

Evaluating the Existence and Nature of the Critical Period Hypothesis in Second Language

Acquisition

John M. Jayasankar

A thesis submitted in partial fulfillment of the requirements

for the degree of Bachelor of Arts in Linguistics

Haverford College Linguistics Department

December 2021

Acknowledgements

First, I would like to thank my thesis advisor, Jane Chandlee, for her continual support, guidance, and advice throughout this process. Jane's provision of essential feedback was paramount. Jane's kindness, patience, and understanding throughout this endeavor was unparalleled. In the same breath, I would like to thank my second faculty advisor, Shizhe Huang, whose motivating words and tenacious demeanor was an equally prominent factor in motivating this work. Additionally, I would like to thank Brook Lillehaugen, who was instrumental in me first discovering my passion for linguistics and whose fervor for learning through exploration continues to inspire me. I also want to express my sincerest gratitude to Eloise Kadlec and Jiangxue Han, my fellow linguistics thesis peers, whose careful and thoughtful comments and critiques of my thesis have shaped its refinement into its current form. Lastly, I would like to thank all the innumerable people who encouraged and energized me along this journey. Finally, I am forever thankful to the Haverford College Linguistics Department for all of the wonderful learning opportunities and memories. I am grateful for the unwavering support throughout my time here.

Abstract

This paper seeks to investigate the existence and nature of the Critical Period Hypothesis (CPH) in Second Language Acquisition (L2A). I conduct an extensive literature review into many studies spanning five decades into many domains of research. I advocate for multiple critical periods (CPs) for various aspects of language acquisition (morphology, syntax, phonology, phonotactics, grammar, semantics, pragmatics) each with their own unique discontinuity between ultimate attainment (UA) and age of acquisition (AoA). I expose gaps and highlight sources of debate within current literature such as the validity of (UA) as a yardstick for evaluating L2A proficiency, problematic statistical methodology for modeling the discontinuities in the AoA-UA function, language acquisition transfer interference from first language acquisition into L2A, individualistic traits such as language aptitude and motivation. I examine methodological differences in existing literature with a particular focus on incorrect assumptions and statistical techniques that lead to false conclusions being drawn about the shape of the age of acquisition (AoA) and ultimate attainment (UA) function, in testing for the CPH. Ultimately, I advocate for the re-analysis of past studies using different methodological techniques to generate new AoA-UA function graphs to discern if there are real discontinuities or not. I hypothesize that correct and repeatable statistical modeling and proper experimental design will facilitate the discovery of multiple CPs that occur in a robust sequential order with unique onsets, offsets, and discontinuities to each CP. I also hypothesize that individuals with common L1s and interlanguage systems share unique predictable CP onset and offsets that are robust within the group. This paper adds to the existing literature by first presenting an updated in-depth

analysis of the current literature and proceeds to discuss how statistical errors in the existing literature may be contributing to the lack of robust evidence for multiple CPs in L2A.

Table of Contents

1	Introduction.....	6
2	Critical Period Definition and Criteria.....	8
3	Discontinuity.....	11
4	Language Acquisition Transfer Interference.....	12
5	Issues with Ultimate Attainment.....	16
6	Ultimate Attainment Measurement Error and Statistical Bias.....	20
7	Single Critical Period or Multiple Critical Periods.....	25
8	Language Aptitude and Motivation.....	31
9	Future Research.....	33
10	Conclusion.....	34
	References.....	39

1 Introduction

People who learn a second language (L2) in childhood are often difficult to distinguish from native speakers who possess it as a first language (L1). Whereas those who begin in adulthood often struggle to fully attain and master all linguistic aspects of a language. Therefore, L2-learners are often qualitatively and quantitatively different in their mastery of the language when compared to traditional native-speakers who grew up learning the same language as their L1. The later a person begins to learn a language, learning becomes more difficult, time-consuming, and mastery seems to be less assured. Furthermore, there is general agreement upon the fact that the age of onset (AoO) of linguistic exposure for any language is negatively correlated with success of language acquisition. However, there is no consensus on whether this is indicative of the existence of a critical period (CP) for second language acquisition (L2A), or explanatory about the nature of such a CP.

It is so far unclear in existing research whether age is a causal factor, or if there are other causal factors relevant to language acquisition. If age was proven to be a causal factor (characterized by a discontinuity in the relationship between age of acquisition [AoA] and ultimate attainment [UA]) then this would be evidence of a CP. However, in the absence of finding such evidence, there is no empirical support for its existence. Consequently, there is great debate about the existence and nature of the critical period hypothesis (CPH) in second language acquisition (L2A). Historically, though, many on the pro-CPH side of the argument advocate for the explanation that there are maturational constraints such as loss of brain plasticity, cognitive limitations, or loss of access to Chomsky's (1986) Universal Grammar (UG) following the onset of puberty. This event results in a discontinuity in L2A mastery (proficiency) at the critical period (CP) thought to be causally related

to age. This discontinuity is a key concept that is discussed in greater detail throughout this paper as discontinuity between AoA and proficiency is necessary to prove the existence of a CP.

In this paper, I first outline the history of the CPH and define the necessary characteristics of a CP. Second, I detail the burden of proof that must be demonstrated through data to establish its existence, namely the discontinuity between AoA and the outcome of language proficiency (most often measured by UA). Third, I discuss the nature and implications of the discontinuity in the AoA-UA function as supporting evidence for the CPH in L2A. Fourth, I consider whether L1A and L2A are unique processes or if there are spillover effects due to transfer interference between these two processes. Fifth, I assert that the yardstick of UA is problematic and inappropriate for L2A analysis and propose the introduction of a new standard more specifically equipped for usage with L2A. Sixth, I expose how various statistical techniques used in the existing literature in L2A studies are inherently flawed. Seventh, I explore the resulting statistical bias that results from measurement error from flawed methodological techniques and experiment design. Eighth, I weigh literature advocating for a single CP against literature advocating for multiple CPs and ultimately give credence to multiple CPs. Ninth, I consider the possibility of language aptitude, individual motivation, and other personal characteristics as relevant variables in the CPH debate. Tenth, I suggest avenues for future research based on my theoretical findings in light of literature in the debate and statistical shortcomings of papers past.

2 Critical Period Definition and Criteria

Before discussing the role of discontinuities, it is necessary to provide background on the critical period hypothesis (CPH) and establish a fundamental and standard definition of a CP. In this section I also discuss numerous proposed onsets, where a CP begins, and offsets, where a CP ends, that have been presented in the literature. The CPH was originally proposed in neuro-linguistic literature by Penfield and Roberts (1959) who argue for a CP onset at birth and offset at nine years of age when neural plasticity subsided. Later Lenneberg (1967), speculates that maturational constraints in the brain limited lateralization of processes to the left/right hemisphere after the offset of a CP, which he proposes was from 2-14 years of age and also proposed that this extends to L2A (Hakuta et al., 2003). Others find that the CP has an offset before six years of age (Hyltenstam and Abrahamsson, 2003; Ruben 1997; Singleton and Lesniewska, 2021). Alternatively, Meisel (2008) proposes a CP offset between three and four years of age, while Dollmann et al. (2020) find that the offset is nine years of age, and Hartshorne et al., (2018) find that the offset is 17 years. Clearly the literature is inconsistent with establishing a particular CP window that is applicable to all aspects of linguistic development morphology, syntax, phonology, phonotactics, grammar, semantics, pragmatics,). Additionally, puberty, which is when the brain undergoes the most maturational changes according to most neurolinguistic research, occurs between the ages of eight and fourteen years of age, and is different across people and sexes as biological females tend to mature faster than biological males (Singleton and Lesniewska, 2021).

