Journal of Language and Communication
Vol. 2, No. 1, 2015, 1-6

Parallel Distributed Processing Model and Second Language Acquisition

Ehsan Abbaspour[1], Mohammad Khatib[2]

Abstract

Based on Connectionist approach, proponents of Parallel Distributed Processing (PDP) believes that language learning involves the activation of nodes and the creation of pathways between neurons within the brain. This article reviews the theoretical underpinnings of PDP in relation to basic issues in SLA research e.g., Universal Grammar, Critical Period Hypothesis, etc.

Keywords: Connectionism, Parallel Distributed Processing, Second Language Acquisition

1. Introduction

Parallel Distributed Processing model (henceforth PDP), as defined by Brown (2007), is a cognitive theory which based on the notion that information is processed simultaneously at several levels of attention. Brown further states that according to PDP model "linguistic performance may be the consequence of many levels of simultaneous neural interconnections rather than a serial process of one rule being applied, then another, then another, and so forth" (p. 31). In this model "the hardware mechanisms are networks consisting of large numbers of densely interconnected units, which correspond to concepts or features" (Pinker & Prince, 1988, p.75).

It must be mentioned that although many scholars consider PDP and Connectionism synonymous (Pinker & Prince, 1988; Richards & Schmidt, 2010), many other scholars believe that Connectionism is the cover term and PDP is one of its approaches (Gass & Selinker, 2008; Hulstijn, 2003; Saville-Troike, 2006). There are several types of connectionist architectures. Hulstijn (2003) notes that the two main classes of connectionism are (1) localist networks, which consist of interconnected symbolic, meaningful categories (such as phonemes, syllables, codas and word stems), and (2) parallel distributed processing (PDP) networks, which consist of sub-symbolic, non-meaningful nodes.

2. Principles of PDP

As Meares (1993) maintains, there are three basic principles that define PDP: parallel processing, neural connections, and the weightedness of connections.

The first and most basic tenet of this theory, states that information processing in the brain is parallel (rather than serial). This means that rather than going step by step, as in rule based systems, information is processed simultaneously.

PDP theory also is based on the belief that symbolic rules are not the basis of language processing, but rather that the neural connections or pathways in the brain are strengthened by the use of the first language and any other languages learned as a child.

The third principle of PDP theory is that all neural connections are not equally joined, but rather they are 'weighted' with some having more weight or in other terms, a stronger connection. Conversely, some

[1]Department of English, Science and Research Branch, Islamic Azad University, Tehran, Iran
[2]Department of English Language and Literature, Allameh Tabataba'i University, Tehran, Iran

connections are more flexible than others, but the level of flexibility may be determined by genetic predisposition or by the influence of input.

These principles in a model based on a study by Seidenberg and McClelland (1989) are represented in Figure 1.

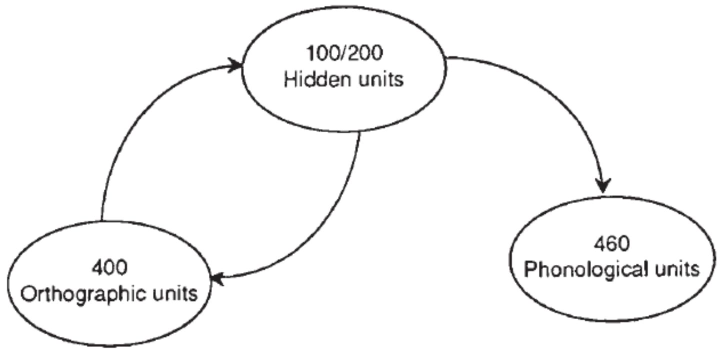

Figure 1: Seidenberg and Mcclelland's (1989, P. 6) Implemented Connectionist Architecture

3. What do Connectionist/PDP Models Claim?

Rod Ellis (2008) enumerates the key percepts of Connectionist/PDP models as follows:

1. Learning is based on simple learning mechanisms.

Connectionism is one of the most influential emergentist models. Emergentism assumes that learning emerges from basic/simple processes.

2. Language is exemplar rather than rule-based.

N. Ellis (2002) claims that knowledge of language is not abstract rules of grammar, but collection of memories of previous experienced utterances.

3. Learning a language involves learning constructions.

N. Ellis (2003, p. 63) defines constructs as "recurrent patterns of linguistic elements that serve some well-defined linguistic function".

4. Learning is a process of gradually strengthening associations between elements.

That is, the association strengths among units are spontaneously modified.

5. Rule-like representations may arise out of the network or associations that learners build.

That is, PDP models can extract regularities in the patterns of interconnections; these are higher-order rules.

