Journal of Applied Linguistics and Language Research
Volume 2, Issue 3, 2015, pp. 111-119
Available online at www.jallr.ir
ISSN: 2376-760X

Poverty of Stimulus: A Detrimental or Facilitative Argument in Favor of an Incremental Connectionist Model

Mehdi Shokouhi
Department of English, Sari Branch, Islamic Azad University, Sari, Iran

Seyyed Abdolreza Moghimi
MA student of TEFL, Sari Branch, Islamic Azad University, Sari, Iran

Abstract
One of the most controversial issues in language acquisition is the notion of poverty of stimulus which led the greatest intellectual and linguist of the century, Noam Chomsky to conclude in favor of an innate ability for human being to pick up a language. But according to the connectionists in general, Chomsky somehow jumped to conclusion in that poverty of stimulus argument is not only not disadvantageous but also facilitative from a connectionist view. As a result, the present paper demonstrates that how Jeffrey Elman reasons that poverty of stimulus is an absolute advantage from a connectionist perspective in general and from an incremental connectionist point of view in particular considering a phenomenon such as language acquisition.
Keywords: connectionism, neural network, learning algorithm, poverty of stimulus argument

INTRODUCTION

Connectionism offers a challenge to *traditional symbolic* models of cognition. According to Gasser (1990), despite the powerful appeal of symbols, rules, and logic, the traditional view suffers from a very unhuman-like brittleness. Linguistic and conceptual entities are assigned in an all-or-none fashion to categories, rules typically apply in a fixed sequence, and deviations from expected patterns are not handled well while in connectionist models the brittleness is avoided due to the fact that the entities that a connectionist system uses to characterize the world are fluid patterns of activation across portions of a network.

Moreover, considering a distinction between the traditional symbolism and the connectionist proposal, Poersch (2005) underlines while symbolism processing, based on digital computers, aims to model the mind as a symbol processor, connectionism highlighting a parallel distributed processing has a different origin in that it attempts to design computers inspired by the brain.

Correspondence: Mehdi Shokouhi, Email: mehdi_shokouhi2002@yahoo.com

Additionally, Poersch (2005) highlights that one of the main goals of connectionism is to provide an account of the mechanisms that support cognitive processing. He also adds that connectionists are interested in describing the internal states of brain activity even though they may view them as fundamentally associative in nature.

According to Redington and Chater (1998), connectionism, based on a neural inspiration, means that the brain consists of a very large number of simple processors, neurons, which are densely interconnected into a complex network. Furthermore, neurons appear to communicate numerical values rather than symbolic messages, and therefore neurons can be viewed as mapping numerical inputs onto numerical outputs. So, a neural network is a massively distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: 1. Knowledge is acquired by the network through a learning process. 2. Inter-neural connection strengths known as synaptic weights are used to store the knowledge.

What's more a neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. The procedure used to perform learning processes is called a "learning algorithm"; the function of this algorithm is to modify the synaptic weights of the network in order to attain a desired design.

More importantly, Feldman and Ballard (1982) highlight, an important aspect of connectionist networks is their ability to learn. Most connectionist models come equipped with a built-in learning algorithm that enables them to learn from their experiences, which is also an issue of paramount significance in the field of artificial intelligence.

Amazingly and interestingly they believe that connectionist models can be trained to perform a wide variety of tasks, for example, predicting the appearance of an object from behind a screen, changing a verb into its past form, predicting the next word in a sentence, categorizing objects, categorizing speech sounds, pronouncing written text. In each case, the learning algorithm fine-tunes the strength of the connections in the network until adult-like performance is achieved.

BASIC FEATURES OF CONNECTIONIST MODELS

According to Rumelhart and Zipser (1985), most connectionist models share the following basic features:

1. The system's memory consists of a network of simple processing units joined by weighted connections.
2. The behavior of units is based on neurons. They sum the inputs they receive on connections and compute an activation, which is a function of the total input.

3. The analogue of long-term memory in other models is the set of weights on the network connections. In learning models, these weights are adjusted as a consequence of processing.

4. Processing is parallel. In connectionist models, as in the brain, there is activity in many places simultaneously.

5. Control is distributed. Unlike traditional cognitive models, connectionist systems have no central executive whose job it is to determine which rule or rules are currently applicable and to execute them. In fact, there are no rules to be executed.

CONNECTIONISM AND LANGUAGE LEARNING

One of the most influential and interesting researches considering the connectionist account of language learning is a thoroughgoing project reported by David Rumelhart and James McClelland (1986 and 1987, as cited in Pinker and Prince, 1988).

Using standard PDP mechanisms, they reported that their model learns to map representations of present tense forms of English verbs onto their past tense versions. It handles both regular and irregular verbs, and more interestingly, productively yielding past forms for novel verbs which were not present in its training set, and above all, it is able to distinguish the variants of the past tense morpheme conditioned by the final consonant of the verb.

