

The Critical Period Hypothesis and Second Language Acquisition: A Critique of Two Empirical Studies

Ivan B. BROWN

Joetsu University of Education

Abstract

The existence of a critical period in second language acquisition has been supported and disputed by a number of scholars and is an issue with wide-ranging implications such as the notion of a universal grammar and second language learning. It has often been defined as the period when the Language Acquisition Device (LAD) is available to a child. However, there have been a variety of conceptions of the critical period, some of which discriminate between effects for different linguistic domains, while others discriminate between various explanations for reduced levels of L2 attainment among older learners. This paper reviews two consecutive and closely related empirical studies. Several aspects of the studies, including their motives, assumptions, methodology, replicability, results and conclusions are scrutinized, and salient points are proposed for future research.

Key Words:

Critical Period Hypothesis	grammaticality judgment
phonology	vocabulary
morphosyntax	ESL EFL

The Studies:

Flege, J.E., Yeni-Komshian, G.H. and Liu, S. (1999)
Age Constraints on Second-Language Acquisition.
Journal of Memory and Language 41, 78-104.

Lin, H.L., Chang, H.W. and Cheung, H. (2004)
The effects of learning on auditory perception of English minimal pairs
by Taiwan University Students.
Journal of Psycholinguistic Research 33(1), 25-49.

Background to the Studies

The critical period hypothesis (CPH) is an issue with far reaching implications for the notion of a universal grammar, psycholinguistics and second language learning. Jackendoff (2002) cited Lenneberg (1967) as defining the critical period as when the Language Acquisition Device (LAD) is available to a child. During this period, the child is able to make use of the LAD as an innate tool to learn its native language and perhaps even a second language. After the critical period, often regarded to end at about twelve years old, the LAD is no longer available and the older child or adult has to make use of more general cognitive logical reasoning faculties to learn a language. In summary, whether or not the critical period exists or not is generally regarded to be a major factor in determining whether or not an innate LAD and universal grammar exist or not.

There are further specific issues related to the CPH. Newport (1990, cited in Jackendoff, 2002) described this phenomenon as not being a sudden cut-off point but an inverse correlation between the age at which one starts to learn the language and ultimate proficiency in that language. In addition, most discussion about universal grammar has concerned first language acquisition, but there has also been a lot of discussion about whether adult learners of a second language use a universal grammar. On the other hand, Jackendoff argued that much of this discussion is based on the misconception that the LAD is a singular entity that cannot be broken down. Instead, it should be thought of as a collection of capacities, some vulnerable to critical period effects, and some not. Jackendoff (2002:263-264) lists the following linguistic domains as not being effected by a critical period: vocabulary, the concatenation of words and simple semantically based principles of word order and the following as having critical period effects: phonology, phrase structure and the inflectional system. Eubank and Gregg (1999) expressed the similar view that the notion of language as a singular entity as an epiphenomenon is not appropriate for scientific enquiry.

Skehan (1998) noted some important practical implications of the critical period hypothesis: realistic expectations of what is achievable in L2, whether L2 should be in the early school curriculum or not, and the investigation and selection of appropriate instructional methodologies. Can there be a critical period for acquisition of a second language as well as first language? One of the most seminal empirical studies specific to this issue was conducted by Johnson and Newport (1991). Accordingly, the relevant differences between first and second language acquisition include the following: (1) An L2 is more often embarked upon at a more advanced developmental stage; and (2) L2 learners already have linguistic knowledge of L1. Additionally, there may be crucial differences in how consecutive bilinguals process language. Hamers and Blanc

(2000) described four possible strategies for aural sentence decoding (e.g., word order for English speakers versus animacy for Italian speakers): differentiation (separate strategies for L1 & L2); forward transfer (L1 to L2); backward transfer (L2 to L1); and amalgamation (a single strategy).

