Parallel Distributed Processing and Universal Grammar: Complementary Concepts in a Biological Explanation for the Critical Period in Language Acquisition

By Mary Meares

'Language offers one of the most complex challenges to any theoretical paradigm.'
Lachter and Bever (1988)

1. Introduction

Linguists studying second language acquisition have discovered and described a critical or sensitive period when a language can be learned completely. Somewhere around puberty, it appears that the ability to learn a language with native-like competence is lost. Especially in the realm of phonetics and phonology, differences can be found in those who learned the language before versus after the teenage years. Linguists have attempted to explain the language learning critical period with various theories, but the biological basis is still not clear. Most theorists however, do believe that there is some sort of biological process that occurs at puberty, limiting further linguistic learning. Traditional linguistic approaches have focused on the mind and what occurs to limit language learning. The goal of this paper is to contrast the traditional explanations for the critical period with a biologically based model, Parallel Distributed Processing. By combining this concept with that of Universal Grammar an attempt is made to describe the process of language development including the critical period phenomena.

2. Traditional Explanations for the Critical Period: the Mind

The concept of a critical period was developed in observation of animal behavior. Some behaviors, like the imprinting of young ducklings to recognize their mothers (or major care givers) are only possible within a certain limited time period (Newport, 1991). If the behavior does not develop during this critical period, the learning of the behavior is no longer possible. Some type of time constraints affect learning of many different behaviors including the development of vision in kittens and song birds learning their distinctive songs. Lenneberg (1967) theorized that language learning undergoes the same type of critical period in humans. Many linguists who have studied the critical period worked to develop research with the goal of describing the parameters and patterns of the critical period (Seliger, 1978). Others have attempted to discover why the critical period occurs in terms of reasons and explanation

(Scovel, 1988; Bever, 1981; Newport, 1991) and it is this that this paper will discuss.

Bever (1981) divided theories explaining the critical period into three different categories. The first was that proposed by Lenneberg (1967) that there was a loss of neurological flexibility at the time of puberty. The second category of theories was that of the language learning capacity being totally filled. This concept is easily disproven by the vast discrepancy in the language abilities of both pre- and post-pubescents. Bever's third category was that of theories where cognitive development interferes with language learning. Krashen's Model and Felix's theory support this third category.

Krashen attempted to explain the critical period in terms of the Monitor and Affective Filter Hypotheses. Adults, according to Krashen (Larsen-Freeman and Long, 1991), have an affective filter that effects their ability to intake language as well as a monitor which they use to monitor and edit their own output. These factors, especially the filter, impose limitations on the adult learner. Although it is possible for this filter to be lowered, this happens completely only in rare optimum circumstances. According to Krashen, the filter, therefore, is the cause of the critical period.

Newport (1991) further questioned the nature of the critical period. She developed two possible scenarios for the diminishing of the ability to learn languages. The first perspective is that 'the underlying learning mechanism itself undergoes maturational decline or decay' (Newport, p. 114). This was Lenneberg's view of the critical period. Newport advocates consideration of an alternate perspective. Rather than a decline, it may be that maturational changes in other processes affect language learning, masking this ability. Felix's theory is an example of this perspective.

Felix (1985) developed the Competition Theory. This theory states that language and problem solving are located in two different modules of the mind. The language module matures first and is active until the problem solving component begins to interfere. The problem solving component matures as individuals enter Piaget's formal reasoning stage, around the age of puberty. At this time the problem solving module tries to take over language learning, but it is not as good at it as the language learning module. This competition is what appears as the critical period according to Felix.

