Neuro-cognitive Processing of Morpho-syntax and Phonology in Late Second Language Learners

Erin Jacquelyn White
Department of Psychology
McGill University, Montreal

May, 2012

A thesis submitted in McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Psychology.

Table of Contents

Abstract ………………………………………………………………… 1

Résumé ………………………………………………………………… 4

Contribution of Authors ………………………………………………… 7

Acknowledgements ……………………………………………………….. 8

General Introduction …………………………………………………… 10

The Critical Period Hypothesis ………………………………………… 11

Multiple Critical Periods for Different Domains of Language ……… 14

The Shape of the Age-L2-Acquisition (AoA) Function ……………... 15

L1-L2 Similarity ……………………………………………………….. 20

Successful Late L2 learners ……………………………………………. 22

The Role of L1 Experience ……………………………………………. 25

Brain-based Studies of Age Effects …………………………………… 26

ERP Studies of L2 Processing …………………………………………. 31

The Present Studies …………………………………………………….. 39

Study 1:

Brain Responses Reveal Proficiency-Based Changes in Second Language Grammar Processing: An ERP Investigation of L1-L2 Differences in Late Second Language Learners ……………………………………………..…. 43

Abstract ………………………………………………………………… 44

Introduction …………………………………………………………….. 45

The present Study ………………………………………………………. 56

Methods ………………………………………………………………… 61

Participants …………………………………………………………….. 61

Stimuli …………………………………………………………………. 65

Procedure ……………………………………………………………… 67

ERP Recordings and Analysis ………………………………………… 68

Analysis of Behavioural Data ……………………………………… 72

Results ………………………………………………………………… 72

Behavioural Results …… 72
ERP Results: Analysis by L1 Groups (All Trials) …… 73
ERP Results: Analysis by L1 Groups (Correct Only Trials) …… 75
ERP Results: Comparison of “All Trials” vs. “Correct Only Trials” 77
Relationship between ERP Results and Behavioural Performance … 79
Discussion …… 80
The presence and magnitude of the P600 reflects L2 proficiency … 81
L1 reading strategies may influence the latency of the P600 …… 87
Future Research Directions …… 92
References …… 96

Tables

Table 1 …… 104
Table 2 …… 105
Table 3 …… 106
Table 4 …… 107
Table 5 …… 108
Table 6 …… 109

Figures

Figure 1 …… 110
Figure 2 …… 111
Figure 3 …… 112
Figure 4 …… 113
Figure 5 …… 114
Appendix 1 …… 115
Connecting Text – Study 1 and 2 …… 116

Study 2:
Native-like Phonological Processing by Late Second Language Learners:
ERP Evidence from Syllable- and Word-Level Processing …… 118
Abstract …… 119
Introduction …… 120

Experiment 1 …………………………………………………………. 127
Method …………………………………………………………….. 134
Participants ……………………………………………………. 134
Stimuli ………………………………………………………….. 137
Procedure ……………………………………………………… 138
EEG Recordings and ERP Analyses ………………………... 139
Results …………………………………………………………….. 142
Accuracy Results …………………………………………… 142
ERP Results …………………………………………………… 143
Analysis of Mastoid electrodes …………………………….. 147
Discussion ……………………………………………………….. 148
Experiment 2 ……………………………………………………….. 154
Method …………………………………………………………….. 159
Participants ……………………………………………………. 159
Stimuli ………………………………………………………….. 159
Procedure ……………………………………………………... 162
EEG Recordings and ERP Analysis ………………………... 163
Results ……………………………………………………………. 165
Behavioural Results ………………………………………… 165
ERP Results …………………………………………………… 166
Correlation Analyses ……………………………………….. 168
Discussion ……………………………………………………….. 169
General Discussion ……………………………………………… 172
References ……………………………………………………………. 178

Tables

Table 1 ……………………………………………………………. 189
Table 2 ……………………………………………………………. 190
Table 3 ……………………………………………………………. 191
Table 4 ……………………………………………………………… 192
Table 5 ……………………………………………………………… 193

