The Critical Period Hypothesis for Second Language Acquisition: Tailoring the Coat of Many Colors

David Birdsong

Abstract The present contribution represents an extension of David Singleton's (2005) *IRAL* chapter, "The Critical Period Hypothesis: A coat of many colours". I suggest that the CPH in its application to L2 acquisition could benefit from methodological and theoretical tailoring with respect to: the shape of the function that relates age of acquisition to proficiency, the use of nativelikeness for falsification of the CPH, and the framing of predictors of L2 attainment.

1 Introduction

David Singleton's (2005) study, "The Critical Period Hypothesis: A coat of many colors", is the second most-cited article ever to appear in *International Review of Applied Linguistics in Language Teaching*. At its core, the piece is a critique of the Critical Period Hypothesis (CPH) as it has been applied in the context of second language acquisition (L2A). Singleton argues that, as an account of constraints on L2A attainment, the CPH is underspecified in the literature. Crystallizing the sometimes vague and decidedly diverse positions advanced by researchers in the CPH tradition, Singleton (2005: 280) writes: "For some reason, the language acquiring capacity, or some aspect or aspects thereof, is operative only for a maturational period which ends some time between perinatality and puberty".

With respect to the notion of 'period', Singleton notes that various researchers have pegged the end of the CP for phonetics/phonology at ages ranging from one year to puberty. As for the affected language learning capacities, Singleton's review of the literature reveals that CP researchers have put forth accounts of deficits in: general language learning ability, non-innate linguistic features, innate

D. Birdsong (✉)
University of Texas at Austin, Texas, USA
e-mail: birdsong@austin.utexas.edu

linguistic features, specific subparts of innate features, and implicitly acquired linguistic features. As concerns the underlying sources of CP effects, Singleton's survey tallies six accounts of a neurobiological nature, four in terms of cognitive development, and four relating to affect and motivation.

Singleton (2005: 280) characterizes with trademark pithiness his notion of 'the manifoldness' of the CPH:

> My conclusion from this exploration is that the CPH cannot plausibly be regarded as a scientific hypothesis either in the strict Popperian sense of something which can be falsified (see, e.g. Popper 1959) or indeed in the rather looser logical positivist sense of something that can be clearly confirmed or supported (see, e.g. Ayer 1959). As it stands it is like the mythical hydra, whose multiplicity of heads and capacity to produce new heads rendered it impossible to deal with.

From Singleton's perspectives on the CPH/L2A, "a coat of many colors" is indeed an apt metaphor.

The present contribution piggybacks on Singleton's work, taking complementary perspectives on mainstream research conducted in service of the CPH/L2A. Adding a fitting metaphor to Singleton's original title, I attempt to show that the coat of many colors might warrant some methodological and theoretical tailoring to accommodate the facts and phenomena associated with age and attainment in L2A.

2 What Critical Periods Look Like

To make a case for a CP in the L2 context, it does not suffice to demonstrate that age of onset of L2 learning (often referred to as *age of acquisition* or *AoA*) and ultimate L2 attainment are related. To qualify as a period, the geometry of the function relating AoA to performance (usually characterized in terms of linguistic proficiency or processing ability) should contain a slope that is bounded at some points along the function.

Many studies have found AoA effects over the full span of AoA's, suggesting unbounded functions (Birdsong 2005). Conversely, non-linearities or inflections in the AoA-attainment function have been interpreted as suggestive of a period, in the sense that changes in slope would mean that AoA-related effects are bounded (Hakuta et al. 2003; Stevens 2004). The logic here is that a significant slope change would be consistent with a qualitative change in sensitivity of the learning mechanism. To suggest that maturational effects are at play, the changes in slope should line up with recognized developmental milestones that are uncontroversially maturational in nature.

In this context, Birdsong and Molis (2001) reanalyzed the L2 proficiency data from Johnson and Newport's (1989) study of Korean and Chinese learners of L2 English. Using a piecewise linear regression model, the reanalysis placed the breakpoint in Johnson and Newport's AoA-proficiency slope at 18 years, i.e. at an AoA beyond puberty. Similarly, Vanhove (2013) applied piecewise regression

analyses to DeKeyser et al.'s (2010) data from Russian immigrants learning L2 English in North America and L2 Hebrew in Israel. Vanhove's reanalysis of DeKeyser et al.'s Hebrew grammaticality judgment results revealed that including an inflection point in the AoA-attainment function did not result in a better fit than a simple linear regression model. In other words, AoA effects were best modeled as a straight-line function, across the full range of AoA. The reanalysis of the English grammaticality judgment results revealed that a model with a breakpoint at around AoA = 16 was a marginal improvement over a simple linear model. However, like the Hebrew data, the slope of the function after AoA = 16 did not flatten, i.e. a decline in performance continued throughout adult AoA.

