Symbolic vs. Connectionist Accounts of SLA

Mohammad Reza Yousefi Halvaei

Department of English, Islamic Azad University, Bonab Branch, Bonab, Iran

Musa Moradi

Department of English, Islamic Azad University, Bonab Branch, Bonab, Iran

Mohammad Hossein Yousefi

Department of English, Islamic Azad University, Bonab Branch, Bonab, Iran

Abstract—Generative Linguists following (Chomsky 1965, 1982, and 1995) have argued that grammar is innate, exist in brain as domain-specific module, and is transmitted by genetic inheritance. They also argued for rule-governed nature of language and language acquisition. They have resorted to many arguments to justify these claims among which, the complexity of language, the poverty of stimulus and the lack of negative evidence can be mentioned (Cook and Newson, 1996). For some decades these theories have been widely accepted as being not controversial and even undeniable. But within the last two decades these ideas have been strongly disputed by emergentists, construction grammarians, associationists, and connectionists. These approaches differ strikingly from other accounts of language learning. They do not believe that language acquisition is the result of internalizing language rules. Instead, in these approaches, the importance is put on construction of associative patterns (Mitchell & Myles 2004). Among these approaches to language the last one, connectionism, is greatly distinguished by others in its research techniques. The development of neural network computer simulations or what has come to be known as Artificial Neural Networks (ANN) has helped researchers in this approach to make stronger claims about the nature of language and language acquisition. This has helped them to move from making abstract and obscurant theories toward entangling with concrete and physical realities. The present paper is an attempt to compare and contrast the symbolic and connectionist approaches to second language acquisition.

Index Terms—symbolic approaches, connectionism, emergentism

I. INTRODUCTION

Hulstijen (2002) explains that symbolic accounts represent knowledge as a collection of symbols accompanied by rules that specify the relationship between them. According to connectionist account, knowledge is represented as sums of tiny information-packed units but rather activation patterns in a neural network.

There has been a wide research interest towards the connectionist approaches to second language acquisition (Ellis N. 1998; Elman 2005; Elman et al., 1996; Ingram, 2007; MacWhinney, 1998; Ingram, 2007; Rumelhart and McClelland (1986). According to Connectionism, the brain is like a computer that consists of neural networks. Learning in this view occurs on the basis of associative processes, rather than the construction of abstract rules. Connectionist approaches to language acquisition explore the representations that can result when simple learning mechanisms are exposed to complex language evidence (Ellis N. 1998, p.645). There are many separate connectionist simulations of the learning of morphology, phonological rules, semantic structures, etc. These stimulations demonstrate that connectionist models can extract the regularities in each of these domains of language and then operate in a rule-like (but not rule-governed) way (Ellis, N. 1998).

Emergentism

Emergentists contend that the innateness assumption of the language instinct hypothesis lacks any plausible process explanation (Elman et al., 1996). It is argued that today's theories of brain function; process and development do not verify the inheritance of structures which might serve as principles and parameters of UG (Ellis, N. 1999). In the Emergentist perspective, interactions occur at all levels, from brain to society, give rise to emergent forms and behavior (Elman et al. 1996; MacWhinney, 1998). These outcomes might be highly constrained and universal, but they are not directly contained in the genes in any domain specific way (Ellis, N. 1999).

Emergentists look at interdisciplinary nature of language. They argue that a complete understanding of language is not going to come from one discipline alone (Ellis, N, 1999). As Cook and Seidlhofer (1995) summarize, language can best be viewed as:

a genetic inheritance, a mathematical system, a social fact, the expression of individual identity, the expression of cultural identity, the outcome of a dialogic interaction, a social semiotic, the intuitions of native speakers, the sum of attested data, a collection of memorized chunks, a rule-governed discrete combinatory system, or electrical activation in a distributed network ... We do not have to choose. Language can be all of these at once. (Cook & Seidlhofer, 1995, p. 4)