6. Processing is carried out in parallel rather than serially.

McClelland, Rumelhart, and Hinton (1986) argue that multiple constraints govern language processing, with semantic and syntactic factors constantly interacting without it being possible to say that one set is primary. They argue that "if each word can help constrain the syntactic role, and even the identity, of every other word, processing must take place not serially but simultaneously on different levels" (p. 7).

7. Language learning is frequency-driven.

As Ellis (2008, p. 468) notes "in real-life learning, large amounts of input are needed to fine-tune the developing system. Learners are sensitive to the frequencies of sequences in the input".

8. Learning is governed by the power law of practice.

Practice improves performance but with a gradually diminishing effect.

9. Role of memory.

The role of working memory in the above-mentioned processes is undeniable.

10. Implicit learning.

As Ellis states, “the learning that results from the induction and subsequent analysis of sequences found in the input occurs without awareness” (p. 469).

11. Explicit learning and knowledge.

Learners need to attend to the input sequence and it is a conscious process.

12. Role of L1.

The development of L2 networks are similar to the networks of L1 acquisition.

4. Simulation of Language Learning in PDP Model

One experiment conducted following the framework of PDP wasMcClelland and Rumelhart’s (1986) simulation ofpast tense learning using a PDP model. This grammatical feature is of considerable interest to language researchers because its naturalacquisition typically involves U-shaped learning behavior (for example, went-goed-wented-went). Overgeneralization errors, such as ‘goed’, indicate that learners must organize theirlinguistic knowledge into ‘rules’. Given that in PDP systems no such a rule exists, past tense learningprovides a rigorous test of the ability of such systems to account for languagelearning.

The PDP system that McClelland and Rumelhart’s used to simulate past tense learning is equipped with (1) an encoding device that operates on the root form of verbs (for example, ‘hope’) and converts them into a set of context-sensitive phonological features (called Wickelfeatures); (2) a ‘pattern associator’ network that takes the output from (1) and computes the past tense form of the verbs, again in terms of Wickelfeatures; and (3) a decoding device that converts the output of (2) into a normal set of phonological features. Large numbers of exemplars of both root and past tense verb forms were fed systematically into the model; they comprised the input.The network showed the same tendency to overgeneralize as a child acquiring L1 does.

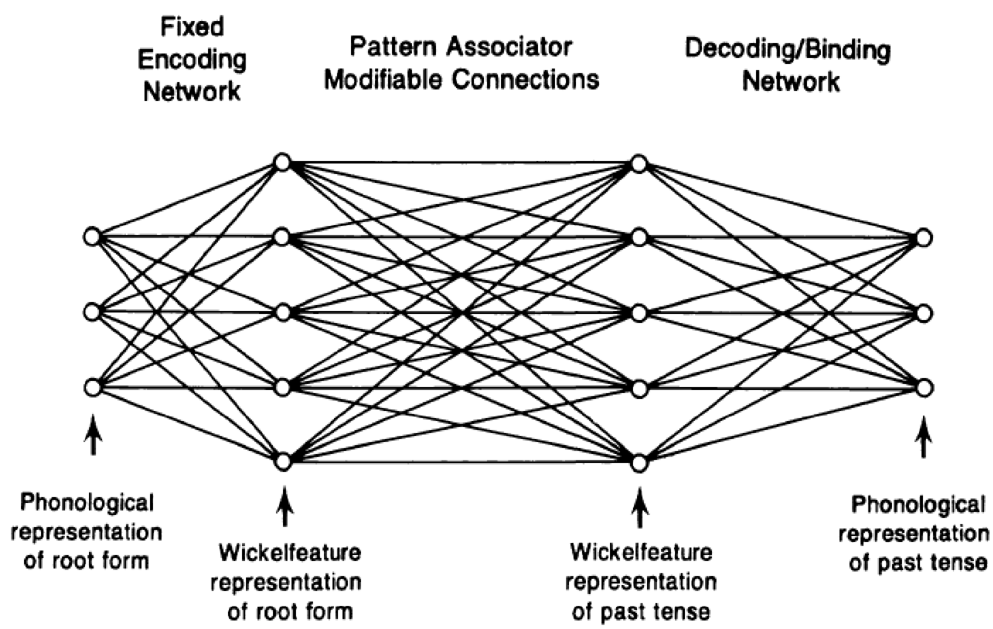

Figure 2: The Basic Structure of Mcclelland and Rumelhart’s (1986) Model

PDP and UG

Contrary to claims by many scholars such as Gasser (1990), Meares (1993) purports that it is possible to see a clear connection between UG and PDP. She believes that it is through the concept of unequal weightedness of neural connections that the concepts of Universal Grammar (UG) can be seen as compatible with PDP. Meares further notes that:

UG can be defined as the constraints on language that humans have, but these constraints by and large are not based on physical constraints. If however, they are constraints based on weights of connections that are present at birth they can be seen as the hardware of the mind. The building of connections through input may then be viewed as the software that conforms to the hardware. UG can be seen as the parameters set within the brain's initial weighted connections. (p. 21).