3. Brain

Part of the difficulty in clearly explaining what happens in the Critical Period lies in the problem of the brain versus the mind. All of the previously described explanations for the Critical Period involve the mind. An alternative is to theorize about the brain and how its functioning affects language acquisition. Although medical science is progressing in its quest for knowledge about the brain, one of the difficulties that linguists and scientists share is the problem of examination of neural processes as they occur. In studying the brain, we can observe language production (or output), evidence gained through use of technology (such as CAT and PET scans), and the physical structure of the brain after death, but not what is occurring in the brain as language is produced. All three of these approaches leave much to be learned about behavior and development of neurons and the connections between them, as well as the roles of different areas of the brain. The development of computers and artificial intelligence have allowed theorists to model human development, however the artificial technology cannot at this time match the human abilities to process and produce language. In spite of this, it is important for linguists to consider what is known about the brain in evaluating the validity of theories of language acquisition. Neurological explanations of language can augment cognitive theories to form a sound model.

Many of the theories and models of second language acquisition discussed in this paper and elsewhere have a way of explaining the critical period, yet most are based on logic, not biology; on mind, not brain. The Language Acquisition Device, the Affective Filter, and so forth are all just metaphors that give explanations based on concrete studies of human language acquisition behavior. There is nothing wrong with developing a metaphor to cognitively explain why something may occur, but it is also important to look at the brain itself and what scientists think neurologically is possible. Jacobs (1988) warns against taking metaphors as literal truth citing their 'deceptive appeal' (p.329). Much is not known about the brain, yet in the literature there are some examples of neurobiological evidence and theory.

4. Neurological Explanations for the Critical Period

One neurobiological explanation has been developed by Jacobs (1988) primarily using neurological evidence rather than observation of behavior and applications of suitable metaphors. Although it is scientifically impossible at this time to know what actually happens inside our heads, Jacobs attempted to reconcile knowledge of the brain with the concept of Universal Grammar. He emphasized the importance of the environment and how input has an important role in 'shaping the actual development of neural tissue' (p.309). The brain has approximately 100 billion neurons, each with 50,000 to 200,000 dendritic spines connecting it to other neurons. After birth neurons do not reproduce, but the connections grow and become more complex as a child develops. A second language learned after these connections are established must utilize the existing neuronal framework. Jacobs acknowledges the metaphorical nature of the term universal grammar and he had difficulty reconciling what form universal grammar might take with what is known of the brain, but does recognize the genetic preprogrammed predisposition for language.

In another study focusing on neurological explanations, Walsh and Driller (1981) maintained that the reason that infants were not ready to use language was that their neural connections were not well established. In the first two years, there is a large amount of growth in the dendrites, the receptive surface of neurons. They also document the fact that most connections are made by the age of six to eight. This is used to explain the difficulty in acquiring native-like phonological proficiency. Instead of new connections being formed, the first language pattern of connections is used to approximate the new language. They conclude, however, that the critical period only holds true for speech not 'higher order linguistic functioning' involving abstraction, verbal cognition and construction (p.16).

5. Combining Physical and Mental Theories

The problem with all of the previously described theories is that they do not take into account both the mental and physical knowledge that is available. What is needed is a theory that explains both what comes out of our mouths (the output produced) and what happens in our heads (the hardware that processes the data). Combining neurological information with knowledge of linguistic output, a new more comprehensive theory has been formed. Building on what is known from both human and computer sciences, Parallel Distributed Processing or connectionism models the brain and attempts to produce the resulting language produced by humans. I propose that this model bridges the biological/cognitive gap. A description of this theory reveals that it is compatible with both biological and cognitive notions.

6. Parallel Distributed Processing: A Biological Approach

Parallel distributed processing (PDP) is a recent theory that attempts to focus on the behavior of the brain. For the purpose of this paper I will describe the 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 in the brain is parallel (rather than serial) processed. This means that rather than going step by step, as in rule based systems, information is processed simultaneously. The knowledge that we have of neurons and dendritic connections supports this principle. 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 connections are more flexible than others, but the level of flexibility may be determined by genetic predisposition or by the influence of input.

PDP provides a clear explanation for the critical period. The development of neural networks occurs easily in children. The critical period is the time when the brain is still physically able to make new connections. Sokolik (1990) 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. This explain the different recovery rates of aphasic children as described by Lenneberg (1967). He 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. This chemical possibility raises questions of possible reversal of the critical period as neurochemical expertise grows.