Connectionists believe that language at any one of the domains (phonology, syntax, etc.) is the result of interactions between language and environmental variables. The sum is a dynamic, complex, non-linear system where the timing of events can have dramatic influence on the development course and outcome (Elman et al. 1996; MacWhinney, 1998). Emergentists claim that rule-like reality can emerge from clearly unregulated behavior. For Emergentists, language is like the majority of complex systems that exists in nature and which empirically delineate hierarchical structure (H. A. Simon, 1969). They believe that the complexity of language emerges from rather simple developmental phenomena being exposed to a complicated environment. Thus they supplant a process description for a state description, study development rather than the final state, and focus on the language acquisition process (LAP) rather than language acquisition device (LAD) (Ellis, N. 1999). As N. Ellis (1998) believe connectionists argue just as the conceptual components of language may derive from cognitive content, therefore, the computational facts about language originate from nonlinguistic processing, that is, from the large number of competing and converging constraints imposed by perception, production, and memory for linear forms in real time (Bates, 1984). In the same vein, Elman (2005) describes emergentists' view as:

In emergentist view the language sits at the crossroads of a number of small phenotypic changes in our species. That interacts uniquely to yield language as the outcome. Here, language is seen as a domain-specific outcome that emerges through the interaction of multiple constraints, none of which is specific to language. Figure 2.1). (Elman, 2005 p. 114)

To put it another way, (N. Ellis 1998) contend that Emergentists believe that the universals of language have emerged, just as the universals of human transport solutions have emerged. They have many examples from other domains to simulate and prove this position. Just as Simon says: "Cars are cars all over the world" (Simon, 1969). As (N. Ellis, 1998, 1999) mentioned universal properties of languages have not originated from some preordained design, rather they have emerged from the constraints imposed by human transport goals, society, physics, ergonomics and the availability of natural resources. Humans have evolved systems for perceiving, and representing different sources of information such as vision, space, audition, touch, motor action and emotion (N, Ellis, 1998). Simple learning mechanisms, operating in and across these systems as they are exposed to language data as part of a communicatively rich human social environment by an organism eager to exploit the functionality of language, suffice to drive the emergence of complex language representations (Simon, 1969).



Connectionism

Although, these claims of Emergentists are plausible; there is obviously too little explanation of the involved processes (N, Ellis, 1998). The reason may lie in the complexity and ambiguity of these individual domains that interact to yield language (Elman, 2005). (N, Ellis, 1998) asserted that when one doesn't properly understand any of the individual domains that interact to yield language, how can one hope to perceive the emergent product of their interaction? Moreover, the interactions are going to be so complex that their unique nature cannot be anticipated before they appear in the linguistic evidence. For these reasons, Emergentists draw on connectionism because it provides a set of computational tools for exploring the conditions under which emergent properties arise (N, Ellis, 1999; McClelland et al. 1986). One can list Connectionism's advantages for this aim: firstly, neural inspiration, secondly, thirdly, distributed representation and control, fourthly data-driven processing with prototypical representations emerging rather than being innately pre-specified, then emphasis on acquisition rather than static description, slow, incremental, non-linear, content-sensitive and finally structure-sensitive learning; generalization and transfer as natural products of

learning; and, since the models must actually run, less scope for hand waving (Churchland & Sejnowski, 1992; McClelland et al. 1986).

Connectionists are of the belief that although language behavior can be explained as being rule-like, this does not imply that language behavior is rule governed (N, Ellis, 1998, 1999; McCleland, Rumelhart, & PDP Group, 1986). Rather, they explore how simple learning mechanisms in artificial neural networks are able to acquire the associations between, say, forms and meanings, joined with their respective reliabilities and validities, and then utilize these associations to produce novel responses through generalization (N, Ellis, 1999; Levy, Bairaktaris, Bullinaria & Cairns, 1995). Connectionist models demonstrate how symbolic associative systems, neither given nor identifiable rules, persuade rule-like grammatical behavior (Miikkulainen, 1993).