Rumelhart and McClelland report the result of their research project as following:

> We suggest implicit knowledge of language may be stored in connections among simple processing units organized into networks. While the behavior of such networks may be describable as conforming to some system of rules, we suggest that an account of the fine structure of the phenomena of language use and language acquisition can best be formulated in models that make reference to the characteristics of the underlying networks. (Rumelhart & McClelland, 1987)

> We believe we have provided a distinct alternative to the view that children learn the rules of English past-tense formation in any explicit sense. We have shown that a reasonable account of the acquisition of past tense can be provided without recourse to the notion of a "rule" as anything more than a description of the language. We have shown that there is no induction problem. The child need not figure out what the rules are, nor even that there are rules whatsoever. (Rumelhart & McClelland, 1986).

> We view this work on past-tense morphology as a step toward a revised understanding of language knowledge, language acquisition, and linguistic information processing in general. (Rumelhart & McClelland, 1986).

A BIT OF CLARIFICATION: CONNECTIONISM OR NEO-BEHAVIORISM

Some critics of connectionism (Fodor & Pylyshyn, 1988; Pinker & Prince, 1988) argue that connectionism is nothing more than a revival of behaviorism dressed up to look like neuroscience.

It has to be mentioned that it is true that connectionist models share with behaviorism a focus on the learning of stimulus-response associations. But, according to Gasser (1990), the differences lie in the concern of connectionists with the internal representations that are constructed between the inputs from and the outputs to the environment and with the specific mental processes that are involved in the construction of these representations. In addition, many connectionist models involve feedback connections which would not be possible in a strict stimulus-response framework.

More importantly, connectionists are also increasingly concerned with the initial structure of the networks they work with, that is, with what could be thought of as innate "knowledge" of a sort.

POVERTY OF STIMULUS: A DETRIMENTAL OR FACILITATIVE ARGUMENT

One of the most controversial issues in language acquisition is the notion of poverty of stimulus which led the greatest intellectual and linguistic of the century, Noam Chomsky to conclude in favor of an innate ability for human being to pick up a language. Chomsky (1997) puts heavy emphasis on the issue that the primary linguistic input underdetermines the language output in that there are some aspects of language that cannot be learnt from linguistic input which he believes to be in a state of poverty.

But according to the connectionist scholars in general, Chomsky somehow jumped to conclusion in that poverty of stimulus argument is not only not disadvantageous but facilitative from a connectionist view as well.

In a series of articles (1990; 1991; 1993; 1995; 1999; 2003; 2004; 2005; 2009) and in a series of 17 high profile conferences held annually on the issue of *Cognitive Science Society*, Jeffrey Elman, a PhD holder in Linguistics from university of Austin at Texas and currently a distinguished professor at the university of California in San Diego reasons how poverty of stimulus is an absolute advantage from a connectionist perspective in general and from an incremental connectionist point of view in particular.

Elman (1990) starts his job with an article under the rubric of "finding the structure in time" by considering time effect in a neural network to explore how time should be represented in connectionist models.

According to Elman, one approach is to represent time implicitly by its effects on processing rather than explicitly as in a spatial representation, which involves the use of recurrent links in order to provide networks with a dynamic memory. Moreover, Elman reports that these networks are able to learn interesting internal representations which incorporate task demands with memory demands and in this approach the notion of memory is inextricably bound up with task processing.

Elman himself, in his 1990's article, confessed that the results of his research project are preliminary in nature in that they are highly suggestive, and often raise more questions than they answer.

As a result, Elman (1991) undertook a research project investigating whether or not connectionist models are capable of complex representations which possess internal structure and which are productively extensible. He reports the important result of his work is to suggest that the sensitivity to context which is characteristic of many connectionist models, and which is built-in to the architecture of the networks does not preclude the ability to capture generalizations which are at a high level of abstraction. Nor is this a paradox.

According to Elman, sensitivity to context is precisely the mechanism which underlies the ability to abstract and generalize. The fact that the networks exhibit behavior which is highly regular is not because they learn to be context-insensitive. Rather, they learn to respond to contexts which are more abstractly defined.

Although, in his two previously articles (i.e., 1990, 1991), Elman tries to be tentative and suggestive considering the results of his papers, however, in his 1993's article, he turns out to be more self-confident and more importantly ground-breaking.

Elman (1993) underlines that it is a striking fact that in humans the greatest learning occurs precisely at that point at time of childhood when the most dramatic maturational changes also occur. Elman believes there are circumstances in which models work best when they are forced to "start small" and to undergo a developmental change which resembles the increase in working memory which also occurs over time in children. This effect occurs because the learning mechanism in such systems has specific shortcomings which are neatly compensated for when the initial learning phase takes place with restricted capacity.

