Assess the Critical Period Hypothesis in Second Language Acquisition

Lihong Du

International School, Changjiang Professional College

No. 918 Xiongchu Avenue, Wuhan 430074, China

Tel: 86-27-8717-0151   E-mail: joanna18@163.com

Abstract

The Critical Period Hypothesis aims to investigate the reason for significant difference between first language acquisition and second language acquisition. Over the past few decades, researchers carried out a series of studies to test the validity of the hypothesis. Although there were certain limitations in these studies, most of their results supported the hypothesis.

Keywords: The Critical Period Hypothesis, Pronunciation, Grammar, Semantic Content

1. Introduction

The ultimate attainment level of Second Language Acquisition (L2A) contrasts sharply with that of First Language Acquisition (L1A) (Birdsong, 1999, p.1). That is: various success versus universal success (Gleitman & Newport, 1995, p.1). This significant difference between L1A and L2A is explained by the Critical Period Hypothesis (CPH). The CPH states that for language acquisition, either first language (L1) or second language (L2), there is a critical period during which it is possible to achieve the same level as natives (Birdsong, 1999, p.1). Once this period is over, the ability to learn language declines (Johnson & Newport, 1989, p.61). In other words, the ultimate attainment level in L2A is determined to a great extent by the age of first exposure to a L2 (Birdsong & Molis, 2001, p.235).

The validity of the CPH has received much attention from numerous researchers. Four cases which are related to the CPH in L1A are "Isabelle" (discovered at age 6, which was considered within the critical period), "Genie" (discovered at age 13, which was beyond the critical period), "Chelsea" (discovered at age 31, which was far too late) and deaf people (Newport (1990) looked at the acquisition of American Sign Language) (for detail, see Gleitman & Newport, 1995, pp.11-15). However, because it is rather difficult to find adults like "Isabelle", "Genie" and "Chelsea" who have not learned a L1 early in life (Gleitman & Newport, 1995, p.13), most of the researchers dealt with the CPH of L2A.

In order to investigate the nature of the relation between the age at which the acquisition of a L2 commences and the ultimate attainment level of the L2 learners, researchers have suggested a number of hypotheses and conducted a lot of studies to test the ultimate attainment of L2 learners in pronunciation, grammar (including word structure and sentence structure) and semantic content.

In this essay, part of these theories and studies are first reviewed and compared, afterwards the CPH is evaluated.

2. Pronunciation

It seems that speaking L2 like natives is the most difficult aspect of L2A for L2 learners. In other words, foreign accent is unavoidable as a result of native language influence. In fact, it is universally agreed that the earlier L2 learners start to learn speech, the better their pronunciation is (Flege, 1999, p.101). However, researchers disagree on how the age of L2 learning relates to the degree of foreign accents as well as which specific factor contributes to foreign accent of the L2 learners.

The CPH means that the specific ability which allows L2 learners to pronounce L2 accurately is reduced or lost over the critical period (Flege, 1999, p.102). In other words, similar performance should be found among early learners. However, the findings of two recent studies, Flege, Munro, and Mackay (1995) and Yeni-Komshian, Flege and Liu (1997) were likely in contradiction to the CPH since they found a near-linear relation between Age of Arrival (AOA) and degree of foreign accents in two groups of participants (both early learners and late learners): 240 Italian-English bilinguals and 240 Korean-English bilinguals respectively (for detail, see Flege, 1999, pp.102-103).

2.1 The interaction hypothesis

Researchers try to seek other possible explanations for the existence of foreign accents since the CPH is widely accepted. There are three hypotheses, namely, *exercise hypothesis*, which implies that foreign accent increases as

AOA increases because L2 learners have stopped learning speech; *unfolding hypothesis*, which predicts that the fully developed L1 phonetic system leads to foreign accents; *interaction hypothesis* attributes foreign accents to the mutual influence of the different phonetic systems of L1 and L2 (Flege, 1999, p.105).