Birdsong (2017) defines a CP as a “bounded maturational span during which experiential factors interact with biological mechanisms to determine neurocognitive and behavioral outcomes.” Linguists agree that robustly, CPs are characterized by four primary aspects. First, CPs have a clearly specified beginning and endpoint for the period. Second, there is a well-defined decline in L2A proficiency and learning rate at the end of a CP (Birdsong, 1999). Third, there is evidence of qualitative differences in learning between LA within and outside a CP. Fourth, there is robustness to environmental variation inside a CP (Hakuta, 2001).

According to Bornstein (1989) and Colombo (1982), CPs imply both a high level of preparedness and natural ability for learning within a specified developmental period to ensure mastery and, a substantial and significant hindrance of ability and inclination for learning outside of this interval (Hakuta, 2001). These conditions necessitate that learning is both qualitatively and quantitatively different inside versus outside a CP. This difference is often attributed to neural plasticity allowing for optimization of specific linguistic processes within windows (Singleton and Lesniewska, 2021). Therefore, when a CP closes at its offset, the belief is that neural plasticity subsides due to neural maturation. In a sense, the linguistic foundation established during a CP is cemented at offset, and more complicated functions build on this now essentially immutable base (Cisneros-Franco et al., 2020). Animals also experience CPs, however unlike CPs in less neurologically complex animals (imprinting in ducks and geese, for example), CPs in humans are less clearly defined in onset and offset across individuals (Purves et al., 2004; Singleton and Lesniewska, 2021). Despite this, learning within a CP is assured to be similar across individuals, normatively described, and governed primarily by endogenous and not exogenous (environmental) factors (Birdsong, 1999).

Learning outside of a CP is qualitatively different in both how learning is accomplished (what part of the brain is involved) as well as quantitatively different (how much a person can learn and at what rate learning occurs). Outside of a CP, learning proficiency is diminished and outcomes are erratic as compared to within a CP (Birdsong, 1999). As a result, theoretically speaking, complete proficient language learning after the offset of a CP is not possible, if the CPH is in fact true. This must be robust across L1A and L2A if there exists a CPH for L2A. Consequently, if the CPH were true, a clear discontinuity between proficiency of acquisition and time of acquisition would be visible at the offset point of a CP (Birdsong, 1999).

The CPH relies on causality as an explanation for the differing success in L2A between younger and older learners, AoA, acquisition age, is the key variable. Therefore, the notion of attributing causality instead of correlation is endemic and problematic to the CPH (Birdsong, 1999), and this is where much pro-CPH research has been criticized. While there is consensus that there is a correlation between the AoA and UA it does not necessarily entail that age is a causal factor in that relation (Birdsong, 1999).

In order to prove that AoA is causally related to UA and that there is a CP, experiments must be able to successfully isolate age as an independent testable variable and disentangle it from all the other extraneous variables to point to age being a causal factor. A causal relationship between AoA and language proficiency (measured here by UA) would affirm the CPH by confirming that there is irreversible unilateral interference by age on linguistic/cognitive processes that govern L2A (Birdsong, 1999). This would necessarily be represented visually as a nonlinear and sudden discontinuity at the end of a CP, and not as a gradual monotonic linear decay on a graph of AoA against L2A proficiency. If age is not a causal factor, though, and

instead is merely a corollary factor, the resulting function will be a linear relationship which refutes the CPH in L2A (Hartshorne et al., 2018; Hakuta, 2001).

3 Discontinuity

Previous research on the role of age in L2A indicates that there is a negative correlation between AoA and UA, this is universally accepted. However, in order to provide incontrovertible evidence for the existence of a CP hinges on the ability to find a discontinuity between AoA and UA. A discontinuity, in studies with large sample sizes, between the effects of age and proficiency of language attainment (as indexed in research by UA) is a necessary feature to determine the existence of a biological constraint. This is a consensus in the field. What is not consensus, however, is the shape of the AoA-UA function and location of the discontinuity. I now turn to discussing different patterns of the AoA-UA function with discontinuities.

There are three primary patterns of the AoA-UA function that meet the condition of different slopes before and after offset of a CP. First, a steep decline of AOA-UA leading to the offset of a CP and afterwards, no age effect (a floor, flat constant slope at minimum of the function) (Birdsong, 2006). Second, a constant level during the window of a CP (a ceiling, flat constant slope) and a steep continual decline after the closure of a CP (Birdsong, 2006). Third, a constant level during the window of a CP (a ceiling, flat constant slope at maximum of the function) followed by a steep decline after the offset of a CP, but that levels off (a floor, flat constant slope at the minimum of the function) at a certain age where a floor is hit, and no age effect is observed after this point (Birdsong, 2006). Most often in the literature, researchers

predict this first pattern for AoA-UA function outcomes, if the CPH is assumed to be true (Vanhove, 2013). Presently, however, most of the studies find no sharp discontinuity in language learning capacity at a specific age, and instead point to a gradual decline (implying age as a corollary factor).

I argue that three shapes of the AoA-UA function generate uncertainties and raise some critical questions due to oversimplification. For example, do all L2-learners who have an AoA within a CP achieve UA in the same amount of time, while L2-learners who do not drop off suddenly following the closure of a CP? Or alternatively, is there a steady decline in efficiency of L2A inversely related to AoA; later AoA within a CP window results in slower UA than earlier AoA within a CP, *ceteris paribus*? Here the efficiency, or speed, to UA is left ambiguous and oversimplified by the second and third primary proposed shapes that posit that UA rate is constant among all who begin during a CP. I question the validity of this assumption as the shape of the AoA-UA function is likely affected by individualistic characteristics such as innate proclivity for language learning, individual motivation to learn, and similarity of L1 to L2. I discuss these issues further in sections 4, 5, and 8. Additionally, I discuss methodological errors about how the AoA-UA function for L2A is estimated and analyzed in section (6). First, I turn to exploring the issue of L1A-L2A transfer interference.

4 Language Acquisition Transfer Interference

In deciphering the nature of the presence of the CPH in L2A, it is essential to distinguish the processes of L1A from L2A. Researchers are divided in terms of considering L2A a unique

independent repetition of L1A (no transfer interference) or instead, a cumulative process that draws on L1A data, patterns, knowledge, and abstractions in order to contextualize L2A (transfer interference present). Transfer interference is inherently a complex phenomenon due to the difficulty in isolating the effects of L1A on L2A in randomized experimental environments. Beyond this, there is a lack of consensus governing if there is varying transfer interference for different aspects of language acquisition (semantics vs. phonetics/phonology, for example). If interference is occurring, L1-learners with closely related L1 to target L2 may achieve a more efficient and speedier L2A whereas those dissimilar L1 from target L2 may find delays and errors in their L2A.

I will now examine the differing views of theorists into the nature of transfer interference. Birdsong (1999) claims that L2A within a CP should show little to no L1A transfer interference because direct access to Chomsky's (1986) UG should override cognitive interference when constructing the system of rules for L2. Birdsong (1999) therefore claims that L1A and L2A are identical processes performed with a 'clean slate' with access to UG and argues for no transfer interference. On the other hand, Hartshorne et al. (2018) argue L2A may reflect interference from L1A if L2A learning rate is non-linear rather than linear with respect to age. In agreement, Galasso (2002) insists that there is no natural 'clean-slate' UG for L2A but instead that L1 interference in L2 is strategically derived by native L1 parameters and not random L1-L2 mismatch. More nuanced even, Bialystok (1997) agrees with the presence of transfer interference but attributes it to being random instead of strategically derived from L1 parameters. Moreover, there is a lack of consensus about the patterns of L1A-L2A interference in the literature.