Furthermore, Elman writes that the poverty of stimulus argument forced linguists to believe that the knowledge of language is biologically predetermined and innate in nature. But Elman (1993) argues that there is another factor in helping account for the apparent ability of learners to go beyond the data. This factor hinges on the simple fact that language learners (children) are themselves undergoing significant developmental changes during precisely the same time that they are learning language. What's more, Elman argues that the poverty of stimulus is not only advantageous but a necessary and prerequisite phase in language learning.

Elman goes on to give an account of his research project to corroborate his claim. Elman (1995) reports of a regimen in which the training input was organized into corpora of increasing complexity. There were five phases in all. In the first phase, 10,000 sentences consisting merely simple sentences were presented. The network was trained on five exposures to this database. At the conclusion of this phase, the training data were discarded and the network was exposed to a new set of sentences. In this second phase, 7,500 of the sentences were simple, and 2,500 complex sentences were also included. As before, the network was trained for 5 epochs, after which performance was also quite high, even on the complex sentences. In phase three, the mixture was 5,000 simple/5,000 complex sentences, for 5 epochs. In phase four, the mixture was 2,500 simple/7,500 complex. And in phase five, the network was trained on 10,000 complex sentences.

At the conclusion of training, the network's performance was quite good, for complex as well as simple sentence. Furthermore, the network generalized its performance to novel sentences as well. This result contrasts strikingly with the earlier failure of the network to learn when the full corpus was presented at the outset. Put simply, the network was unable to learn the complex grammar when trained from the outset with the full "adult" language. However, when the training data were selected such that simple sentences were presented first, the network succeeded not only mastering in these, but then going on to master the complex sentences as well.

In this regard, Elman (1999) concludes that in one sense, this is a pleasing result, because the behavior of the network partially resembles that of children. Children do not begin by mastering the adult language in all its complexity. Rather, they begin with the simplest of structures, and build incrementally until they achieve the adult language.

Elman (2003) puts a nice analogy forward which he believes at the same time to be a dis-analogy. Elman believes there is an important disanalogy between the way in which the network was trained and the way children learn language. In this simulation, the network was placed in an environment which was carefully constructed so that it only encountered the simple sentences at the beginning. As learning and performance progressed, the environment was gradually enriched by the inclusion of more and more complex sentences.

Elman emphasizes that if it is not true that the child's environment changes radically, what is true is that the *child* changes during the period he or she is learning language. A more realistic network model would have a constant learning environment, but some aspect of the network itself would undergo change during learning.

Moreover, Elman(2005 a and b), reasons that When learning proceeds in an incremental fashion, either because the environment has been altered or because the network itself is initially handicapped, the result is that the network only sees a subset of the data. When the input is staged, the data are just the simple sentences. When the network is given a limited temporal window, the data are the full adult language, but the ***effective*** data are only those sentences, and portions of sentences, which fall within the window and these are the simple sentences.

More importantly, Elman believes the simple sentences, contain only three of the four sources of variance (grammatical category, number, and verb argument type) and there are no long-distance dependencies. As a result, the network is able to develop internal representations which encode these sources of variance. When learning advances (either because of new input, or because improvements in the network's memory capacity give it a larger temporal window), all additional changes are constrained by this early commitment to the basic grammatical factors.

Indeed, Seen in this light, the early limitations on memory capacity and linguistic data assume a more positive character as it was taken in a Nativist school in that there is no poverty of stimulus at all in that the input that child receives is rich enough but it is the limited memory capacity which act as a filter on the input, and focus learning on just that subset of facts which lay the foundation for future success.

However, Elman (2009) provides a clearer picture of his idea proposing an incremental connectionist model. Elman believes the incremental learning strategy is an example of how a system can learn a complex domain by having *better initial data*.

In the case of learning a language, the language problem is hard for the network to learn because crucial primitive notions are obscured. This makes it difficult to learn the primitive representations. But the important issue here is that we have a Catch-22 problem. The network is also unable to learn about the complex grammatical structures because it lacks the primitive representations necessary to encode them. These difficulties are compounded by the network's early commitment to erroneous hypotheses, and its tendency to ossify (harden, to be inflexible) over time.

Incremental learning, according to Elman's over all proposals, solves the problem by presenting the network with *just the right data* (i.e., data which permit the network to learn the basic representational categories) and *at just the right time* (i.e., early on, when the network's plasticity is the greatest).

More importantly, a key aspect to the solution, as far as its possible relevance to the human case, is that there is a natural mechanism available for doing the filtering. Elman believes by starting with an immature memory which allows the system to process only simple sentences, the network constructs a scaffolding for later learning. As time prog***re***sses, the gradual improvement in memory capacity selects more and more complex sentences for processing.