The rejection of linear rules is fundamental to PDP. The symbolic, serial approach which can predict phonological linguistic output is not compatible with the speed at which information is processed. Although they may be used to describe what is occurring to the phonemes and the resulting output, linear rules do not explain how the brain manipulates the information. The neural networks of connection in PDP attempt to show how humans can 'exhibit rule-like behavior without explicit rules' (Gasser, 1990). This rule-like behavior seems to be part of phonology (the aspect of language affected by the critical period) but not syntax. Another argument for rejection of the linear processing is that in a linear system, one missing step results in a stop in processing. The parallel system allows other connections to take over for the missing connection. This may explain the possibility of other neural units partially taking over for damaged ones, although with lessened efficiency.

Gasser's description of the PDP views learning as 'the unsupervised association of pattern elements with one another' (p.179). He also takes care to distinguish connectionism from behaviorism, noting the existence of a true feedback connection rather than just a stimulus-response relationship. One of the examples that Gasser presents as support for connectionism is the success in using computer simulations to model learning. Computers can be set up to follow the same developmental processes found in some morpheme acquisitions studies in humans. There have been some difficulties with computers however, in deciding how to set up the system to simulate the brain. These models, and this theory, do explain transfer and can explain the patterns of language that exist, but Gasser maintains that they are incompatible with Universal Grammar and explain the critical period only in terms of the lack of ability of the adult brain to build new connections.

7. PDP and Universal Grammar

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. 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, or may act to get around it, as a savvy programmer may do. Contrary to Gasser's claim, it is possible to see a clear connection between UG and PDP. UG can be seen as the parameters set within the brain's initial weighted connections. This may be based on binary choices, although the determination to establish a binary system may be a result of rule-based logic, rather than biological predisposition. The processing of input (PDP) then develops connections within the genetic framework, conforming to the specifications of the language heard.

An example from phonology is that of syllable structure. Branching rhymes, nuclei, and onsets are all possible, however there seems to be some restraint on the combinations of these branchings. Phonological studies (Kaye, 1989) have examined languages of the world and found that if the onset in a particular language branches, that the rhyme will also branch. This has been used as a possible example of UG, illustrating the constraints in our heads (usually meant as mental constraints, rather than physical) that limit syllable structure. Based on the linguistic evidence available, it seems to be impossible to have a language with a branching onset but not a branching rhyme. This is an established part of the connections present at birth. As a child learns language phonology, other complementary connections are built up around the established constraint system. An example of this is the input received on cluster constraints and acceptable violations of the sonority hierarchy that are present in languages learned before puberty. After the nerve growth factor disappears, new connections are difficult to form and thus language learning is restrained by the existing phonological networks.

The concept of linguistic markedness may also be based on connections in the brain. Markedness is the concept in phonology that some combinations of features are more likely to occur in any language than other combinations (Kaye, 1989). The less likely a combination is, the more marked it is. When this concept is incorporated into the PDP theories, markedness is seen as the closeness of a neural pathway to a biologically pre-determined connection. The less marked a feature is, the more substantial connection is present prior to input in the first language. More marked features travel along less natural pathways, and therefore may require more input for the pathway to develop.

An analogy can be used to illustrate the roles of UG, markedness, and PDP in language learning. The mass set of neurons in the brain can be compared to a landscape. As the earth was formed, numerous features developed. Rivers, mountains, and forests appeared over time. Within this landscape there are pathways where there were naturally no trees. Some of these may provide bi-directional choices or forks in the road. These naturally treeless pathways are the genetically predispositioned possible connections or UG. Other areas have many trees and underbrush. Over time, new paths can be developed by cutting trees and undergrowth with an ax or machete (nerve growth factor), but after a certain time it becomes impossible to cut more because the tool is lost or dull. This leaves the landscape relatively set with further travel possible only along its established pathways. Second languages learned after puberty must travel along the closest paths available from the first language, although the natural pathways still exist. Pathways closest to natural pathways are less marked that those further away.