It is also important to consider interference in relation to age, that is if transfer interference predictably increases or decreases in frequency and magnitude with age. If there is a positive correlation between transfer interference and age, then it would follow those older individuals would have relatively more contamination in their L2A proficiency than younger individuals. It is important to consider that this L2A contamination is not inherently negative as a very closely related L1-L2 match may aid in L2A, despite being contaminated from linguistic elements from L1A. Conversely, L2A contamination may be harmful to learning if there is greater L1-L2 mismatch between one's own native language and the target language. Moreover, isolating the presence of L1A-L2A contamination is difficult due to the still opaque understanding of how language learning is processed and how the brain treats spillover effects (akin to a 'black box' scenario in machine learning). Linguistic data input goes into our neural network and language outcome data is the output. Yet the exact sequence and nature of the algorithms between input and output are not well understood for language learning both in humans and artificial intelligence. This is especially pertinent for neural networks, where input data undergoes complex transformations in numerous layers of processing, and output data may behave in unpredictable ways even given the same stimuli as input. I turn to investigating this phenomenon more in section (7).

If an experimental design were to isolate the effects of L2A contamination, then it would be direct evidence for transfer interference regardless of the nature of L1-L2 transfer interference. Additionally, if L1-L2 interference was definitively proven to increase with age it would support the existence of some abstract UG as access to some innate program would diminish with age (at the offset of a CP) and increase reliance on past L1 knowledge. It would be

worthwhile to examine the role and nature of L1-L2 transfer interference for individual aspects of language. If the amount of L1-L2 transfer is different for phonology versus syntax after the respective CPs per say that would be noteworthy in dictating the reliance on a bioprogram/UG for each individual aspect and the implications of L1-L2 transfer within each of the many CPs. Again, formulating a type of experimental design aimed at isolating the effects of transfer interference (if any) requires careful consideration and attention to detail to remove any potential biases.

In a similar vein, Krashen's Input Hypothesis (Krashen, 1987, 1988) states that for L2A learning is maximized by receiving 'comprehensive' linguistic input from L2A stimulus that is just beyond an individual's current level of grammatical competence. Krashen's (1987, 1988) hypothesis is often criticized for vagueness, oversimplification, and overclaims of linguistic input (White, 1987; Lightbown and Spada; 2006; Gregg, 1984,), and it assumes that the process of L2A is similar in nature to L1A (essentially a repetition). Understanding how L2A is organized in the brain is critically important to understanding the role and presence of L1A transfer interference in L2A. Without knowing for sure that there is no transfer interference, it is far more dangerous to assume there is no contamination of L2A by L1A. Therefore, without evidence of no transfer interference, L2 must be treated as a recapitulation of L1A with bias from L1A. Data into L2A supports this, despite no study being able to demonstrate insurmountable proof. Treating L2A as contaminated is at worst over-precautionary and may weaken the significance of results or the robustness of empirical studies due to the variable of L1 being variable across all individuals. However, it is reckless to continue as many studies have by

assuming the processes are identical and that L2A suffers no spillover effects from L1A. I will now discuss the issues with UA that transfer interference motivates.

5 Issues with Ultimate Attainment

Since the relevant discontinuity of interest is between age of acquisition (AoA) and ultimate attainment (UA) it is necessary to discuss the metric of UA and why UA is inappropriate for usage in L2A, before delving into discussion about the nature of the discontinuity. While UA has been the primary variable of interest in many of these studies the lack of consensus governing the CPH is problematic. Additionally, since identifying a L2-learner who attains UA after the offset of a CP would invalidate the CPH, this may motivate researchers to increase the standards of the UA to an artificially higher level, akin to native-speakers, in order to attribute or lack of success in acquisition to the CPH (Vanhove, 2013). This is especially troubling for bilinguals and multilinguals as their linguistic inventories and production behavior differs from monolinguals (Muñoz and Singleton, 2011; Cook, 1992).

The UA metric can be thought of as being a behavioral cumulative representation of susceptibility to linguistic stimulus, and thus UA is integrative to the susceptibility function (Newport, 1991; Vanhove, 2013). UA is also influenced by other factors such as intensity of linguistic stimulus, duration of linguistic stimulus, among others. Therefore, as Vanhove (2013) points out, all else equal, UA decreases as susceptibility decreases. However, UA decreasing is not alone evidence for the CPH. There needs to be a discontinuity between AoA and UA. However, if all else is not equal, if there is variable duration or intensity of linguistic stimulus,

for example, then the UA function may take on a variety of different forms. There is significant controversy in existing literature surrounding the usage of UA or ‘nativelikeness’ when judging the success of L2A, due to the increasingly high and nearly unattainable level to which it holds L2 speakers. The usage of UA in L2A evaluation is problematic for several reasons. First, the standard of UA is inherently subjective and implies that there is a standard and uniform level of language proficiency that characterizes a native speaker, when in reality it is understood that native speakers may demonstrate a wide variety of linguistic variation in accent, vocabulary, semantics, grammatical structure, tone, and syntax; all of which do not diminish their status as native speakers with full comprehension of a language (Dabrowska et al., 2020; Singleton and Lesniewska, 2021). Thus, it cannot be assumed that all native speakers are homogenous in linguistic inventory and production, this is an oversimplification error. In addition, native speakers constantly make errors when speaking their native L1, this is well documented and robust across many languages and age groups (Dabrowska, 2012).

UA also does not take into consideration the possibility or nature of transference between L1A-L2A. Instead, UA as a metric ignores this as UA treats L2A as an identical replication of L1A. To this point, researchers continue to assert that we must not compare multilinguals to monolinguals as this is not a useful or fair comparison due to the nature of transfer interference and dissimilarities between L1A and L2A processes. As evidence for this, Cook’s multi-competence perspective (2002, 2016) asserts that multilingual speakers cannot have their non-L1 (L2, for example) judged against a native speaker’s L1 of the same language due to transfer interference. To add to this sentiment, Singleton and Lesniewska (2021) propose that interactions between various learned languages for multilinguals is a complex issue that is

not totally understood, but evidence of transfer interference makes comparison moot. This phenomenon, attributed to Larry Selinker (1972) is referred to as an ‘interlanguage system’ that characterizes a dynamic linguistic system in which there exists a neural repository of all language learned and based on L1, and L2, L3, L4, ..., Li learners draw upon the features of their L1 to approximate the target language but usually overgeneralize or create incorrect linguistic phrases. Naturally, Selinker’s (1972) theory assumes that there is an underlying latent neural structure that becomes activated when a learner is attempting to learn another language other than L1. Birdsong (2008) elaborates that it is the nature of bilingualism to produce variable results when compared to monolinguals for a particular language due to a certain mutual interaction between languages (Singleton and Lesniewska, 2021). This belief has been researched more in-depth and supported generally by Abrahamsson (2012), Wei (2018), Singleton (2020), and Jarvis and Pavlenko (2008) as well as Singleton and Lesniewska (2021).

Further, UA is unfair as it implies native-like (L1-like) mastery of all aspects of language, morphology, syntax, phonology, and phonotactics, grammar, semantics, and pragmatics, yet even very young children before the offset of a CPH in L2 differ from native speakers in their level of competencies of the lexico-grammatical dimension (Hyltenstam and Abrahamsson, 2000). It is rare to find L2-learners who can use the target language as well as native speakers of the language (Hoang-Thu, 2009). Also, young children exposed to L2 often end up speaking the L2 with a non-native accent which is deemed as not meeting UA criteria fully. Even learners who started receiving L2 input, are often found to end up without UA in their target language (Christie, 2012).