Connectionist approaches to language acquisition explore the representations that can result when simple learning mechanisms are exposed to complex language evidence (N, Ellis, 2003). In connectionist approach the hypotheses about the emergence of representation is tested by assessing the efficacy of these implementations as computer models consisting of a number of artificial neurons connected in parallel forms (N, Ellis, 1998, 19999; Miikkulainen, 1993).

Connectionism likens the brain to a computer consisting of neural networks. Learning in this view occurs on the basis of associative processes, rather than the construction of abstract rules (Ingram, 2007). According to this paradigm, the human mind looks for associations between elements and creates links between them. These links become stronger as these associations keep recurring and they also make relations with other connections between elements. Connectionists claim that learners make use of regularities in the language input and extract probabilistic patterns on the basis of these regularities.

II. CONNECTIONIST MODELS AND NEURAL NETWORKS

Connectionist networks are referred to as neural networks because they have a number of the basic characteristics of a biological neural network (Gregg, 2003). A connectionist model includes a number of simple processing units (artificial neurons) that are interconnected by their inputs and outputs. To decide whether or not to launch a processing unit, it integrates the influences that operate upon it at a specific point of time. This is because there is a analogy with the way neurons in a biological network behave (Gregg, 2003; Ingram, 2007). In these systems learning is achieved through modifying the connection weights on synapses -the points of contact between processing units (Ingram, 2007).

Localist and distributed networks

Early computational simulations of language learning are known as localist networks. It is because their functional architecture was so that in it each unit had a designated task. In other words, in these networks each node in the network is regarded as functionally distinct element. In order to be able to learn new elements these networks should be rewired to encompass new elements (Miikkulainen, 1993).

Distributed networks are flexible and rigorous neural network architectures, where the linguistic elements are not operated in particular nodes, but distributed across activation patterns of the whole system (Ingram, 2007). Elman's recurrent network which is widely used in the field of language acquisition is an example of distributed networks.

Artificial Neural Networks

Generally, a biological neural network includes of a number of chemically connected or functionally `distinct neurons. A single neuron, presumably, connected to a number of other neurons and the whole number of neurons and connections in a network may be extensive (Gregg, 2003). Connections, also called synapses, are formed from axons to dendrites, and other connections are plausible. Apart from the electrical signaling, there are other forms of signaling that originate from neurotransmitter diffusion, which have an effect on electrical signaling (Elman, 2006; Gregg, 2003).

An artificial neural network (ANN), also called a simulated neural network (SNN) or commonly neural network (NN) is a joined group of artificial neurons that utilizes a computational model for the purpose of information processing on the basis of a connectionist model to computation (Arshvasky, 2006). In most cases an ANN is an adaptive system that makes some changes in its structure based on external or internal information that runs through the network (Hagan, 1996).

In more practical ways, neural networks are non-linear statistical data modeling or decision making tools (Arshvasky, 2006; Ingram, 2007). They can be utilized to simulate complicated relationships between inputs and outputs or to find regularities in data (Arshvasky, 2006).

An artificial neural network composes of a network of simple processing factors (artificial neurons) which can portrait complicated global behavior, selected by the connections between the processing elements and element parameters (Arshvasky, 2006). Artificial neurons were first put forwarded in 1943 by Warren McCulloch, a neurophysiologist, and Walter Pitts, a logician. In a neural network model simple nodes, which can be referred as variously "neurons", "Processing Elements" (PE) or "units", are joined together to shape a network of nodes — therefore the term "neural network". Whereas, a neural network does not have to be adaptive, its practical application comes with algorithms which designed to adjust the strength (weights) of the connections in the network in order to produce a desired signal flow (Arshvasky, 2006).

In an ANN every neuron has an associated activation value, often between 0 and 1 approximately analogous to the firing rate of real neuron (Elman, 2006; Gregg, 2003). Psychologically speaking, meaningful objects can then be represented as models of this activity across the whole set of artificial neurons. The units in the artificial network are

normally multiply interconnected by connections with variable strengths or weights. These connections allow the level of operation in every unit to enhance the level of activity in all the units to which it is joined (N, Ellis, 1998, 1999, 2003). The connection strength are modified by an appropriate learning algorithm, in such a way that when a specific model of activation emerges across one population it can lead to a desired model of activity emerging in another set of units (N, Ellis, 1998, 1999; Gregg, 2003).