Vanhove's study suggests that piecewise regression models, which have been used only infrequently in L2 attainment studies, are appropriate for determining whether the timing and geometry of the AoA-attainment function conform to assumptions of what a CP should look like.[1] Made-to-measure analytical methods may be required to suitably fit the coat to the function.

3 Nativelikeness and the CPH/L2A

Long (1990) stipulates that the way to falsify the CPH in the L2A context would be to find a single late learner who is indistinguishable from an adult monolingual native. The operational logic goes something like this: the absence of observed nativelikeness is due to maturational factors, and nativelikeness can disconfirm the CPH/L2A.

On a complementary view of non-nativelikeness, many researchers point out that non-monolingual-likeness in both the L1 and the L2 is a defining characteristic of bilingualism (early and late) (for a review, see Ortega 2009: 26–27). For example, VOT values of the L1 may extend toward those of the L2, just as VOT values of the L2 may extend toward those of the L1 (see e.g. Fowler et al. 2008). Among bilinguals, effects of maturation (in the sense of biologically determined declines in learning ability) cannot straightforwardly explain the fact that syntax, lexicon, and phonology of the *L1* are altered in bilingualism, and have features reflecting contact with and use of the L2 (see e.g. Cook 2003). Non-monolingual-nativelikeness in the

---

[1] Granena and Long (2013) applied multiple linear regression analyses to the relationship of Chinese natives' AoA to their attainment in L2 Spanish morphosyntax, phonology, and lexis and collocation. For each of these three linguistic domains, including breakpoints in the model revealed a small (5 %) but statistically significant increase in variance accounted for, as compared to the variance accounted for in a model with no breakpoints. According to the authors, the fact that the improvement was so small "could mean that the less complex (i.e. more parsimonious) model with no breakpoints is already a good enough fit to the data or, alternatively, that a larger sample size is needed to compensate for the loss of degrees of freedom and to minimize the risk of overfitting" (2013: 326–327).

L1 of bilinguals cannot be due to maturationally induced impairment of a presumed language learning mechanism, inasmuch as the L1 has been fully acquired, before the end of maturation.

Arguably, the fact that the L1 can be influenced by the L2 in adulthood is evidence for maturationally conditioned representational *plasticity*. In other words, non-monolingual nativelikeness in the L1 is suggestive of a capacity to learn language in adulthood. For example, 'speaking with an accent in the native language' is common among immigrants returning to their homeland for visits, as are noticeable changes of accent among individuals who move across dialect boundaries within a single country. Such permeability of the L1 would not be possible if the neural systems underlying phonetic perception and production were not plastic. To fully clothe the big-picture facts about late L2 and late L1 learning, the CPH/L2A coat might benefit from some broadening through the shoulders.

4 Scrutiny Across the Board

According to Long (1990) and Hyltenstam and colleagues (e.g. Hyltenstam and Abrahamsson 2003; Abrahamsson and Hyltenstam 2009), there are two key elements of the linkage of nativelikeness to the CPH/L2A. One is the requirement that the nativelikeness in the L2 must be observed 'across the board', that is, with respect to all L2 linguistic features and processes, for it to be sufficient to falsify the CPH. The other is that the evidence for (non)-nativelikeness (be it, presumably, behavioral or brain-based) should be uncovered from close scientific scrutiny, lest some evidence be overlooked. Thus, on this view, an individual who appears nativelike to the casual observer or on coarse or too-easy performance measures is insufficient evidence for rejecting the CPH. In sum, falsification of the CPH/L2A would require 'scrutinized nativelikeness' (Abrahamsson and Hyltenstam 2009) on a comprehensive set of linguistic measures.

There is a sensible rationale for psycholinguists to look beyond what is noticed by the untrained ear. With sensitive measures, our understanding of linguistic behaviors—especially inter-group and inter-individual differences—is enhanced. In the L2 context, as in scientific inquiry generally, the precision of information available from granular observation is valuable and welcome. From this perspective, there is no argument with scrutiny. The concern is with the application of evidence for non-nativelikeness—be it obtained by scrutiny or by any other methodological orientation—to theory. Monolingual-bilingual differences are inevitable, and more differences are sure to emerge from challenging tasks and fine-grained analyses than from simple tasks and coarse analyses. But it is not clear that non-monolingual-like behaviors and brain functions are decisive for CPH/L2A theory. Given what is known about reciprocal L1-L2 influences in bilinguals' behaviors, evidence for non-nativelikeness—be it detected on the street or under microscopic examination, be it present in outer patches or inner pockets, in bolts of cloth or in buttonholes—does not compel, uniquely, a maturational explanation.