The consensus for much L2A research into the CPH finds the UA metric to be fundamentally flawed and detrimental to research design studies in evaluating proficiency for L2A. While I agree that it is imperative to have a metric of comparison in testing for the effects of age on L2A, L1-UA is an unachievable standard in L2A. L1-UA and L2-UA are not comparable and should be treated as distinct because the monolingual L1 standard is UA not the L2A standard (Cook, 2002). Therefore, I advocate for a unique and separate metric for evaluating L2A as the effects of transfer interference and the natural differences between monolingual (L1) and multilingual acquisition (L2, L3, L4..., Li) of the same language is qualitatively different even at the same level of proficiency. There is an assumption in all research into L2A that the learner is striving toward some stateable goal, which has thus far been native-like near perfection. However, it is known that linguistic perfection is not possible since performance is flawed by human error in production and subjective judgments (Birdsong, 1999).

Lastly, with the current UA metric, subjectivity is a relevant issue. Different subjective evaluators with differing criteria may evaluate the same level of achievement from the same person differently. This cannot therefore be a trusted benchmark. This is problematic as the standard must be more objective and robust for evaluation. Furthermore, while there is an endpoint and ceiling in L2A that is reached like L1A, and some learners may be able to achieve L2-UA there must be a more realistic empirically testable standard than the current standard of UA.

6 Ultimate Attainment Measurement Error and Statistical Bias

In order to evaluate the existence and nature of the CPH, there needs to be a clearly defined scope of a CP age range along with testable and falsifiable hypotheses. However, there is a lack of consensus on both the former and latter in the literature (Vanhove, 2013). The primary variable of interest has been AoA with respect to acquisition attainment level. The literature so far has not considered rate of acquisition or other variables because individualistic differences such as innate language sensitivity may contribute to rate differences and even adult learners have been shown to exhibit faster acquisition initially than some children which may be attributed to several factors, namely more awareness and directed attention to learning (Vanhove, 2013). Nevertheless, three factors contribute to measurement error and statistical bias.

First, researchers have been using different age ranges in their samples and are not consistent across studies with sampling from a particular age range. Second, researchers have been using the problematic standard of UA as a benchmark for L2A mastery. Third, as a result of using the problematic measure of UA for L2A studies, the effects of age are mismeasured and biased. This measurement error causes bias because researchers are holding L2-learners to the standard of native L1-learners of the same target language which is a primarily an unattainable standard for L2-learners. Therefore, this not only overestimates the effect of age generally as it is classifying those who may in fact have L2A mastery as not reaching UA, it also constrains the window of a CP. To elaborate, if UA is a binary and takes a value of 1 if a learner has achieved mastery and takes a value of 0 if a learner does not achieve mastery, and the CPH is in fact true then there is some age point at which UA will switch from all values of 1 (learners all achieving UA) to 0 (learners all failing to achieve UA) among learners at a specific age. Since UA is not a

fair metric and is largely unattainable for L2-learners, this means that in the data, UA may begin returning values of 0, indicating failure to achieve UA mastery, at an earlier age than if we were to use a different metric such as a more appropriate catered measure of L2-mastery. Such a metric may relax assumptions that L1-learning and L2-learning are synonymous and that outcomes should be similar between L1A and L2A. Furthermore, using a more appropriate method, I hypothesize the window of time between onset and offset of a CP would extend as the current UA metric biases the AoA-UA function and overestimates the effect of age at younger ages than may actually be the case.

Three primary methodologies in research have been used to find evidence of a discontinuous function supporting the ‘flat slope’ prediction, however all three are statistically flawed. The first uses statistical tools such as t-tests, chi-squared tests, and analysis of variance (ANOVAS) to compare group means. The second uses comparison of correlation coefficients between AoA-UA for various groups. The third uses regression analysis with UA as the outcome variable and AoA as the regressor variable. As Vanhove (2013) points out, the first two categories build their argument on statistical fallacies and are invalid. In contrast, regression models are valid so long as they are well-designed with appropriate controls and modeling.

The first category is problematic because researchers, such as Johnson and Newport (1989) and Abrahamsson and Hyltenstam (2009) split up the continuous variable of AoA into discrete interval age bins which are often arbitrary and inconsistent across studies. This dichotomization reduces the statistical power and data available from the variable when reducing it to group means and variance (Vanhove, 2013). It is made worse by the fact that age bin intervals are not consistent across studies as there is no agreed-upon standard, so results are

incorrectly measured in these studies and inherently variable across studies due to age bin cutoff subjectivity. Secondly, when using these methods there is a fallacious assumption that a non-significant result indicates that group means are identical (that UA performance is identical between groups) and that significant test results indicate group means are non-identical (that UA performance is not similar between groups and is different enough). The issue in this assumption is that when adopting the null hypothesis that there is no difference between groups, the statistical power of the results should be significantly higher than the standard in social sciences (Sedlmeier and Girgerenzer, 1989), however in practice, these power computations are generous and differences between groups are going undetected (Abrahamsson and Hyltenstam, 2000; Johnson and Newport, 1989; Vanhove 2013). Consequently, it is inappropriate to use these tests for evaluating whether there is no difference between groups.

The second category treats AoA as a continuous variable but is still problematic because computing correlation coefficients between AoA and UA rests on the assumption that differences in correlation coefficients between groups is representative of different slope parameters between groups (Vanhove, 2013). Slope and correlation coefficients are not synonymous and are often confounded in the literature. Johnson and Newport (1989) made this mistake as many other studies have. The difference between slope and correlation coefficients is as follows: slope is the amount by which the y-axis value changes when increasing the x-axis value by one unit, it can theoretically take any value from negative and positive infinity. In contrast, correlation coefficients indicate the strength of the linear relationship between two variables and takes on a continuous value from -1 to 1. Further, 0 indicates (weakest/no relationship) and the interval endpoints -1 (perfect negative linear relationship) and 1 (perfect positive linear relationship)

increase in strength as moving away from zero in both directions. The relationship between slope and correlation coefficient in a simple linear function is that slope is equal to the correlation coefficient multiplied by the ratio of the sample standard deviations of the y-and-x-variables (Vanhove, 2013). In order to compare the slope of the AoA-UA function, researchers must focus on slope coefficients and not correlation coefficients.

The third category is the only suitable methodology in using regression models with AoA as a regressor and UA as the outcome variable is the most pertinent approach to obtaining proper and relevant estimates to the AoA-UA function shape. However non-linearities naturally follow from specific regression models, and thus cannot be used to attribute any discontinuities to differences in AoA-UA function (Vanhove, 2013). An important consideration when discussing regression models is the principle of parsimony. The principle of parsimony states that models should be as simple and basic as possible, preferring fewer parameters, linear models, and less assumptions. However, the difficulty is linear models are not always the best unbiased estimators of the AoA-UA function.

Here I will discuss several techniques past researchers have used to attempt to correct for the shortcomings of the traditional linear model. For example, Stevens (1999) uses a logarithmic transformation of the AoA variable in a logistic regression model, which results in inherent non-linearities from this model (Vanhove, 2013). Alternatively, Bialystok and Hakuta (1999) use locally weighted scatterplot smoothing (LOWESS), which guarantees no sudden jumps in fitting the dependent variable to the independent variable (Vanhove, 2013) to estimate the AoA-UA function, which is an appropriate strategy for analyzing the AoA-UA function. Another promising technique is using a cubic regression model which allows for more precise

approximation of the function and more smoothness than a linear interpolation. This allows for the shape of the function to vary more realistically. However, the cubic model has been criticized because pinpointing a specific point of discontinuity is impossible. This is because the slope changes continuously (first derivative test) at a rate that changes continuously (second derivative test) (Vanhove, 2013). Therefore, models such as Flege et al. (1999) that combine both the benefits of the cubic in more precisely modeling the shape of the function, with the benefits of the linear specification, in being able to locate a discontinuity at a point, are sound models. Comparatively, Birdsong and Molis (2001) use multiple piecewise linear regressions to fit to their data to check whether any discontinuities were significantly explained well enough to include to offset the loss of parsimony (Vanhove, 2013). Birdsong and Molis (2001) find that there is a significant discontinuity at 18 years in age (Birdsong and Molis, 2001; Vanhove, 2013) when checking their model against a simple linear model.