The interaction hypothesis was tested by Flege, Frieda, and Nozawa (1997) and Yeni-Komshian, Flege and Liu (1997). Their findings seemed to support the idea that L1 and L2 influence mutually (the interaction hypothesis), while disconfirmed the CPH. In detail, the former study found foreign accents among the participants who had learned English as young children and the relation between foreign accents and amount of L1 use. While the later study found only one bilingual out of 240 Korean-English bilingual participants (consisted of both early learners and late learners) pronounced both languages without a detectable foreign accent (Flege, 1999, pp.106-108).

2.2 The relation between production and perception

The particular L2 vowels and consonants produced inaccurately by L2 learners may account for foreign accents. One of the explanations for inaccurate pronunciation of particular vowels and consonants is Bever's psycho-grammar hypothesis, which proposes that L2 learners' ability to produce sounds according to their perceptual representation of the sounds is lost beyond the critical period (Flege, 1999, pp.108-110). In other words, L2 learners' perception and production ability are separated over the critical period. Another testable explanation is the Speech Learning Model (SLM), which suggests that because of the mutual influence of L1 and L2, L2 learners' perception of L2 sounds may be more accurate than their production (Flege, 1999, pp.108-109).

Oyama (1973) study tested Bever's hypothesis on the sentence level, he found that the participants who pronounced English more accurately can understand English sentences better (Flege, 1999, p.111). Consequently, his finding disapproved Bever's hypothesis. For there are possible factors which may affect the findings of Oyama's study, Meadow, Flege and MacKay (1997) replicated his study, they found same results (for detail, see Flege, 1999, p.111). Moreover, Flege, Bohn, and Jang (1997) study tested Bever's hypothesis by studying participants' perception and pronunciation of English vowels, they found that more accurate perceiver can produce vowels more accurately (Flege, 1999, pp.112-116). Similar studies were conducted on English consonants by Flege (1993) and Flege and Schmidt (1995), they found that participants' overall pronunciation of English and their comprehension were significantly correlated (Flege, 1999, pp.116-118). Obviously, the above findings conflicted with Bever's hypothesis.

Flege (1998a) tested the SLM, he found that the more experienced L2 learners' perception of L2 sounds is better than the less experienced L2 learners, while the two groups of English learners pronounced English sentences with equally strong foreign accents (Flege, 1999, pp.112-113). Undoubtedly, the results of his study supported the SLM.

2.3 Category formation for L2 sounds

According to the SLM mentioned above, due to the influence of L1, the possibility of establishing new categories for L2 vowels and consonants decreases as the age of learning L2 increases. As a consequence, the accuracy of producing L2 decreases (Flege, 1999, p.119). In other words, L2 learners would produce L2 vowels and consonants which are dissimilar to L1 sounds more accurately than those sounds which are very similar to their native languages.

The conclusions of two recent studies, Flege, Mackay and Meador (1998) and Flege, Schmidt and Wharton (1996) (for detail, see Flege, 1999, pp.119-124), confirmed the hypothesis that production accuracy of L2 sounds was constrained by category formation for L2 vowels and consonants.

3. Grammar

Apart from pronunciation, grammar is another aspect of L2 acquisition from which the researchers try to seek evidence to test the CPH.

3.1 Johnson & Newport (1989)

Johnson and Newport (1989) is one of the most influential studies which explored the correlation between the age of first immersion in a L2 and the ultimate attainment of L2 grammar.

The participants were 46 native Chinese or Korean speakers. Their age of arrival (AOA) in the United States was taken as the age of first immersion in English. These participants were asked to give grammaticality judgments for 276 sentences which were supposed to test their knowledge of English syntax (consisted of determiners, case, gender, number, particle movement, auxiliaries, yes/no questions, wh-question, word order) and morphology (such as past tense, plural, third person singular, present progressive, etc) (for detail, see Johnson & Newport,

1989, pp.68-77).

The participants were divided into two groups: *early arrivals* (the participants who arrived in the USA before age 15) and *late arrivals* (the subjects who arrived in the United States after age 17) (Johnson & Newport, 1989, p.69). It proved that AOA was the better measure than other experiential variables (length of exposure, amount of initial exposure, age of English classes, years of English classes and motivation to learn in classes) and attitudinal variables (identification, self-consciousness and motivation) (Johnson & Newport, 1989, pp.84-85).