Birdsong (1999:5-6), in fact arguing against the CPH, points out that there is actually no single CPH, but several based on different explanations, such as: (1) decrease in neural plasticity; (2) loss of access to LAD; (3) maladaptive gain or maturation; (4) "use it then lose it" (LAD is unnecessary after critical period and thus dismantled by the brain); (5) "use it or lose it" (the LAD is dismantled if not exercised); and (5) "learning inhibits learning" (the connectionist hypotheses). Bialystok and Hakuta (1999), also arguing against the critical period hypothesis, likened age to an intervening variable in a design, where apparent age effects are in fact caused by other variables related to age and not age itself. This is a key issue in the two experimental studies reviewed here.

1) Flege, Yeni-Komshian & Liu (1999)

This study was motivated by similar potential theoretical and practical implications as those described above such as educational policy for example. Disparate and divergent results of previous studies were cited for age effects on L2 acquisition, so the aim was to evaluate the critical period hypothesis (CPH), in relation to specific linguistic domains. The study also set out to scrutinise several variables potentially confounded with age of onset.

The experimental method involved selecting 240 native speakers of Korean who differed according to age of arrival (AOA) in the USA but had all lived there for at least eight years, and were generally organised into 10 sub-groups (K groups) according to AOA. They completed a language background questionnaire, their pronunciation of English was evaluated by native speaker listeners (NEs), and they took a grammaticality judgement test (GJT) based on that devised by Johnson and Newport (1989) but with added features such as a separate rule-based set and a lexically based set. In the foreign accent evaluation procedures, it was interesting to note that the recordings of the Koreans' speech were made with a sampling frequency of 22.05 kHz; approximately half the sampling rate of that found in normal commercial CD sound quality. Space does not permit a detailed speculation of how this sub-standard fidelity would influence the speech accent evaluations, but it was probably related to a need for economy of processing time and data storage.

The overall effects of AOA on foreign accent ratings and GJT scores were similar (p.86). However, all 10 K groups differed significantly from the NEs for

phonology, whereas only AOAs 7-23 (not 2-6) differed in GJT scores. The relationship between AOA and foreign accent appeared to be linear near the end of the critical period, which failed to provide support for the CPH for phonology (p.87). The relationship between AOA and GJT scores was non-linear at the age of about 12, which supports a CPH for morphosyntax (p.88). The scores for grammatical sentences were generally higher than for ungrammatical sentences. Possible causes of this suggested by Flege *et al.* were a general bias towards giving “Yes – grammatical” answers and less grammatical determinacy in adult learners. However, there is arguably a potential statistical validity problem with a multiple-choice question with only two possible answers, in that there seems to be a significant statistical likelihood (theoretically 50%, if pure chance) of a participant achieving a correct response by default. This issue will be discussed further in the last section of this review.

GJT scores for rule-based and lexical-based sets with AOA of less than 12 did not differ significantly, whereas those with AOAs greater than 12 had significantly lower scores for lexical than for rule-based sets. Flege *et al.* suggest that this may be due to “bottom-up” acquisition in contrast to generalisable rules, which can be learned through general reasoning through formal education. Although lexically-influenced phrase structures are not the same entity as vocabulary, this result seems to be incongruent with Jackendoff’s statement that there are no critical period effects for vocabulary, since vocabulary and lexically influenced phrase structures are both associated with the lexicon.

The study considers different explanations of why there should be more of an AOA effect on phonology than on morphosyntax, and whether acquisition of phonology and lexical aspects of morphosyntax share anything in common. It is plausible that both phonology and lexical aspects of morphosyntax require a bottom-up, data-driven approach. They also conjecture whether the presence of native speaker input has a significant effect. Although foreign accents grew stronger and scores on the grammaticality judgement test decreased as AOA increased, underlying bases of these effects differed importantly. When confounding variables were controlled, AOA effects on phonology but not morphosyntax remained significant.

With regard to validity, some sentences in the GJT may not have tested the intended grammatical structure or rule (p.89), where ungrammaticality could have been attributed to more than one part of the sentence. Further possibilities for assessing morphosyntax will be discussed in the final section of this review. Regarding the correlation between self-assessed language use and L2 performance, this may not necessarily be a frequency of use effect on

performance. Relatively infrequent use of L2 may be an effect of poor performance in L2, rather than its cause.