There are different standard architectures of model; each one is appropriate to a particular kinds of classification. The most common models have three different layers: the input layer of units, the output layer, and an intervening layer of hidden units (so-called because they are hidden from direct contact with the input or the output) (N, Ellis, 1998, 19999, 2003). The presence of these hidden units enable more complex input and output mapping to be learned than would be possible if the input units were directly connected to the output units (N, Ellis, 2003; Elman et al. 1996). The most common learning algorithm is back propagation, in which, on each learning trial, the network compares its output with the target output, and propagates any difference or error back to the hidden unit weights, and in turn to the input weights, in a way that reduces the error (N, Ellis, 2003). The utility of artificial neural network models may lie in the fact that they can be utilized to infer an operation from observations and also to employ it (N, Ellis, 1998, 1999, 2003). This is particularly useful in applications where the complexity of the data or task makes the planning of such a function impractical.

III. EMPIRICAL STUDIES ON CONNECTIONISM

One of the earliest works in this approach was the study done be Rumelhart and McClelland (1986). They devised a model to simulate the learning of English past tense on the basis of associative patterns. Their model used a computer that made generalizations based on the input that was presented to it.

This model was not only able to acquire the correct past tense endings of English verbs, but most importantly, it made some overgeneralization errors, similar to those that English children make. In other words it was able to simulate the famous U-shaped learning curve of English past tense. This model was criticized (Pinker and Prince 1988) on the basis of some differences on the rate of exposure to input or the rate of learning but the important point was just the ability of the system to acquire these rules because these networks are just very tiny models of a very extensive real network. The application of the model has now been extended beyond the realm of morphology to phonology, syntax and the lexicon (N.C. Ellis, 2003)

Sokolik and smith (1992) used a coercionist network to investigate the assignment of gender to French nouns. The system used the orthographical clues to decide which gender should be assigned to the nouns. In, French for example, nouns ending in -ette or -tion are feminine, while nouns ending in -eur or -on are masculine. Although this is not always true but studies indicate that French children also use these clues to assign gender.

Their system learnt to determine correctly the gender of a number of French nouns. The model was also able to generalize from that learning experience data and assign gender to already unstudied nouns with a high degree of reliability. This system assigned gender through relying on the orthography of the nouns, to the exclusion of any other cues such as adjective or pronoun agreement, or semantic clues. So sokolik and smith (1992) concluded that the model was able to assign grander accurately on the basis of the regularities (associative patterns) it had observed in the input.

N.C. Ellis and Schmidt (1997) investigated the study of English past tense morphology done by Rumelhart and McClelland (1986). Based on their research they had claimed that a connectionist model reproduced very closely the way in which children acquire the past tense in English. Their study had been criticized by pinker (1991), who had argued that only irregular verbs are learned by associations and regular verbs are learnt as a symbolic rule.

Ellis and Schmidt (1997) used a connectionist network to study the adult acquisition of plural morphology. To do this they devised an artificial language and presented it to their adult participants. They also used it as data to their connectionist network. They found that the results from their connectionist network were very similar to that of adult learners. So they came to the conclusion that associative patterns suffice to explain the acquisition of plural morphology and there is no need for the dual route presented by pinker (1991).

Connectionist models of language have their historical antecedents in learning theory, in psychology, and in much older philosophical tradition of empiricism and associationist views of mind. But what sets contemporary connectionist models of language apart from their behaviorist and empiricist forebears is that they take the form of computational simulations, rather than being purely paper and pencil models. (Ingram 2007 p.79)

In computational simulations, the performance of the model is evaluated by comparing it with the performance of human language performance.