PDP and CPH

Sokolik (1990) believes that PDP provides a clear explanation for the critical period. He describes a hypothetical nerve growth factor that allows children to make new neural connections. At puberty, the availability of this factor diminishes, preventing new connections from occurring in adulthood. Sokolik (1990) found that children under between the ages of two and ten recovered their use of language. At this age, children would still be able to make new neural connections. Children over twelve, however, had only a partial recovery. Based on PDP theory, these older children would have to adapt other neural networks since they no longer are able to produce new pathways.

Shortcomings of PDP Model

According to Gregg (1996, p. 58) nonmodular approaches to L2 acquisition research have been "largely motivated by a misguided concern for communicative competence, there has been little in the way of explicit nonmodular theorizing".

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 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. Saville-Troike, (2006) also asserts that "researchers from several approaches to SLA which focus on learning processes are taking a renewed look at how frequency influences learning" (P. 81).

Another criticism leveled at PDP models is that a connectionist computer program does not say anything about how the human mind works (Fodor& Pylyshyn,1988).

Norman (1986) also believes that one of the weaknesses of PDP models is concerned with the type-token problem. The type-token problem is to be able to handle different instances of the same concept, sometimes at the same time. Thus, if the system has the knowledge that "John eats a sandwich" and that "Helen eats a sandwich," the system has to treat these as different sandwiches. This capability is not easy for PDP systems.

5. Conclusion

Although there are numerous PDP simulations of language learning, there are a lot of unresolved issues regarding SLA which are hard to be explained by PDP models. In addition, according to Sokolik (1990) the ideas underlying PDP modelsare hardly new ones, because it is rooted in connectionist ideas and run deep in the fields of psychology and neurology around one decade. So, attempting to apply PDP principles to varied fields can be considered as a new idea.

Finally, a combination of symbolic and sub-symbolic processing as what Hultijn (2002) calls a 'hybrid' model may provide an effective model for explaining language learning in general and SLA in particular.

References

Brown, H. D. (2007). Principles of language learning and teaching (5^{th} ed.).White Plains,NY: Pearson Education.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. Studies in Second Language Acquisition,24(2), 141-188.doi: 10.1017/S0272263102002024

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 languageacquisition (pp. 63-104). Oxford: Blackwell Publishing.

Ellis, R. (2008). The study of second language acquisition (2nd ed.). Oxford, UK: Oxford University Press.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. Cognition, 28(1), 3-71.doi: 10.1016/0010-0277(88)90031-5

Gass, S. M., &Selinker, L. (2008). Second language acquisition: An introductory course (3^{rd} ed.). NY: Routledge.

Gasser, M. (1990). Connectionism and universals of second language acquisition. Studies in Second Language Acquisition, 12(2), 179-199.doi: 10.1017/S0272263100009074

Gregg, K. R. (1996). The logical and developmental problems of second language acquisition, In T. K. Bhatia & W. C. Ritchie (Eds.), Handbook of second language acquisition (pp. 50-75). London: Academic Press, Inc.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. Second Language Research, 18(3), 193–223.doi: 10.1191/0267658302sr207oa.

Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. Computer Assisted Language Learning, 16(5), 413-425.doi: 10.1076/call.16.5.413.29488

McClelland, J. L., & Rumelhart, D. E. (1986). On learning the past tenses of English verbs. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition(pp. 216-271). Cambridge, MA: MIT Press.

McClelland, J. L., Rumelhart, D. E., &Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland& the PDP Research Group (Eds.), Parallel distributed processing:Foundations (pp. 3-44). Cambridge, MA: MIT Press.

Meares, M. (1993).Parallel distributed processing and universal grammar: Complementary concepts in a biological experiment for the critical period in language acquisition, Working Papers in Linguistics, 2, 17-23.

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J.L. McClelland, D.E. Rumelhart & the PDP Research Group (eds.),Parallel distributed processing: Psychological and biological models (pp. 531-547). Cambridge: MIT Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. Cognition, 28(1), 73-193. doi: 10.1016/0010-0277(88)90032-7

Richards, J. C., & Schmidt, R. (2010). Longman dictionary of language teaching and applied linguistics (4th ed.). London: Longman.

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96, 523–568. doi: 10.1037/0033-295X.96.4.523

Sokolik, M. E. (1990). Learning without rules: PDP and a resolution of the adult language learning paradox. TESOL Quarterly, 24(4), 685-696.doi: 10.2307/3587115