And so it is with across the board nativelikeness. Since bilinguals are not like monolinguals in either of their languages, it is hard to argue that comprehensive nativelikeness, scrutinized or not, should be held up as the gold standard for falsifying the CPH/L2A.

If the idea is to look around for non-nativelikeness in bilingualism, then non-nativelikeness will eventually be found. If the follow-on idea is to stipulate that across-the-board nativelikeness is what is required to disconfirm the CPH, then the CPH is invulnerable to falsification. This being the case, the coat would need some letting out in the chest to accommodate the Kevlar vest underneath.

5 Framing the Issues

A study by DeKeyser (2000), entitled "The robustness of critical period effects in second language acquisition", investigates the roles of factors such as AoA, language learning aptitude, and years of schooling in predicting L2 English grammaticality judgment (GJ) accuracy by 57 Hungarian immigrants to the US. A look at each of these factors in turn is revealing.

- *AoA*. For all participants, AoA was predictive ($r = -0.63$, $p < 0.001$). On the other hand, breakout correlations with groups divided by early arrivals (AoA < 16; n = 15; $r = -0.26$ ns), and late arrivals (AoA 17–40; n = 42; $r = -0.04$ ns), revealed no significant declines at either pre- maturational or post-maturational AoA epochs. Thus, definitional evidence for a critical period, in the form of pre-maturational declines in proficiency, is not found. DeKeyser acknowledges this failure to replicate the pre-maturational AoA effects observed by Johnson and Newport (1989) (the items used in DeKeyser's grammaticality judgment task were a slightly modified subset of those used by Johnson and Newport). DeKeyser considers this discrepancy "hard to interpret" (2000: 513), and goes on to develop an explanation based on putative artifacts of sampling (2000: 514).

- *Aptitude*. DeKeyser administered to all participants a Hungarian-language adaptation of Carroll and Sapon's (Carroll and Sapon 1959) *Modern Language Aptitude Test*. The average aptitude score of all participants was a low 4.7 out of a possible 20. DeKeyser divided the 57 participants into a high aptitude group (n = 15) whose aptitude scores were 6 or higher, and an average- or low-aptitude group consisting of 42 participants. To clarify, the 15- and 42-participant breakouts for high aptitude and average/low aptitude, respectively, were not the same participants as the groups of 15 early arrivals and 42 late arrivals. Across all 57 participants, aptitude was not predictive of GJ scores ($r = 0.13$ ns). The reported correlation of aptitude with GJ scores for early arrivals was not significant either ($r = 0.07$ ns). However, for late arrivals, a significant positive correlation of aptitude and GJ scores was observed ($r = 0.33$, $p < 0.05$). DeKeyser had predicted that adult learners would not

score within the range of early AoA participants unless they had high language learning aptitude. The combination of: a significant positive correlation of aptitude and performance among late arrivals, a non-significant correlation of aptitude and performance for early learners, the performance near ceiling of early learners, and an examination of 5 (of 6) higher-aptitude late learners whose GJ scores were within the range of early learners, leads DeKeyser to the following generalization: "Whereas the younger acquirers in the present study all reached a native or near-native level regardless of aptitude, only the adults with above average aptitude eventually became near native" (2000: 515). "Aptitude plays a role for adult learners" (2000: 515) in the sense that, on L2 proficiency measures, high aptitude trumps, or compensates for, high AoA. Thus, the basting that sews together the AoA variable and proficiency is the interaction of AoA and an additional learner variable, language learning aptitude: aptitude conditions performance among late learners, but not among early learners. This is a notable finding, to the extent that its interpretation allows for rationalization of high GJ scores among late learners. However, what is also notable, and what the DeKeyser study does not adequately investigate in its data, is a clear-cut set of relationships involving the education variable.