Figures

Figure 1 …………………………………………………………… 194
Figure 2 ……………………………………………………………. 195
Figure 3 ……………………………………………………………. 196
Figure 4 ……………………………………………………………. 197
Figure 5 ……………………………………………………………. 198
Figure 6 ……………………………………………………………. 199
Figure 7 ……………………………………………………………. 200
Appendix 1 ……………………………………………………………. 201
General Discussion ……………………………………………………… 202
General Summary …………………………………………………… 203
Neuro-cognitive Changes and L2 Learning …………………………… 205
Age of L2 Acquisition and L2 Proficiency Effects on L2 Neuro-Cognitive Processing ………………………………………………………………… 209
Future Directions ……………………………………………………….. 212
Summary: The State of Maturational Accounts of L2 Acquisition …….. 219
General References ……………………………………………………… 222

Abstract

The role of age of acquisition in determining level of ultimate second language (L2) attainment is one of the most longstanding and controversial issues in the field of L2 acquisition. In particular, it is unclear whether there are limits to the brain's ability to process a L2 learned after puberty. The research for this dissertation examined the neuro-cognitive processes that late (i.e., post-puberty) L2 learners use when processing L2 morpho-syntax and phonology in order to investigate: (1) how the neuro-cognitive bases of L2 processing change as late learners become more proficient in their L2 (Study 1); and (2) to what extent high proficiency late L2 learners recruit similar neuro-cognitive processing mechanisms as native speakers (Study 2).

Study 1 was a longitudinal study that tracked the neuro-cognitive changes associated with L2 learning in order to examine how L2 morpho-syntactic processing is shaped by a learner's first language (L1) background and L2 proficiency. Korean- and Chinese late-L2-learners of English were tested at the beginning and end of a 9-week intensive English-L2 course. Event-related potentials (ERPs) were recorded while the learners read English sentences containing violations of regular past tense. This grammatical structure operates differently in Korean and does not exist in Chinese; previous research findings would predict that this would be difficult for these learners to acquire and process. By the end of L2 instruction, significant P600s were observed for both L1 groups that were not present at the start of instruction. Across all participants, larger P600 effects at session 2 were associated with higher levels of behavioural performance on an online grammaticality judgment task. These findings suggest that the neuro-cognitive processes underlying the P600 (e.g., “grammaticalization”) are modulated by individual levels of L2 behavioural performance and learning.

Study 2 was a cross-sectional study that examined the neuro-cognitive basis of L2 phonological processing in late learners. French speakers who were late L2 learners of English were tested in their processing of an English-specific phonetic contrast that is notoriously difficult for native French speakers to acquire: /h/ versus /Ø/. The L2 learners were classified into intermediate and high proficiency groups on the basis of their pronunciation skills and then, along with native English speakers, were tested on their processing of the non-native phonetic contrast in two stimulus/task conditions. In Experiment 1 the contrast was presented as syllables in an attended discrimination task (odd-ball paradigm). Three ERP components were examined to investigate automatic (MMN) and attention-driven stages of phonetic processing (N2b, P3b). In Experiment 2, the contrasts were presented as words and pseudo-words in a task that directed attention away from phonetic analysis. The N400 pseudo-word effect was used to index successful and automatic discrimination. High proficiency L2 learners displayed similar ERP effects as native speakers, indicating native-like (and automatic) processing, whereas intermediate proficiency L2 learners showed evidence of attention-driven but not automatic neuro-cognitive processing. The results suggest once late L2 learners advance to high levels of L2 proficiency they process non-native phonetic contrasts using similar (automatic) neuro-cognitive processes as native speakers.

In sum, the results of these studies suggest that a certain degree of plasticity remains in the neural systems supporting L2 grammar and phonological processing, even in adult learners. Native-like neuro-cognitive processing appears to be available to late L2 learners at relatively high levels of proficiency when processing certain aspects of their L2. Results are discussed with respect to the neuro-cognitive changes that are associated with L2 acquisition in late learners and whether there are maturational constraints that limit L2 acquisition and processing for late L2 learners.