Moreover, the approach by Birdsong and Molis (2001) is most convincing to me as it combines the usage of multiple different piecewise regression models, that preserve the principle of parsimony, in conjunction with a test to analyze whether any breakpoint found contributed enough to be included to explain the data over a single standard linear model. Essentially these tests whether the discontinuity found is significant enough to be included in the model while recognizing the tradeoff with parsimony. This weighted technique of comparison is appealing as it is the most statistically sound and is informative about the location of the discontinuity and the shape of the AoA-UA function. I advocate for the usage of this empirical strategy in future research as well as for re-analysis on existing studies. If reconstruction of existing studies' methodology to match this framework is not possible then I instead advocate for alternative

techniques as in Vanhove's (2013) reanalysis of DeKeyser et al. (2010). In the reanalysis, Vanhove (2013) uses the original data and corrects for methodological errors by parsing apart the statistical fallacies and incorrect assumptions in the original paper descriptively and alters methodology where possible.

7 Single Critical Period or Multiple Critical Periods

Within the pro-CPH side of the CPH debate, researchers are divided based on disagreements about the specific length (onset and offset of CP), the existence of one CP versus multiple CPs, and the underlying nature of what is responsible for causing a CP to be offset at a specific age. Much of the literature insists that there is one singular overarching CP for all L2A, others though opt for the usage of 'sensitive' or 'optimal' periods (Columbo, 1982), and some suggest that only certain domains of language are subject to CP restrictions such as syntax, grammar, and phonology whereas lexical semantics (acquisition of new words and meanings) is unbounded by maturational CP restraints (Steinhauer, 2014).

Anti-CPH researchers, on the other hand, claim there is no evidence for an abrupt discontinuity in L2A after the reported offset but instead a gradual consistent decline continuing past reported the offset (Abello-Contesse, 2008; Bialystok, 1997). They do not refute age of onset being correlated to outcome of proficiency in acquisition, but instead claim that it cannot age cannot be proven as a causal mechanism. To this end, anti-CPH researchers criticize the nature of the work of pro-CPH researchers for methodological errors in conducting experimental data and interpreting results. These methodological errors in estimating the shape of the AoA-UA curve and the effects of age lead to inconsistent and incorrect results that in

conjunction with assuming one CP instead of multiple CPs leads to the ‘needle in a haystack’ type problem. The age range being researched is far too large and acquisition is not segmented by language aspect to yield multiple CPs which are more manageable to test under controlled experimental conditions.

Many critics of the CPH attribute the inconclusiveness in establishing a relationship between AoA and UA to poor methodological instrumentation of raw numbers in judgment tests with limited variables and lab-based experimental settings that eliminate the element of naturalness in speech language (DeKeyser et al., 2010). Additionally, Hulstijn (2014) asserts that the prevalence of quantitative research in approach to L2A is an inherently wrong tactic because it is too positivist, and we need an interpretivist approach instead. Therefore, while future research needs to be more observationally reliable, more empirically testable and more methodologically sound, I assert that researchers are simply getting it wrong by overgeneralizing all of language acquisition to a single CP instead of considering multiple CPs. To this end, anti-CPH researchers who advocate for no CP in L2A on account of no definitive onset and offset can be contested by conceptualizing that it is simply a problem of characterizing too many differing onset and offset points for various linguistic functions and maturational timelines (multiple CPs) to one overarching onset and offset (a single CP). This would explain the lack of consensus for a consistent singular CP onset and offset for all language acquisition domains that is robust among all individuals cross-linguistically.

In considering the existing literature among advocates for multiple CPs, Granena and Long (2013), posit that there exist three ‘sensitive’ periods. They hypothesize that the CP for phonology has the earliest offset, followed by lexis, and finally, syntax. Two recent studies seem

to confirm the multiple CP theory: Hartshorne et al. (2018) as well as Dollmann et al. (2020). Hartshorne et al. (2018) advocate that the CP offset occurred at the age of 17 based on a test on syntax (in L2 English), whereas Dollmann et al. (2020) argue that the CP offset was much earlier, at the age of 9, by utilizing the metric of degree of foreign accent (in L2 German bilinguals) (Singleton and Lesniewska, 2021). Additionally, earlier studies have also yielded solid evidence for multiple CPs: for example, Long's (1990) review of Johnson and Newport (1989) determine different ages of decline in plasticity for syntax versus phonology and among other aspects of language. This contrast in plasticity between formal phonological and grammatical versus semantic aspects is consistent with evidence suggesting separately developing neural-linguistic acquisition systems and multiple CPs with differing respective onset and offset points.

Adding to these findings Newman et al. (2001) find differences between phonological and grammatical aspects of language acquisition as compared to the acquisition of semantic and lexical aspects, the former appears to show a negative correlation in acquisition proficiency with increasing age, whereas the latter does not display evidence of a negative correlation with age (a necessary feature of a CP is negative correlation with age and a discontinuity just before and after offset). This seems to suggest that there exists multiple CPs for some aspects of language acquisition while perhaps sensitive periods, or no defined CP at all for others. Qingxin (2012) adds to this by indicating that the differing rate and timing in loss of plasticity with age is the factor dictating onset and offset in multiple CPs. Additionally, Ruben (1997) finds evidence for multiple CPs based on different linguistic domains, notably remarking the onset for phonetics/phonology 6 months of age, the offset for phonetics/phonology is age 1 year, the offset

for syntax is age 4 years, and the offset for semantics is age 15/16 years. In a more recent paper building on their 2013 paper, Long and Granena (2018) find evidence of multiple CPs reaffirming and further specifying Ruben's (1997) order of CPs from earliest onset/offset combination to latest onset/offset combination: phonology, lexis, collocations, morphology, syntax, in that order. Moreover, there has been data spanning decades affirming multiple CPs in a predictable order with remarkably similar onset and offset periods.

In the pro-CPH literature many other proponents of multiple CPs have proposed different CPs with different onset/offset points and with various specifications. While the inability to agree on onset/offsets for various aspects of language paired is disheartening to the multiple CP theory. For example, even studies which claim to support the multiple CPH find that there are always exceptions to the general trend of inability to reach native-like proficiency after a specific offset point, that some late L2-learners can achieve such level of proficiency (Singleton and Lesniewska, 2021). The counterfactual to this is if the multiple CPH theory is in fact true and various CPs occurs in a predictable sequence, then it would be impossible to encounter L2-learners who have a native-like syntax but an imperfect command of phonetics/phonology, if the CP for phonetics/phonology precedes the CP for syntax (Singleton and Lesniewska, 2021). This sentiment is echoed by Krashen's (1987, 1988) Natural Order Hypothesis that states that specific grammatical structures in language are always acquired earlier while others are always required later, without expect to environmental variation. Hypothetically, from this past example, this is not evidence against the multiple CPH in general but more specifically against a specific and predictable sequential series of multiple CPs occurring. However, the goal is to be able to meaningfully model and predict the sequential series of multiple CPs. Otherwise, if CPs are

variable both in chronological order and duration (duration between onset and offset), then I would argue that this would hardly be a helpful hypothesis at all.