- *Years of schooling.* With the data provided in Appendix A of the DeKeyser chapter, I conducted correlations of years of schooling with performance on the grammaticality judgment task. I found that, over all AoA (n = 57), years of schooling significantly correlate with grammatical proficiency ($r = 0.45$, $p < 0.001$). Education also predicts GJ scores among late learners (n = 42; $r = 0.51$, $p < 0.01$) as well as among early arrivals (n = 15; $r = 0.78$, $p < 0.001$).[2] With learners separated into aptitude groups, my analysis reveals that education is again predictive of proficiency. For the 15 high aptitude participants, years of schooling correlate significantly with GJ scores ($r = 0.564$, $p < 0.05$). Likewise, for the 42 low- to average-aptitude participants, education predicts proficiency ($r = 0.43$, $p < 0.01$). Meanwhile, education and aptitude are not correlated over all AoA ($r = 0.03$ ns), nor among early arrivals ($r = 0.006$ ns), nor among late arrivals ($r = 0.08$ ns), suggesting the independent contributions of education and aptitude. To summarize, years of schooling predict GJ results across all relevant correlations. Importantly, unlike AoA and unlike aptitude, the 'education effect' is systematic: significant correlations are not restricted to certain AoA spans or certain aptitude levels.

The DeKeyser (2000) narrative is about finding a connection between AoA and L2 proficiency that is consistent with the CPH/L2A. But by framing the study around the 'robustness of critical period effects', the most robustly predictive factor in proficiency—education—is neglected (see Hakuta et al. 2003 on the role of education in L2 proficiency over AoA).

---

[2] DeKeyser (2000: 515) erroneously reports that the correlation of years of schooling and GJ scores is $r = 0.006$ ns, for early arrivals, and $r = 0.08$ ns, for late arrivals. In fact, these reported coefficients reflect correlations of years of schooling with *aptitude*; see discussion to follow.

Researchers in SLA have an interest in knowing what factors account for L2 proficiency in a sampled population. This interest is not limited to explanations of high-aptitude late learners' proficiency as a function of assumptions of the CPH/ L2A. A more fundamental concern is accounting for L2 proficiency globally, over all AoA and over all aptitudes. Perhaps the coat's palette might include a few neutral tones alongside the many bespoke hues.

6 Conclusion

The CPH coat of many colors, pointedly so named by David Singleton, has a history going back to Penfield and Roberts (1959) and Lenneberg (1967). Over the ensuing years the garment has graced the torso of many a modish scholar. The present contribution has suggested that a gusset here, a gather there, might mean the difference between a well-worn coat and one that is worn well.

References

Abrahamsson, N. and K. Hyltenstam. 2009. Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59: 249–306.

Ayer, A. J. 1959. History of the Logical Positivist movement. In *Logical Positivism*, ed. A. J. Ayer, 3–28. New York: Free Press.

Birdsong, D. 2005. Interpreting age effects in second language acquisition. In *Handbook of bilingualism*, eds. J. Kroll and A. DeGroot, 109–127. Oxford: Oxford University Press.

Birdsong, D. and M. Molis. 2001. On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44: 235–249.

Carroll, J. B. and S. M. Sapon. 1959. *Modern Language Aptitude Test: Manual.* New York: Psychological Corporation.

Cook, V. 2003. *Effects of the second language on the first.* Clevedon, UK: Multilingual Matters.

DeKeyser, R. M. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22: 499–533.

DeKeyser, R., I. Alfi-Shabtay and D. Ravid. 2010. Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31: 413–438.

Fowler, C. A., V. Sramko, D. J. Ostry, S. A. Rowland and P. Hallé. 2008. Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics* 36: 649–663.

Granena, G. and M. H. Long. 2013. Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29: 311–343.

Hakuta, K., E. Bialystok and E. Wiley. 2003. Critical evidence: A test of the Critical-Period Hypothesis for second-language acquisition. *Psychological Science* 14: 31–38.

Hyltenstam, K. and N. Abrahamsson. 2003. Maturational constraints in SLA. *The handbook of second language acquisition*, eds. M. H. Long and C. J. Doughty, 539–588. Malden, MA: Blackwell.

Johnson, J. S. and E. L. Newport. 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.

Lenneberg, E. H. 1967. *Biological foundations of language*. New York: Wiley.

Long, M. H. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12: 251–285.

Ortega, L. 2009. *Understanding second language acquisition*. London: Hodder Education.

Penfield, W. and L. Roberts. 1959. *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

Popper, K. 1959. *The logic of scientific discovery*. New York: Basic Books.

Singleton, D. 2005. The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching* 43: 269–285.

Stevens, G. 2004. Using census data to test the critical-period hypothesis for second-language acquisition. *Psychological Science* 15: 215–216.

Vanhove, J. 2013. The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PLoS ONE*. 8(7): e69172. doi:10.137/journal.pone.0069172