The major criterion for evaluating a connectionist model is not so much 'Is it a good analogue of how the brain is wired? But rather, can it simulate interesting and non-obvious aspects of the process under study. (Ingram, 2007 p.81)

IV. THE IMPORTANCE OF EARLY LIMITATIONS OF LEARNING MECHANISMS

One of the most interesting studies in this approach is the research done by J.L. Elman (1993). Elman (1993) in his paper mentions two noteworthy differences between humans and other species. These human's exceptional capacity to learn and the unusually long time it takes to reach maturity. He considers the first difference –unusual capacity to learn-

positive and the other difference –long time to reach maturity- negative. To provide a reason why the evolutionary pressure has not cut this long period, he suggests a theory. His theory is that, these two differences are related to each other. In other words, in order to for humans to posses this unusual capacity to learn, they should pass through such a long period to reach maturity. He proposed the theory that, the limitations in learning mechanisms such as memory span and attention in childhood is a factor that facilitates learning language. Elman proposed that in humans, learning and development interact in an important and non-obvious way. Maturational changes may provide the enabling conditions which permit learning to be most effective. (Elman, 1993)

In order to investigate this claim, he devised an ANN and made some simulations. What he did in short terms is that: He made 10000 sentences with varying length and complexity. Hi did the study in two different phases and some simulations in each of these phases. In first phase, he maintained the learning mechanism (ANN) constant. He used all 10000 sentences (without any grouping based on their complexity) as input data for the network and observed the success of the network. Then he divided the sentences into four groups based on their complexity, each consisting of 2500 sentences. This time, again maintaining the learning mechanism constant, he fed the sentences in each group by the order of complexity of the sentences in each group. The success of the network in learning was much better from when all the sentences were fed at the same time. This was something that seemed logical and is one of the principles of human learning, something close to the concept of Krashen's i+1. But the problem is that this incremental input for children is unrealistic. There is good evidence that children are exposed to language input at its full range in their environment. So the interesting part of the research began. In second phase of the study, he did not sort the sentences according to their length and complexity. But he interfered in the learning mechanism (ANN) itself. In this phase, similar to the first phase, he did four simulations but instead of maintaining the network constant and increasing the complexity of the input, he handicapped the network in four varying degree. At first simulation, he deprived the network severely of some of its memory capacities and presented all the data. At second and third simulations, he increased the memory capacity of the network by reducing the obstacles to the network and again presented all the data. At last simulation he did not interfere at all in the network, so let the network to utilize all its memory capacities. So in short terms, in this phase, he began with a limited learning capacity and gradually increased it. The results were much better from when the network had all its learning capacity from the beginning. He concluded that if the learning mechanism was permitted to undergo "maturational change" (hence, increasing its memory capacity) during learning process, then the outcomes are as good as if the environment had been progressively complicated.

V. CONCLUSIONS

The theoretical underpinnings and tenets of Connectionism were touched upon in sharp contrast to symbolic account of second language acquisition this short paper. Moreover, a number of advantages of this model to language acquisition were enumerated. At the end of the paper, mention was made of few empirical studies in the domain of connectionism. What remain for the L2 researchers and applied linguistics is to investigate different structures and different domains languages.