Studies on immigrants who migrate to another country have provided important naturalistic evidence into understanding the CPH. For example, Asher and Garcia (1969) find that the best predictor of L2A English proficiency is age of arrival in the United States as opposed to length of residence in the United States. This gives credence to the CPH more broadly in that the age of arrival (age of onset of English L2A) is a more relevant predictor of proficiency outcome than duration spent learning (duration of linguistic stimulus). The implication that arises is thus that after a certain age, duration of learning has negligible effects on acquisition mastery. Similarly, Patkowski (1980) finds that age of arrival is negatively correlated with English syntax proficiency for L2-learners who migrate to the United States, such that those who arrive later have worse outcomes than those who arrive earlier. Likewise, Piske et al., (2002) find that age of arrival is negatively correlated with native-likeness of vowel production for Italian immigrants migrating to the United States to learn English as L2. These results taken together all seem to indicate that the earlier an individual migrates to the place where the target non-native language is spoken, the more likely an individual is to achieve native-like proficiency in that language, everything else held constant. It is important to note that several other relevant factors determine proficiency of acquisition in L2-learning among migrants, such as L1-L2 similarities, language aptitude, and motivation to learn, however these natural results provide some critical insight into the CPH debate.

Computational linguistics and machine learning models also support the theory of multiple CPs. Similar to the proposed mechanism in humans, deep artificial neural networks

exhibit loss of information following a rapid increase in information in the early phases of training. This is often referred to as loss of ‘information plasticity’. A study by Achille et al. (2018) finds that the first few epochs are critical for the creation of strong connections, and these do not seem to change with additional training. In fact, deep neural networks exhibit CPs in which a temporary stimulus deficit can impair the development of a skill, and this impairment depends on the duration of the stimulus deficit period and size of the neural network (Achille et al., 2018). Similarly, Parisi et al. (2019) challenge the notion of ‘lifelong learning’ by analyzing neural networks in both biological agents and artificial intelligence models. They find that through a careful review of machine learning models with continuous streams of information and dynamic structures allowing for strategic allocation of network layers to accommodate new knowledge, that there exists CPs and asymptotic returns to learning. In summary, even the most complex and novel algorithms have limitations to learning windows and a maximum volume of information capable of being processed. This is important as these results indicate that CPs are robust not only to biological beings but also to artificial intelligence due to the structure of information processing and the nature of learning dynamics.

Moreover, I advocate for the existence of multiple CPs in L2A. I employ a Fermi-esque paradox on L2A, if there was only a singular CP, existing research likely would have already identified a definitive onset and offset points due to the allocation of research and resources to this debate, regardless of the large age range studied in the literature. This is not the case, though, as the idea of only one CP is macro-constrained and overgeneralizes the process of acquisition to a broad interval that is not well-defined. If we analyze each linguistic aspect as a microelement of a greater L2A process we can begin to understand the existence of many age-related

constraints based on individual aspects of language, many CPs within this one umbrella CP and hence there will never be any single onset or offset for all language aspects, as it is too broad. Just as developmental processes do not all occur simultaneously or at a single universal onset and offset, I argue L2A is no different. Evidence from iterative learning processes both in humans and machine learning models support this claim. Kirby et al. (2014) find that language learning is robustly iterative and results in systematically structured behavior in which linguistic elements necessarily build on one another in a predictable format, giving credence to Krashen's (1987, 1988) Natural Order Hypothesis. Combining these iterative learning findings with research indicating multiple CPs for deep artificial neural networks points to there being multiple CPs for human language learning. Furthermore, I maintain that there is an onset and offset for multiple CPs varying on type of linguistic aspect; syntax, phonetics/phonology, semantic, and so on. The reason we have yet to be able to reach consensus on onset and offsets of CPs is due to the fact that much of the existing research has been using flawed methodological approaches and searching for a singular CP, so the AoA-UA function is being incorrectly approximated, and this is biasing results.

8 Language Aptitude and Motivation

A common problem researchers face when investigating the CPH is the inability to isolate or account for innate differences between individuals with respect to language learning. The most complex and vexing of these individual characteristics is language aptitude and motivation. However Krashen's Affective Filter Hypothesis (1987, 1988) also states that there are a number of other affective variables such as self-confidence, anxiety, and other personality

traits in addition to motivation and language aptitude that are responsible for playing a facilitative, but not a causal role in L2A. Krashen (1987, 1988) hypothesized that individuals with high motivation, high extroversion, high self-confidence, high self-image, and low levels of anxiety are better suited for more successful L2A, while individuals with low motivation, introversion, low self-confidence, low self-image, and high levels of anxiety fare much worse with successful learning. Krasher (1987, 1988) claimed these negative traits cause the affective filter to activate which inhibits the success of L2A. Contrary to this, the CPH claims that all individuals have an innate ability to overcome language-learning barriers as evidenced by receiving differing amounts of linguistic stimulus in L1A results in mastery of L1 robustly. However, I argue that this is certainly not robust for L2A. Inborn language aptitude is also the argument evidenced for individuals who can achieve mastery outside of a CP. Some theorists even purport those greater aptitudes for language learning may act to combat the effects of a CP closing and language learning may take place outside of the CP, the two forces seem to counter one another. According to the scope of the CPH, mastery outside of the CP window should be theoretically impossible.

A necessary criterion of a CP is that learning outside of a CP must both be qualitatively and quantitatively different from learning within a CP (as evidenced by a discontinuity). Therefore, it may in fact be that individuals with high levels of aptitude are able to learn differently or are able to access different areas of the brain outside of a CP than individuals with lower levels of aptitude. This learning is likely qualitatively different from learning within a CP, if this is true. If this is the case, then this is not problematic for the CPH. However, the conflict with the CPH is in the quantitative differences in learning within and outside a CP. If those with

high levels of aptitude can overcome the offset of the CP without experiencing a discontinuity in acquisition success with respect to age, then this would invalidate the CPH entirely. However, otherwise the CPH would be unaffected by the level of aptitude for language learning as described above. The interaction between language aptitude and a CP is not well-understood and requires more research to isolate how much of an effect an individual's language learning aptitude has on acquisition both within and outside a CP for L2A. Other factors such as motivation have been relevant to success of acquisition. Motivation is relevant as well, but studies have not robustly been able to isolate the effects of motivation as measurement is extremely difficult. Similarly, controlling for language aptitude which is an inherent trait is not able to be done in experiments both because measuring inherent aptitude is not easy as it is not tangible or transparent, and controlling for it is thus not feasible in a meaningful way.

9 Future Research

Moving forward there is an opportunity both to re-analyze past studies and interpret as well as correct for incorrect methodology in addition to conducting novel experiments using appropriate methodology, representative samples, and alternatives to the UA metric. For example, Vanhove's (2013) re-analysis of DeKeyser et al. (2010). Vanhove (2013) uses non-parametric scatterplot smoothers and piecewise linear regressions in conjunction with a myriad of robustness checks in-line with testing the original hypotheses by DeKeyser et al. (2010). To this end, better methodology is applied retroactively to a past study, while preserving the dataset, however the standard of UA and sample used in the data cannot be changed now retroactively. Therefore, it is imperative to establish common foundational standards in research

as to avoid this problem in the future of misaligned standards and variables. While more attention should be given to reanalyzing past studies with more apt methodology to re-estimate the shape of the AoA-UA curve and discontinuities, I advocate that more focus be directed to designing future studies. Future studies allow for a new opportunity to capture all the omitted variables and shortcomings of previous research in beginning new with unbiased data, sampling techniques, metrics of evaluation, and statistical modeling methodology. Moreover, a goal of future research should be to unify research to an agreed-upon set of standards and definitions to ensure all research starts with the same parameters. Collaboration and open communication between researchers in identifying an optimal strategy based on fact is a key aspect to identifying the real relationship between AoA and UA. In addition, easy replication of methodology is also another critical component of research that has been lacking in existing research into the CPH. This begins with thorough documentation. One potential avenue left relatively unexplored for future research is integrating machine learning models with more and more ‘human-like’ neural networks and experimenting with different linguistic input durations, learning environments, and varying characteristics such as aptitude, motivation, and simulated IQ and observing results.