2) Lin, Chang and Cheung (2004)

This study seemed to be motivated substantially by the authors' views of the state of and recent trends in language education in Taiwan, in particular the associated educational policies and commercial pressures. They raised strong objections to the financial burdens imposed on language students and tax payers by profit-seeking language schools, expensive immersion programmes and unqualified native speakers.

In concordance with Flege *et al.*, they cite disparate results from previous studies, including that of Flege *et al.*, and argued that in many studies, there had been misconceptions about second language acquisition pertaining to: (1) acquisition speed as opposed to ultimate attainment; (2) age differences and neurobiology; and (3) a lack of emphasis on cases in which adults master L2. They also argued that there was a need for more research into the CPH with regard to specific linguistic domains, and accordingly, their study focuses on auditory perception.

Lin *et al.* state that Flege *et al.* found no age effects on morphosyntax, presumably referring to that study's finding that only effects on foreign accent remained significant when confounded variables were controlled. Moreover, they argue that findings on adults' speech accents have been inconclusive, perhaps partly alluding to the fact that although Flege *et al.* still found AOA effects on foreign accent ratings when other variables were controlled in the matched sub-group analyses, their discontinuity test failed to support the CPH for phonology, and their pre/post correlation test also failed to support CPH. Their finding of AOA effects on foreign accent also depended on the validity of the matched sub-group analyses.

In fact, Lin *et al.* criticize the use of matched sub-group analysis by Flege *et al.*, arguing that it is "not an ideal method to control all possible confounding variables," (p.27) since other variables can still be left uncontrolled. They maintain that variables confounded with AOA will always be an inescapable part of research into CPH and L2 acquisition in the milieu of English as a second language, since the amount of exposure to the English language that participants experience in an English-speaking country cannot be controlled. They make a case for researching CPH and second language acquisition in an English-as-a-Foreign-Language (EFL) milieu. They maintain that in EFL, "English learning is

highly constrained in the classroom and often not mandatory,” (p.27). Therefore, EFL in Taiwan is ideal milieu where “it is possible to separate different effects such as length of learning and starting age,” (p.27). However, surely even EFL learners in non-English speaking countries are exposed to a considerable amount of English outside the classroom through the media, games, imported products and tourists, etc. This is not necessarily a valid assertion.

Although Lin *et al.* state that previous findings on age and speech accent are inconclusive, and give a detailed account of the distinction between the continuous and categorical modes of auditory perception, they do not give a clear argument of why they chose to examine auditory perception, or precisely which mode (or combination of modes) of it they aim to test. Their argument seems to imply a supposed link between the ability to distinguish between phonemes that are not in an adult learner’s native language, and that adult’s ability to produce these phonemes in speech. While there is indeed likely to be a relationship between the two, this relationship is demonstrably not necessarily a simple one.

The experiment in this study involved testing Taiwan university students’ auditory perception of isolated English minimal pairs with and without noise background to obtain the intervallic variable SCORES. The auditory test was rigorously constructed with random distribution of groups of three words, each including a double of one of the words of a minimal pair, so that participants had to find the odd word. The participants were also interviewed on their early experiences in EFL. Data from this was collected to measure the following intervallic and ratio variables: age of formal or informal exposure to the English alphabet (ALPHAGE); age of onset of formal EFL education (FORMALAGE); age of immersion experience (IMMERAGE); number of years of immersion experience (YRSIMM); and number of years of formal EFL learning (YRSFORMAL). Whether or not the participants had been taught by a native speaker or not was also measured as a nominal variable.

Some of the questions asked of this data seemed to be misconstrued or be statistically invalid. For example, having collected intervallic/ratio information about previous immersion experience, this was then forced into being a nominal variable concerning whether or not students had had immersion experience or not (questions on p.33; table of results on p.35). They do not state a threshold or cut-off point for classifying students as with or without immersion experience. Given that language immersion can take place with varying levels of intensity (e.g., hours per day) and for varying lengths of time, this would seem to produce unreliable results. The same issue could apply to the nominal variable “with or without native speaker instructor(s)” in the same analysis.