REFERENCES

- [1] Arshavsky, Y.I. (2006). ""Scientific roots" of dualism in neuroscience", Progress in Neurobiology. *Prog Neurobiol*, Jul: 79(4): 190-204.
- [2] Chomsky, N. (1965). Aspects of a theory of syntax .Cambridge, MA: MIT Press.
- [3] Chomsky, N. (1982). Some concepts and consequences of the theory of government and binding. Cambridge, MA: MIT Press.
- [4] Chomsky, N. (1995). The Minimalist Program. Cambridge, MA: MIT Press.
- [5] Cook, v. and Newson, M. (1996). Chomsky's Universal Grammar. An Introduction. Oxford: Blackwell.
- [6] Cook, G., & Seidlhofer, B. (Eds.). (1995). Principles and practice in applied linguistics. Oxford: Oxford University Press.
- [7] Ellis, N.C. and Schmidt, R. (1997). Morphology and longer distance dependencies: Laboratory research illuminating the A in SLA. *Studies in Second Language Acquisition*, 19, 145-171.
- [8] Ellis, N.C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91–126.
- [9] Ellis, N.C. (1998). Emergentism, connectionism and language learning. Language Learning, 48, 631-664.
- [10] Ellis, N.C. (2003). Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In Doughty, C. and Long, M. (eds.), *The Handbook of Second Language Acquisition*. Oxford: Blackwell Publishing, 63-103.
- [11] Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking innateness: A connectionist perspective on development. Cambridge, MA: MIT Press.
- [12] Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48, 71-99.
- [13] Elman, J.L. (2005). Connectionist models of cognitive development: where next? Trends in cognitive sciences, 9(3), 111-117.
- [14] Hagan, M.T., H.B. Demuth, and Beale, M.H. (1996). Neural Network Design, Boston, MA: PWS Publishing.
- [15] Hulstijen, J. H. (2002). Towards a unified account of the representation, processing and acquisition Of second language knowledge". Second Language Research. 18/3: 193- 223.
- [16] Ingram, J.C.L. (2007). Neurolinguistics, An Introduction to Spoken Language Processing and its Disorders, Cambridge University Press.
- [17] Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99(2), 349-364.

- [18] Levy, J.P., Bairaktaris, D., Bullinaria, J.A., & Cairns, P. (Eds.). (1995). Connectionist models of memory and language. London: UCL Press.
- [19] McClelland, J.L. and Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*. 6, 465–472.
- [20] MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- [21] MacWhinney, B., & Bates, E. (Eds.). (1989). The crosslinguistic study of Sentence processing. New York: Cambridge University Press.
- [22] MacWhinney, B. (1997). Second language acquisition and the competition model. In A.M.B.de Grootand J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp.113–144). Hillsdale, NJ: Lawrence Erlbaum.
- [23] MacWhinney, B. (Ed.). (1998). The emergence of language. Hillsdale, NJ: Lawrence Erlbaum.
- [24] Marcus, G.F. (1995). Discussion: The acquisition of the English past tense in children and multilayered connectionist networks. *Cognition*, 56, 271-279.
- [25] Miikkulainen, R. (1993). Subsymbolic natural language processing. Cambridge, MA: MIT Press.
- [26] Mitchell, R. and Myles, F. (2004). Second language learning theories. Oxford University Press Inc.
- [27] Plunkett, K. and Marchman, V.A. (2002). Learning from a Connectionist Model of the Acquisition of the English Past Tense, MIT Press.
- [28] Preston, D. (1989). Sociolinguistics and second language acquisition. Oxford: Blackwell.
- [29] Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month old infants. Science (5294), 1926-1928.
- [30] Simon, H.A. (1969). The sciences of the artificial. Cambridge, MA: MIT Press.
- [31] Trueswell, J., Tanenhaus, M. K., & Kello, C. (1993). Verb specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(3), 528-553.
- [32] Ungerer, F., & Schmid, H.J. (1996). An introduction to cognitive linguistics. Harlow, UK: Addison Wesley Longman.



Mohammad Reza Yousefi Halvaei holds a BS degree in electronics from Urmia University in Iran. He also holds a MA degree in ELT from Tabriz University in Iran. His area of interest includes connectionism, computational linguistics and discourse analysis.



Musa Moradi has taught English for undergraduate students for more than 15 years. He has presented and published a number of papers. His fields of interest are: SLA, and the role of interactional feedback in L2 acquisition. For the time being, he is a faculty member at Islamic Azad University of Bonab Branch, Iran.



Mohammad Hossein Yousefi is currently doing his PhD at the Islamic Azad University of Khorasgan (Isfahan), Iran. He also is a member of faculty at Islamic Azad University of Bonab Branch, Iran. His main research interests are; Task-based Language Teaching, Cognitive Complexity, and SLA.