10 Conclusion

The CPH is a hotly debated topic in L2A. Despite the vast amount of literature investigating this topic, there is still very little consensus on findings. We do not have many concrete answers after nearly six decades of research. If anything, the volume of research and variety of the purported results fuels the debate more and more. The more we learn the less we know about the reality of language learning. In many ways the ability to learn language still

represents a ‘black box’ model as there is still a vast amount the literature seeks to understand. It may be the case that there is no robust set of CPs across individuals, and we are in fact attempting to quantify and constrain a process that is unconstrained. Yet it is much more convincing to me that the current findings in the research are to be expected.

If we take Cook’s (2016) multi-competence theory in addition to Selinker’s (1972) interlanguage theory, it makes sense that there are different CPs for L2A based on the proximal similarity between a learner’s L1-to-target-L2. Since both theories assert that a learner’s L1 exists in constant exchange with L2 and all subsequent languages, it naturally follows that the similarity between L1 and target L2 influences the rate of fossilization as well as the ceiling of proficiency for every individual learner. Thus, CPs for L2A are not taken to be robust across all individuals, due to the nature of language learning processes in the brain, but rather robust across L2 learners with the same L1 or perhaps the same interlanguage set of languages, to be more specific. This is because in L2A, L1-L2 patterns are in constant ebb and flow and thus these take priority over other factors at play governing a CP. This notion rationalizes the patterns in the data: differing CP points for identical experiments performed on samples with various L1s. The discrepancies in CP offset timing reflect L1-L2 similarity/dissimilarity. Therefore, it is the result of a broad and overgeneralized UA standard in addition to these multi-competence and interlanguage theories that explains the resulting inability to find a consistent robust offset for all L2-learners who share a common native-language or interlanguage system. It is in fact evidence for the CPH that there are different CP offsets based on unique L1-L2 pairings and interlanguage systems. Moreover, more research needs to be done into segmenting CPH experiments in L2 based on the target L2’s similarity to the learner’s mastered L1. The linguistic proximal

difference (similarity) between these two will directly correlate to the timing of the offset being sooner (later). The more different a target language from the native language, the shorter the offset for the target language, and the more similar, the longer the offset. More empirical research needs to be conducted isolating similar native-speakers and those with similar interlanguage systems when analyzing the CPH in L2A. This requires proper methodology.

Implementing proper methodology is a critical step to discovering the truth about the nature of the CPH in L2A. Much of the research on both the pro-CPH and anti-CPH sides of the debate have centered around the ability (or inability) to discover a discontinuity in the AoA-UA function. In evaluating the discontinuity problem there is no more pressing issue than questioning the reliability of data gathered and methodological processes in estimating results. Even among researchers who agree upon a standard CPH definition, scope, and have explicit predictions about outcomes with testable hypotheses, there still exists debate about what conclusions can be drawn from empirical data that is collected as a result of suspect methodology.

Much of the existing literature that makes claims on either side of the debate relies on faulty assumptions and weak and irrelevant methodological models and designs that do not allow for such conclusions to be drawn from the empirical evidence. Namely, models relying on group mean comparison and/or correlation coefficient comparisons are unable to draw conclusions about the existence of discontinuities in the AoA-UA function due to statistical fallacies associated with drawing conclusions about slope from those such techniques. Alternatively, techniques such as fitting piecewise and linear regressions, locally weighted scatterplot smoothing, and even cubic functions have proven valid for analysis of the AoA-UA function.

Proper methodology and careful analysis ensure that results can be compared across studies. To this end, I recommend moving forward, only large sample sizes of greater than 100 minimum per group be drawn, in addition to standardization of estimation (à la Vanhove, 2013) to be able to compare results from various studies to seek robustness in determining the true nature of the CPH in L2A. Re-analysis of past studies is a critical aspect to seeking robustness and validation in results. While there is much work still to be done in reconsidering the construction of estimation techniques in past research, there is optimism ahead as much of the original data is preserved and original hypotheses, samples, methodology, and results documented, this means that re-analysis is possible on a meta-scale. With eyes to the future, we can learn from the methodological mistakes and shortcomings of the past and aim to truly uncover the nature of the CPH in L2A.

I advocate for multiple CPs for morphology, syntax, phonology, phonotactics, grammar, semantics, pragmatics, each CP having a unique onset and offset period that is robust across all individuals who share a native language and all those with similar interlanguage systems. Each different group with these shared characteristics will have a set of unique, predictable, and shared CPs in language acquisition. For example, all native Spanish-speakers learning French will share the same CPs but will differ from all native English-speakers learning French, who will instead have their own unique and predictable set of CPs.

The lack of empirical evidence in the literature supporting this claim is largely due to methodological flaws, namely researchers using inappropriate modeling techniques. In conducting re-analysis armed with the hypotheses of multiple CPs and segmenting piecewise function regressions properly to detect multiple CPs, I predict there will be evidence of multiple

CPs in the research that were previously undetected. As it stands though the CPH in L2A is unproven empirically and instead relies entirely on theoretical mechanisms. Ultimately, the goal is for researchers on all sides of the debate to work collectively to set standards for methodology, scope, and process when conducting research, and to be united in the eternal quest for answers. In doing so, I am confident that answers about the nature and existence of the CPH in L2A will be illuminated. It is of great importance to seek answers to this question as there are critical implications to formal language instruction and policymakers to ensure that if there are nuanced intervals of time for maximized learning and after which learning is diminished or impossible, that programs adapt around such intervals for the sake of the betterment of learning.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The Robustness of Aptitude Effects in Near-Native Second Language Acquisition. *Studies in Second Language Acquisition*, 30: 481–509.

Abrahamsson, N., & Hyltenstam, K. (2009). Age of Onset and Nativelikeness in A Second Language: Listener Perception versus Linguistic Scrutiny. *Language Learning*, 59: 249–306.

Abrahamsson, N. (2012). Age of Onset and Nativelike L2 Ultimate Attainment of Morphosyntactic and Phonetic Intuition. *Studies in Second Language Acquisition*, 34: 187–214.

Achille, A., Rovere, M., & Soatto, S. (2018). Critical Learning Periods in Deep Neural Networks. *ICLR 2019*.

Asher, J. J., & Garcia, R. (1969). The Optimal Age to Learn a Foreign Language. *The Modern Language Journal*, 53(5).

Awad Al Dawoody Abdulaal, M. (2021). Krashen’s Monitor Model Revisited with Some Linguistic Evidence for the Homogeneity Hypothesis. *TESOL International Journal*, 16(2), 177-197.

Bialystok, E., & Hakuta, K. (1999). Confounded Age: Linguistic and Cognitive Factors in Age Differences for Second Language Acquisition. *Birdsong* [76], 161–181.

Bialystok, E. (1997). Effects of Bilingualism and Biliteracy on Children's Emerging Concepts of Print. *Developmental Psychology*, 33(3), 429.

Bialystok, E. (2002). On the Reliability of Robustness: A Reply to DeKeyser. *Studies in Second Language Acquisition*, 24: 481–488.

Birdsong, D. (1999). Second Language Acquisition and the Critical Period Hypothesis. New Mahwah: Erlbaum.

Birdsong, D. (2006). Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning*, 56: 9–49.

Birdsong, D. (2008). Second Language Acquisition and Ultimate Attainment. *The Handbook of Applied Linguistics*. New York: Wiley, 82–105.

Birdsong, D. (2017). Critical Periods. *Oxford Bibliographies in Linguistics*. New York: Oxford University Press.