The test with no background noise resulted in a ceiling effect; therefore, a test with white background noise was also conducted. This produced the necessary difficulty but also resulted in an interesting age effect on perception of consonants, especially consonant clusters containing the liquid /r/ and /l/ phonemes. This effect on consonants is actually unsurprising if one considers that both white noise and consonants (by varying degrees, due to air turbulence caused) have by nature inharmonic spectra, and superimposition of the two would result in blurred distinctions. Vowels, on the other hand, have more harmonic spectra, and would tend to stand out in relief against a white noise background.

In their conclusion, they related their findings to the original aims, and noted some potential problems in their research design, for example that confounded auditory cues (phonological patterns even in isolated words) could have had some influence in the participants' scores.

Implications for Future Studies

These experiment designs and their findings raise a number of issues. Fortunately, both articles described the experimental procedures in enough detail to be replicated or adapted in the future. One of the main issues concerned the method of assessing acquisition of morphosyntax. Flege *et al.* did this with a grammaticality judgement test that consisted of pairs containing one grammatical sentence and one ungrammatical sentence. Another possible approach would involve questions that consisted of perhaps three or four sentences, with perhaps only one grammatical one or one ungrammatical one. Participants would then choose the appropriate sentence. This would accommodate more distracters, and therefore more statistical validity. The challenge in this would be how to select the group of sentences in order to ensure that the question was targeting a specific feature of morphosyntax, to enable detailed analysis of the scores according to specific features of morphosyntax.

Another issue is the variety of milieux, such as ESL verses EFL settings. Perhaps a kind of meta-study, or preliminary study of environmental conditions could be conducted to quantify effects in these milieux. Longitudinal studies of a number of students in each context could be conducted, with the participants keeping structured diaries of their learning and exposure. This kind of study probably would not be statistically rigorous enough to prove or reject hypotheses, but could generate hypotheses and inform experimental design.

Other linguistic domains could also be investigated: prosodic auditory perception and even pragmatic or discursive competence, for example. Needless

to say, serious challenges would lie in selecting criteria and frameworks for assessment and controlling variables. Participants could also be matched according to performance in one domain and measured in another to test for correlations and relationships between the different areas of acquisition.

References

- Bialystok, E. & Hakuta, K. (1999) Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D.Birdsong (ed.) (1999).
- Birdsong, D. (ed.) (1999) *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.
- Birdsong, D. (1999) Introduction: Whys and why nots of the critical period hypothesis for second language acquisition. In D.Birdsong (ed.) (1999).
- Eubank, L. & Gregg, K.R (1999) Critical periods and (second) language acquisition: divide et impera. In D.Birdsong (ed.) (1999).
- Flege, J.E., Yeni-Komshian, G.H. & Liu, S. (1999) Age Constraints on Second-Language Acquisition. *Journal of Memory and Language* 41, 78-104.
- Hamers, J.F. & Blanc, M.H.A (2000) *Bilinguality and Bilingualism*. 2nd Ed. Cambridge: Cambridge University Press.
- Hinton, P.R. (1995) *Statistics Explained: A Guide for Social Science Students*. London: Routledge.
- Jackendoff, R. (2002) *Foundations of Language*. New York: Oxford University Press.
- Johnson, J. & Newport, E. (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of ESL. *Cognitive Psychology* 21, 60-99.
- Johnson, J. & Newport, E. (1991) Critical period effects on universal properties of language: The status of subadjacency in the acquisition of a second language. *Cognition* 39, 215-258.
- Lenneberg, E. (1967) *Biological Foundations of Language*. New York: John Wiley and Sons.
- Lin, H.L., Chang, H.W. & Cheung, H. (2004) The effects of learning on auditory perception of English minimal pairs by Taiwan University Students. *Journal of Psycholinguistic Research* 33(1), 25-49.
- Newport, E. (1990) Maturational constraints on language learning. *Cognitive Science* 14, 11-28.
- Skehan, P. (1998) *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.