Résumé

L’étendue du rôle de l’âge d’acquisition en tant que facteur déterminant du niveau de maîtrise d’une langue seconde (L2) figure parmi les questions les plus débattues dans la recherche sur l’acquisition de L2. Plus particulièrement, la question de savoir si le cerveau est limité dans sa capacité de traiter une L2 après la puberté demeure irrésolue. La recherche conduite dans le cadre de la présente thèse a pour but d’examiner les processus neurocognitifs qu’utilisent les apprenants tardifs d’une L2 (après la puberté) dans le traitement de sa morphosyntaxe et de sa phonologie dans le but de comprendre les facteurs suivants : (1) comment les bases neurocognitives du traitement de L2 changent-elles à mesure que les apprenants tardifs atteignent des niveaux supérieurs de maîtrise de L2 (étude 1) ; et (2) dans quelle mesure les apprenants tardifs de L2 de niveau avancé recrutent-ils des mécanismes neurocognitifs similaires à ceux des locuteurs natifs (étude 2).

La première étude est une recherche longitudinale visant à retracer les changements neurocognitifs associés à l’apprentissage de L2 afin de comprendre comment le traitement morphosyntaxique de L2 est influencé par la langue maternelle (L1) de l’apprenant et de son niveau de maîtrise de L2. Des apprenants tardifs de l’anglais d’origine coréenne et chinoise ont été testés en début et fin d’un cours intensif d’anglais L2 d’une durée de neuf semaines. Des potentiels évoqués (PÉs) ont été enregistrés alors que les apprenants lisaient des phrases en anglais contenant des violations de la règle du passé régulier. Cette structure grammaticale fonctionne différemment en coréen et est inexistante en chinois, et

les données issues de recherches précédentes permettaient de prédire qu'elle poserait des problèmes d'acquisition et de traitement chez ces apprenants. A la fin de la période de cours, des ondes P600 significatives pouvaient s'observer dans les deux groupes L1, lesquelles n'étaient pas présentes au début de la période de cours. Chez tous les participants, les effets P600 plus larges observés lors de la deuxième session étaient associés à des niveaux plus élevés de performance dans une tâche « on-line » de jugements de grammaticalité. Ces données suggèrent que les processus neurocognitifs sous-tendant la P600 (c'est à dire la « grammaticalisation ») sont modulés par des niveaux individuels de performance et d'apprentissage.

La seconde étude est une recherche transversale visant à examiner la base neurocognitive du traitement phonologique de L2 chez des apprenants tardifs. Des locuteurs francophones tardifs dans l'apprentissage de l'anglais L2 ont été testés dans leur traitement d'un contraste phonétique spécifique à l'anglais et reconnu pour être difficile à maîtriser par des locuteurs francophones : /h/ versus /Ø. Les apprenants tardifs ont préalablement fait l'objet d'un classement en groupes « intermédiaire » ou « avancé » sur la base de leur niveau de prononciation. Leur traitement du contraste phonétique non-natif a ensuite été enregistré en même temps que celui de locuteurs natifs dans deux conditions stimulus/tâche. Dans la première expérience, le contraste était présenté sous forme de syllabes lors d'une tâche attentionnelle de discrimination (paradigme « odd-ball »). Trois composantes PÉs ont été examinées afin d'étudier les étapes automatiques (MMN) et attentionnelles (N2b, P3b) du traitement phonologique. Dans la seconde expérience, les contrastes étaient présentés sous forme de mots et pseudo-mots dans une tâche permettant de diriger le faisceau attentionnel en dehors de l'analyse phonétique. L'effet N400 observable au niveau des pseudo-mots a été utilisé comme marqueur de discrimination réussie et automatique. Les effets PÉs chez les apprenants L2 très compétents étaient similaires à ceux des locuteurs natifs, indiquant un traitement de type natif (et automatique), tandis que les PÉs chez les apprenants L2 intermédiaires indiquaient un traitement neurocognitif de type attentionnel mais non automatique. Ces résultats suggèrent que des apprenants tardifs atteignant des niveaux élevés de compétence en L2 traitent des contrastes phonétiques non natifs au travers de processus neurocognitifs similaires à ceux de locuteurs natifs.