Birdsong, D. (2018). Plasticity, Variability and Age in Second Language Acquisition and Bilingualism. *Frontiers in Psychology*, 9: 81.

Birdsong D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-Language Acquisition. *Journal of Memory and Language*, 44: 235–249.

Bornstein, M. H. (1989). Sensitive Periods in Development: Structural Characteristics and Causal Interpretations. *Psychological Bulletin*, 105, 179–197.

Chomsky, N. (1986). Knowledge of Language: Its Nature, Origin, and Use. *Cultural Critique*, 3, 10.

Christie, F. (2012). Language Education Throughout the School Years: A Functional Perspective. *Language Learning*, 62, 1–247.

Cisneros-Franco, J. M., Voss, P., Thomas, M.E., & de Villers-Sidani, E. (2020). Critical Periods of Brain Development. *Handbook of Clinical Neurology*, 173, 75–88.

Colombo, J. (1982). The Critical Period Concept: Research, Methodological, and Theoretical Issues. *Psychological Bulletin*, 91, 260–275.

Cook, V. (2002). *Portraits of the L2 User*. Clevedon: Multilingual Matters, 1–28.

Cook, V. (2016). Premises of Multicompetence. *Cambridge Handbook of Linguistic Multi-Competence*. Cambridge: Cambridge University Press, 1–25.

Cook V. J. (1992). Evidence for Multicompetence. *Language Learning*, 42: 557–591.

Dabrowska, E., Becker, L., & Miorelli, L. (2020). Is Adult Second Language Acquisition Defective? *Frontiers in Psychology*, 11: 1839.

Dabrowska, E. (2012). Different Speakers, Different Grammars: Individual Differences in Native Language Attainment. *Linguistic Approaches to Bilingualism*, 2: 219–53.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-Linguistic Evidence for the Nature of Age Effects in Second Language Acquisition. *Applied Psycholinguistics*, 31: 413–438.

Dollmann, J., Kogan, I., & Weißmann, M. (2020). Speaking Accent-Free in L2 Beyond the Critical Period: The Compensatory Role of Individual Abilities and Opportunity Structures. *Applied Linguistics*, 41: 787–809.

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

Galasso, J. (2002). Interference in Second Language Acquisition: A Review of the Fundamental Difference Hypothesis: Transferring the ‘Pro-drop’ Parameter from Spanish to English.

Granena, G., & Long, M. H. (2013). Age of Onset, Length of Residence, Language Aptitude, and Ultimate L2 Attainment in Three Linguistic Domains. *Second Language Research*, 29(3), 311–343.

Gregg, K. (1984). Krashen’s Monitor and Occam’s Razor. *Applied Linguistics*, 5: 75-100.

Grosjean, F. (1989). Neurolinguists, Beware! The Bilingual is Not Two Monolinguals in One Person. *Brain and Language*, 36: 3–15.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical Evidence: A Test of the Critical-Period Hypothesis for Second-Language Acquisition. *Psychological Science*.

Hakuta, K. (2001). A Critical Period for Second Language Acquisition? *Critical Thinking About Critical Periods*. Paul H Brookes Publishing Company, 193–205.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A Critical Period for Second Language Acquisition: Evidence from 2/3 Million English Speakers. *Cognition*, 177: 263–77.

Hoang-Thu, T. (2009). The Critical Period Hypothesis in Second Language Acquisition. *Alliant International University*, 1–17.

Hyltenstam, K., & Abrahamsson, N. (2000). Who Can Become Native-Like in A Second Language? All, Some, or None? On the Maturational Constraints Controversy in Second Language acquisition. *Studia Linguistica*, 54(2), 150–166.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational Constraints in SLA. *The Handbook of Second Language Acquisition*. Malden: Blackwell, 539–88.

Jarvis, S., & Pavlenko, A. (2008). Crosslinguistic Influence in Language and Cognition. *Sociolinguistic Studies*. London: Routledge, 6(3).

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated Learning and the Evolution of Language.

Current Opinion in Neurobiology, 28: 108-114.

Krashen, S. D. (1987). Principles and Practice in Second Language Acquisition. Prentice-Hall International.

Krashen, S. D. (1988). Second Language Acquisition and Second Language Learning. Prentice-Hall International.

Lenneberg, E. H. (1967). *The Biological Foundations of Language*. New York: Wiley.

Lightbown, P. M., & N. Spada. (2006). How Languages are Learned (3rd ed.). China: Oxford University Press.

Long, M. H. (1990). Maturational Constraints on Language Development. *Studies in Second Language Acquisition*, 12, 251-285.

Meisel, J. M. (2008). Child Second Language Acquisition or Successive First Language Acquisition? *Current Trends in Child Second Language Acquisition. A Generative Perspective*. Amsterdam: John Benjamins, 55–80.

Muñoz C., & Singleton, D. (2011). A Critical Review of Age-Related Research on L2 Ultimate Attainment. *Language Teaching*, 44: 1–35.

Newman, A. J., Bavelier, D., Corina, D., Jezzard, P., & Neville, H. J. (2001). A Critical Period for Right Hemisphere Recruitment in American Sign Language Processing. *Nature Neuroscience*, 5(1), 76–80.

Newport E. L. (1991). Contrasting Conceptions of the Critical Period for Language. *The Epigenesis of Mind: Essays on Biology and Cognition*, Hillsdale, NJ: Lawrence Erlbaum, 111–130.

Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., & Wermter, S. (2019). Continual Lifelong Learning with Neural Networks: A Review. *Neural Networks*, 113, 54–71.

Patkowski, M. S. (1980). The Sensitive Period for the Acquisition of Syntax in a Second Language. *Language Learning*, 30: 449–68.

Penfield, W., & Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton: Princeton University Press.

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2002). The Production of English Vowels by Fluent Early and Late Italian-English Bilinguals. *Phonetica*, 59: 49–71.

Purves, D., Augustine, G., Fitzpatrick, D., Hall, W. C., LaMantia, A., McNamara, J. O., & Williams, S. M. (2004). *Neuroscience* (3rd ed.). Sunderland: Sinauer.

Qingxin, Z. (2012). Is There A Critical Period For Second Language Acquisition? *2012 International Conference on Education Technology and Computer*, 43, 1-5

Ruben, R. J.. 1997. A Time Frame of Critical/Sensitive Periods of Language Development.

Acta Otolaryngologica, 117: 202–5.

Sedlmeier P., & Gigerenzer, G. (1989). Do Studies of Statistical Power Have an Effect on the Power of Studies? *Psychological Bulletin*, 105: 309–316.

Selinker, L. (1972). Interlanguage. *IRAL - International Review of Applied Linguistics in Language Teaching*, 10(1–4).

Singleton, D., & Leśniewska, J. (2021). The Critical Period Hypothesis for L2 Acquisition: An Unfalsifiable Embarrassment? *Languages*, 6: 149.

Singleton, D. (2020). The Translanguaging Conundrum. *Bilingualism: Hungarian and Non-Hungarian Contexts*. Veszprém: Pannon Egyetem, 175–84.

Steinhauer, K., White, E. J., Titone, D., & Genesee, F. (2015). Phonological Processing in Late Second Language Learners: The Effects of Proficiency and Task. *Bilingualism: Language and Cognition*, 20(1), 162–183.

Vanhove, J. (2013). The Critical Period Hypothesis in Second Language Acquisition: A Statistical Critique and a Reanalysis. *PLoS ONE*, 8(7), e69172.

Wei, L. (2018). Translanguaging as a Practical Theory of Language. *Applied Linguistics*, 39: 9–30.

White, L. (1987). Against Comprehensible Input: The Input Hypothesis and the Development of Second Language Competence. *Applied Linguistics*, 8: 95-110.