8. A Response to Criticism of PDP

Reaction to the connectionist models have not been uniformly enthusiastic. Computer models have limitations and are viewed by some as having been manipulated to give the 'correct' answer. Lachter and Bever (1988) criticized these models, showing that rule-based representations were built into the problem. I maintain that humans have some built-in structure as well, although parallel based rather than rule-based. Our skill in the young but developing field of computer science limits the ability of technology to model the mind totally, however, the computer models that have been used to simulate human language learning have shown that parallel processing based on weighted connections has similar output to human manipulation of phonemes. Lachter and Bever did, however, recognize the importance of these models as tools for increasing understanding of associations and relationships between behavior and underlying structural representations. This critique in fact supports the relationship between UG and PDP. If UG is the 'hardware' of the computer, then a propensity for language is built into our genetic code. The difficulty is in determining UG and building it into computer models in connectionist modes rather than rule-based ones. Much remains to be done in exploring the potential of both these theories and models.

Jacobs' skepticism about UG's physical form within the brain has been met in this paper with a discussion about weightedness of connections. His acknowledgement of preprogramming can be seen as connections that exist at birth. If UG is defined as these pre-existing connections, it becomes more than just a metaphor and addresses Jacobs' concerns.

9. Conclusion

There is obviously a need for more study of second language acquisition and the critical period combining both linguistic theory and neurological knowledge. Just as our vision of the mind is somewhat modular, so are our theories. The question that remains is whether it is possible to explain the behaviors of both the mind and the brain with one unified theory. Parallel Distributed Processing does this, explaining both the output of the mind and the structure of the brain with one theory. The compatibility of PDP with neurological information certainly illustrates the need to consider connectionism as a plausible model. Further development of Parallel Distributed Processing may provide even more convincing arguments for the value of combining the two directions of explanations into this one theory.

References

Bever, T. 1981. Normal acquisition processes explain the critical period for languagelearning. In K. Driller (Ed.), *Individual differences and universals in language learning aptitude* 176-198 Rowley, MA: Newbury House.

Felix, S.W. 1985. More evidence on competing cognitive systems. *Second Language Research* **1**:47-72.

Gasser, M. 1990. Connectionism and universals of second language acquisition.*Studies in Second Language Acquisition* **12**:177-199.

Hatch, E., Shirai, Y., and Frantuzzi, C. 1990. The need for an integrated theory: connecting modules. *TESOL Quarterly* **24**:697-714.

Jacobs, B. 1988. Neurobiological differentiation of primary and secondary language acquisition. *Studies in Second Language Acquisition* **10**:303-337.

Kaye, J. 1989. *Phonology: A cognitive view.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lachter, J. and Bever, T. 1988. The relation between linguistic structure and associative theories of language learning—a constructive critique of some connectionist learning models.

Cognition **28**:195-247.

Larsen-Freeman, D. and Long, M. 1991. *An introduction to second language acquisition research.* New York: Longman Group.

Lennenberg, E. 1967. *Biological foundations of language learning.* New York: John Wiley and Sons.

Newport, E.L. 1991. Contrasting conceptions of the critical period for language. In Carey and Gelman (Ed.) *The Epigenesis of Mind,* 111-130. Hillsdale, N.J.:Erlbaum Associates.

Schumann, J. 1990. Extending the scope of the acculturation/pidginization model to include cognition. *TESOL Quarterly* **24**:667-684.

Scovel, T. 1969. Foreign accents, language acquisition, and cerebral dominance.*Language Learning* **19**:245-253.

Seliger, H. 1978. Implications of a multiple critical period hypothesis for second language learning. In W. Ritchie (Ed.), *Second Language Acquisition Research.* NY: Academic Press.

Sokolik, M.E. 1990. Learning without rles: PDP and a resolution of the adult language learning paradox. *TESOL Quarterly* **24:**685-696.

Walsh, T. and Diller, K. 1981. Neurolinguistic considerations on the optimum age for second language learning. In Diller, K. (Ed.) *Individual Differences and universals in language learning aptitude.* Rowley, Massachusetts: Newbury House Publishers, Inc.