Johnson and Newport found a linear decline relationship between AOA and performance among early arrivals that arrived in the USA after age 7. However, the researchers didn't find a similar trend among late arrivals, they found significant individual differences instead (Johnson & Newport, 1989, pp.77-80).

As dealing with the relation between AOA and grammar rule type, the researchers noticed that all of the rule types correlated significantly with the AOA of the participants. Specifically, acquisition of word order and the present progressive was significantly easy for all the participants regardless of their AOA, while determiners and plural morphology appeared to be the most difficult rule type for the late arrivals (Johnson & Newport, 1989, pp.86-89).

Johnson and Newport concluded that AOA was correlated with ultimate performance in the grammar of L2, and the relationship could be generalized to other L2 learners whose L1 and L2 were other languages than Korean or Chinese and English (Johnson and Newport, 1989, p.89, p.93).

3.2 Birdsong & Molis (2001)

One of the main findings of Johnson & Newport (1989) was that late L2 learners couldn't achieve nativelikeness despite what their L1s were. In doubt of this conclusion, Birdsong and Molis carried out another study in 2001 (Birdsong & Molis, 2001, p.238).

They chose 61 native speakers of Spanish and used almost the same procedures and materials (274 sentences instead of 276, since one pair of sentences was considered not appropriate) as Johnson & Newport (1989) study (Birdsong & Molis, 2001, p.239).

By comparison with Johnson & Newport, Birdsong and Molis found different results. That is, they didn't find a linear decline relationship between AOA and performance among early arrivals, while they found considerably age effect on performance among late arrivals (Birdsong & Molis, 2001, pp.239-240).

In summary, the overall performance of the participants was significantly better than that of the subjects of Johnson & Newport (1989) (Birdsong and Molis, 2001, p.243, p.246). In other words, Spanish speakers performed better than Korean or Chinese speakers. Moreover, Birdsong and Molis did find one late arrival whose score fell within the range of natives (Birdsong & Molis, 2001, p.244).

Birdsong and Molis pointed out that one factor could artificially influence subjects' performance: whether or not the materials employed in the study fully represented the grammar of the target language (Birdsong & Molis, 2001, p.245). As to the generalizability claimed by Johnson and Newport (1989), Birdsong and Molis argued that participants' L1s did affect their ultimate attainment of L2 learning based on the considerable difference in performance of the participants in two studies (Birdsong & Molis, 2001, p.246). Thus, Birdsong and Molis concluded that the above results should be restricted to this study only (Birdsong & Molis, 2001, p.246).

3.3 Birdsong (1992)

Because there are exceptional late L2 learners who appear to perform like natives, researchers try to find out whether there are competence differences in grammar between native speakers and those exceptional L2 learners as well as in which aspects of grammar they differ (Birdsong, 1992, p.707).

The Coppieters (1987) study was the most representative study in answering the two questions mentioned above. Coppieters chose 20 native French speakers (NS) and 21 near-native speakers of French (NNS) with a variety of L1 backgrounds. The subjects were asked to give grammaticality judgments to 107 sentences which supposed to reflect grammatical rules of French. After comparing NS' performance with that of NNS, Coppieters concluded that there were competence difference between two groups of subjects and their competence differences could be divided into two categories: Universal Grammar (UG) type structures and functional or cognitive aspects of grammar (Birdsong, 1992, pp.708-711).

However, other researchers pointed out numerous flaws in Coppieters' (1987) study (i.e. the materials, the methodology used in the study as well as the composition of the subject groups) (for detail, see Birdsong, 1992, pp.711-716). Thus, Birdsong replicated this study to reexamine the two main questions mentioned at the beginning of this section in 1992.

Birdsong chose 20 near native speakers of French whose L1 was English (ENS) and 20 native speakers of French (FNS). The subjects were asked to give grammaticality judgments to 76 French sentences (included only appropriate variables in Coppieters' study, and new structures which represented more grammatical rules) (Birdsong, 1992, pp.717-720).

Unlike Coppieters, Birdsong found that although ENS' performance differed significantly from FNS' performance, some ENS' (15 out of 20) scores fell within the range of FNS (Birdsong, 1992, pp.723-724). Obviously, the results suggested that Birdsong's subjects performed considerably better than the NNS of the Coppieters (1987) study and the participants of Johnson & Newport (1989) study (Birdsong, 1992, p.724).