With this perspective, we can conclude that the limited capacity of infants assumes a positive value. As Elman (1993) asserts "limited capacity acts like a protective veil, shielding the infant from stimuli which may either be irrelevant or require prior learning to be interpreted. Limited capacity reduces the search space, so that the young learner may be able to entertain a small number of hypotheses about the world". (p.95)

AN ANTI-NATIVIST AND PRO-INCREMENTAL CONNECTIONIST CONCLUSION

Elman (1999) in a chapter under the rubric of "Origins of language: A conspiracy theory" edited by MacWhinney, writes that incremental connectionism is a new phenomenon based on providing a new interpretation of the notion of poverty of stimulus in that many people have interpreted the fact that language-learning occurs with greatest success (e.g., learners achieve native fluency) during childhood as evidence for a Language Acquisition Device which operates only during childhood. Once its job is done, it ceases to function. But the incremental connectionism suggests rather that the ability which children have for learning language derives not from a special mechanism which they possess and adults do not, but just the reverse. It is children's *lack* of resources which enables them to learn languages fluently.

Indeed, Elman's reasoning and interpretation sheds more light on the conspicuous and notorious notion of the poverty of stimulus and adults failure in achieving the full proficiency. Based on Elman's overall research studies and reports, it can be concluded that developmental limitations can impose constraints which are crucial for achieving a target behavior, and these developmental limitations arise from biological factors, the network described from an incremental connectionist point of view is "innately constrained" to discovering the proper grammar.

But it has to be underlined that this is a very different sort of innateness envisioned by the pre-wired linguistic knowledge outlined by the Chomskyan School as language acquisition device. The incremental connectionism, in a comparison with the Nativism, can provide more legitimate answers to lots of our questions since it is not only explanatory but observational at well. For example, why children starts from very simple structures and then move on to the more complex ones, which is interpreted as the in-built acquisitional order from a Nativist perspective, can be justified as children do so because this is the most they can do or using Elman's terminology, this is the most their memory capacity window let them do. And this should not be taken as there to be a biologically predetermined in-built system.

Another issue is the evolution that the children arrive at considering their language ability which is again answered by the Nativist that it is because of the already device over there but it seems that this sort of answer has become some sort of bothering buzzword in that the answer to all of the issue considering the notion of the language word be this innate ability. However, from an incremental connectionist perspective the evolution happens as a result of the harmonious and hand-in-hand cooperation of the learning mechanism and the human maturational development for providing, using Elman's phrase, more windows for more complex phenomenon which turns out to be an evolution in our outlook.

Finally, it has to be highlighted that while Chomskyan notion of innateness refers to a sort of in-built syllabus which the children have to undergo in order to pick up a language, the incremental connectionism favors a built-in syllabus which the children arrive at.

REFERENCES

Chomsky, N. (1997). Language and problems of knowledge. *Teorema, 16*(2), 5-33.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211. doi:10.1016/0364-0213(90)90002-E

Elman, J.L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7,* 195-225. doi:10.1007/BF 00114844

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71-99. doi:10.1016/0010-0277(93)90058-4

Elman, J.L. (1995). Language as a dynamical system. In R. Port & T. van Gelder (Eds.), *Mind as Motion: Dynamical Perspectives on Behavior and Cognition* (253-288). Cambridge: MIT Press.

Elman, J.L. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed.), *The emergence of language* (1-27). New Jersey: Lawrence Earlbaum Associates.

Elman, J. (2003). Development: It's about time. *Developmental Science, 6*, 430-433. doi:10.1111/1467-7687.00297

Elman, J.L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Science, 8*, 301-306. doi:10.1016/j.tics.2004.05.003

Elman, J.L. (2005a). Computational approaches to language acquisition. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (726-732).Oxford: Elsevier.

Elman, J.L. (2005b). Connectionist models of development: Where next? *Trends in Cognitive Science, 9*, 111-117. doi:10.1016/j.tics.2005.01.005

Elman, J.L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science, 33,* 1-36. doi:10.1111/ j.1551-6709.2009.01023.x

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6*, 205-254. doi:10.1016/S0364- 0213(82)80001-3

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3-71. doi:10.1016/0010-0277(88)90031-5

Gasser, M. (1990). Connectionism and universals of second language acquisition. *Studies in Second Language Acquisition, 12*(1), 179-199. doi:10.1017/S0272263100009074

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,*73-193. doi:10.1016/0010-0277(88)90032-7

Poersch, J. M. (2005). A new paradigm for learning language: Connectionist artificial intelligence. *Linguagem & Ensino, 8*(1), 161-183.

Redington, F.M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*(3), 129-191. doi:10.1080/016909698386500

Rumelhart, D. R., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9,* 75-112. doi:10.1016/S0364-0213(85)80010-0