionist models make good neuropsychological sense (Segalowitz and Bernstein, 1997, as cited in Segalowitz & Lightbown, 1999 [11]). However, as Ellis (1999 [2]) has suggested, one does not necessarily have to view connectionism as an attempt to model neural architecture for it to be useful (as cited in Segalowitz & Lightbown, 1999 [11]).

Connectionist analysis can, instead, be understood as a mathematical tool that reveals how the process of simple associative learning can contain hidden within it the basis for establishing response patterns that appear to be rule-governed in the absence of explicit rule representation. This use of connectionism as a tool is analogous to, say, the way one uses Fourier analysis to analyze complex wave forms to reveal underlying hidden, simpler acoustic patterns (e.g., the component frequencies, formant structures of speech)(Segalowitz & Lightbown, 1999 [11]).
How the brain actually performs the equivalent analysis is another question.

References

[1] Ahlsen, E. (2006). *Introduction to neurolinguistics.* Amsterdam: John Benjamins Publishing Company.
[2] Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63-104). Oxford: Blackwell Publishing.
[3] Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205-254.

[4] Field, J. (2004). *Psycholinguistics: The key concepts*. London & NewYork: Routledge. Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). New York & London: Routledge.

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

[6] Gregg, K. R. (1996). The logical and developmental problems of second language acquisition, in T. K. Bhatia & W. C. Ritchie (Eds.), *Hand book of second language acquisition* (pp. 50-75). London: Academic Press, Inc.

[7] Jordan, G. (2004). *Theory construction in second language acquisition*. Amsterdam & Philadelphia: John Benjamins Publishing Company.

[8] MacDonald, M. C., & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Hand book of psycholinguistics* (pp. 581-613). London: Elsevier Inc.

[9] Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). London: Longman.

[10] Saville-Troike, M. (2006). *Introducing second language acquisition*. Cambridge: Cambridge University Press.

[11] Segalowitz, N., & Lightbown, P. M. (1999). Psycholinguistic approaches to SLA. *Annual Review of Applied Linguistics, 19*, 43–63.

[12] Williams, J. N. (2005). Associationism and connectionism. Retrieved November 2010 from people.pwf.cam.ac.uk/jnw12/associationism.pdf

[13] Zalewski, J. (2010). A connectionist–enactivist perspective on learning to write. In J. Arabski & Wojtaszek (eds.), *Neurolinguistic and psycholinguistic perspectives on SLA* (pp. 93-106). Toronto: Multilingual Matters.