In addition, the different performance between FNS and ENS in respect of different grammatical variables didn't seem to fall into significant types: UG structures and the structures which were outside the domain of UG (Birdsong, 1992, p.740).

Birdsong was of the opinion that AOA of ENS predicted the differences between the performance of the ENS group and the FNS group. In other words, there was a significant effect of AOA on the performance of ENS group since all of the ENS subjects were immersed into French after puberty (Birdsong, 1992, p.741).

However, because of the existence of the limitations of this study (the materials employed in this study couldn't fully reflect the grammar, the subjects only represented one L1 background) (Birdsong, 1992, pp.742-743), Birdsong failed to find concrete answer for the two questions mentioned at the beginning of this section.

4. Semantic content

It was hypothesized that functional subsystems which were related to processing semantic content of the sentence and structure of the sentence were affected differently by delays in first immersion in L2 (Weber-Fox & Neville, 1999, pp.23-24).

This hypothesis was tested by utilizing a large group of Chinese-English bilinguals (for detail, see Weber-Fox & Neville, 1999, pp.24-33). It was found that although the age of immersion in a L2 still appeared to be a critical factor which determined the linguistic competence of L2 learners, the late learning groups' specific linguistic competence (syntactic competence and semantic competence) seemed to be differentially affected by that factor. Specifically, the late learners seemed to be much slower in interpreting the grammatical content of the sentences than the semantic content of the sentences (Weber-Fox & Neville, 1999, p.33). Similar divergence was found in processing closed-word class words (such as nouns, verbs, adjectives, related to semantic content of the sentence) and open-class words (such as determiners, articles, related to sentence structure) (Weber-Fox & Neville, 1999, pp.33-35). Furthermore, by analyzing averaged Event-Related Brain Potential (ERP) waveforms over left and right parietal sites for monolinguals and bilingual groups (Weber-Fox & Neville, 1999, pp.30-32), the researchers found evidence from electrophysiology which supported this hypothesis.

Obviously, the above findings backed up the hypothesis mentioned at the beginning of this section.

5. Discussion and conclusion

Overall, it seemed that as for pronunciation, the interaction hypothesis and the Speech Learning Model ran counter to the CPH, while the Category Formation hypothesis backed up the CPH (Flege, 1999); as for grammar, Weber-Fox and Neville (1999) study (they proved the hypothesis that functional subsystems related to semantic and grammar processing were differentially affected by late immersion in L2) (Weber-Fox & Neville, 1999), Johnson and Newport (1989) study, Birdsong and Molis (2001) study and Birdsong (1992) study all supported the CPH; as for semantic content, Weber-Fox & Neville (1999) study provided evidence for the hypothesis that functional subsystems related to processing semantic content and grammatical structure were affected differently by delays in L2A (Weber-Fox & Neville, 1999).

Obviously, most of the studies confirmed the CPH. However, there are considerable differences in these studies.

To begin with, the researchers didn't agree on the exact cutoff age. In detail, the age taken as puberty in Flege et al (1995) was 15, while the cutoff ages in Johnson and Newport (1989) and Birdsong and Molis (2001) were the same: 16 years old. Furthermore, Bialystok and Hakuta suggested that if the cutoff age in Johnson and Newport (1989) was postponed to 20, better results would be achieved (a linear decline relationship between AOA and the participants' scores would be significant for early arrivals as well as late arrivals) (Birdsong & Molis, 2001, p.241).

Second, the L1-L2 paring of the participants were different, which affected the results of the studies as claimed by Birdsong and Molis (Birdsong & Molis, 2001, p.246). Specifically, the participants in Johnson and Newport (1989) were English learners who were Korean or Chinese native speakers, while the subjects in Birdsong and

Molis (2001) study and Birdsong (1992) study were Spanish English learners and English French learners respectively.

Furthermore, the most considerable divergence among these studies was the results of Birdsong and Molis (2001) and Johnson and Newport (1989), say, Johnson and Newport found the linear decline relationship between AOA and the subjects' scores only among the early arrivals (AOA>7), while Birdsong and Molis found the same correlation only in the group of late arrivals. As Birdsong and Molis (2001) used the same materials, procedures and analyses as Johnson and Newport (1989), the only difference was in the subjects' L1 (Birdsong & Molis, 2001), different L1s seemed to lead to divergence in ultimate attainment of L2A. However, it was inappropriate to conclude that different L1s was the only factor which resulted in the marked difference in the results of the two studies.

Finally, as Birdsong and Molis pointed out, whether or not nativelikeness could be found among late L2 learners as well as how many of these exceptional learners was enough were crucial factors that could disconfirm the CPH (Birdsong & Molis, 2001). Johnson and Newport (1989) and Flege et al. (1995) didn't find any participants who started to learn L2 after puberty scored within the range of natives (Johnson & Newport, 1989; Flege, 1999, p.104), while Birdsong and Molis (2001) and Birdsong (1992) found subjects among late learners that achieve nativelikeness (one and fifteen respectively) (Birdsong & Molis, 2001; Birdsong, 1992). In a word, no agreement was reached about nativelikeness in determining the truth of the CPH so far.

On the other hand, there were a number of limitations in these studies.

First of all, none of the researchers had considered quality of input when they carried out these studies to test the CPH in pronunciation (foreign accents) (Flege, 1999). However, as for pronunciation, quality of input was probably a rather crucial factor which determined the ultimate performance of L2 learning in pronunciation (with foreign accents or not) since it was more difficult for late L2 learners who had already learned speech in their home country (their teachers spoke English with foreign accents) before they arrived in the United States to speak like natives than for those who hadn't learned any English before.

In addition, the influence of methodological issues on L2A is outstanding (Han, 2008, p.171). The above involved studies were carried out at different time, which might have accounted for different results of the studies. Take Johnson and Newport (1989) and Birdsong and Molis (2001) for example, as Birdsong and Molis conducted their study 12 years later than Johnson and Newport, better teaching methods might have been employed in teaching their participants due to the widely accepted rapid development in L2 teaching methodology.

Last but not least, none of the researchers have carried out a study to investigate the ultimate performance of L2 learners in pronunciation and grammar at the same time and this kind of study could be rather valuable for the researchers who are exploring the validity of the CPH and the nature of the effect of the CPH on L2 learning.

In conclusion, taking account of the differences among these studies stated above as well as the limitations in every study, it was possibly unreasonable to assess the CPH by comparing different studies so far.

References

Birdsong, David. (1992). Ultimate Attainment in Second Language Acquisition. *Language*, 68. 706-705.

Birdsong, David. (1999). Whys and Why Nots of the Critical Period Hypothesis for Second Language Acquisition. In D. Birdsong (ed.) *Second Language Acquisition and the Critical Period Hypothesis*. Lawrence Erlbaum Publishers. pp. 1-21.

Bridsong, David. and Molis, Michelle, (2001). On the Evidence for Maturational Constraints in Second Language Acquisition. *Journal of Memory and Language*, 44. 235-249.

Flege, James E. (1999). Age of Learning and Second Language Speech. In D. Birdsong (ed.) Second *Language Acquisition and the Critical Period Hypothesis*. Lawrence Erlbaum Publishers. pp. 101-131.

Gleitman, Lila R. and Newport, Elissa L. (1995). The Invention of Language by Children: Environmental and Biological Influences on the Acquisition of Language. In L. Gleitman and M. Liberman (eds.) *An Invitation to Cognitive Science: Language*. MIT Press. pp. 1-24.

Han, Zhaohong. (2008). *Fossilization in Adult Second Language Acquisition* (1st ed.) Foreign Language Teaching and Research Press. pp. 44-176.

Johnson, Jacqueline S. and Newport, Elissa L. (1989). Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology*, 21,

60-99.

Weber-Fox, Christine M. and Neville, Helen J. (1999). Functional Neural Subsystems Are Differentially Affected by Delays in Second Language Immersion: ERP and Behavioral Evidence in Bilinguals. In D. Birdsong (ed.) *Second Language Acquisition and the Critical Period Hypothesis*. Lawrence Erlbaum Publishers. pp. 23-38.