En somme, les résultats de ces études suggèrent qu'un certain niveau de plasticité demeure dans les systèmes neuraux sous-tendant le traitement grammatical et phonologique d'une L2, y compris à l'âge adulte. Un traitement neurocognitif de type natif semble être accessible à des apprenants tardifs d'une L2 à des niveaux relativement élevés de compétence en L2. Ces résultats sont examinés dans le cadre de la question des changements neurocognitifs associés à l'acquisition d'une L2 chez des apprenants tardifs et des contraintes développementales limitant l'acquisition et le traitement d'une L2 chez des apprenants tardifs.

Contributions of Authors

The two studies included in this dissertation were written as manuscripts and were co-authored by Fred Genesee and Karsten Steinhauer. Study 2 was also co-authored by Debra Titone. Drs. Genesee, Steinhauer and Titone contributed in a supervisory capacity to these studies by assisting in the development of the hypotheses and conceptual framework for each study, and it editing of the manuscripts. I contributed the initial ideas and research questions. The stimuli used in both studies were developed by me in collaboration with Dr. John Drury and Dr. Steinhauer (Study 1) and in collaboration with Dr. Genesee, Dr. Steinhauer and Dr. Titone (Study 2). I was responsible for participant recruitment, data collection, conducting analyses, and I drafted and revised the manuscripts. For both studies, Dr. Gensee, Dr. Steinhauer assisted in the interpretation of the data and provided valuable feedback on the manuscripts; Dr. Titone also provided feedback on Study 2.

I supervised two undergraduate honours studies during the course of my doctoral studies, Wing Yee Chow and Masha Westerlund. They were involved in testing participants, developing stimuli and analysing ERP data. Their involvement is recognized in the Acknowledgements sections. I have presented the results of both studies at various conferences. Study 2 has been submitted for publication to *Journal of Memory and Language.*

Acknowledgements

Looking back on this Ph. D. journey, it is clear that have arrived where I am because of the support I received from many individuals along the way.

To Fred and Karsten: I can't say enough how fortunate I have been to have had the opportunity to work with both of you, learn from you, and grow intellectually and professionally with you as my guide. Your passion and dedication to research, ability to simultaneously see the large picture as well as the details has been inspiring. You gave me the direction and structure to focus and root my ideas as well as the independence and creative feedback to let them soar. Thank you, both of you, for all of those hours you spent revising, and re-revising, my work; for everything I learned in our meetings in your office and on the back steps at Beatty Hall. Thank you for you guidance, continual support and belief in me.

To Deb: Thank you for including me into your family of grad students, for sharing ideas, for your smart feedback, for helping me to grow my confidence. You have been a role model of how an accomplished academic, and mom, can succeed. Thank you for everything you have taught me.

To Zoey (Wing Yee) and Masha: I was so fortunate to have had the opportunity to work with you on your honour's projects. Your hard work, dedication, and smart ideas put the studies here in motion. Thank you.

To Rhonda Amsel: who never hesitated to open her doors to help any student in the department who needed her advice and expertise. Thank you.

To my office mates who became my friends: thank you for the camaraderie, for the feeling like we were not alone in this process but working side by side towards similar goals, for all of the coffee, yoga, nights out and ideas we shared together. To Inbal, Efrat, Monika and Julie especially who were there all along the way.

To my family, Mom, Dad, Steph, Baba, Dave, She, Carlannah, Drew: Thank you for your encouragement, support and unwavering confidence in me. For providing me with the opportunities that led me to start these projects in the first place. To Mom especially, for guiding the way. I am here because you believed in me and listened to my heart even when I didn't.

To Nora and Allison: You have become like sisters to me. Thank you for always being there.

To Julien: for all of those evening and weekends in the park, tickle fights, and even those early mornings in the rocking chair. For helping me reach that delicate balance between work and life. For teaching me so much as I watched you learn.

Thank you.

Lastly, I wish to express my appreciation for the financial support I received from grants awarded to me and my supervisors from the Quebec and Canadian funding agencies NSERC, SSHRC and FQRSC and for support from the Center for Research on Brain, Language and Music (CRBLM).

General Introduction

In an era of globalization, the study of second language (L2) acquisition is not only important for theoretical reasons, but also because it has widespread practical implications. For example, according to recent Canadian censuses, approximately 20 % of Canadians speak a language other than English or French as a first language (L1), with numbers reaching up to 40 % in large urban centres (Statistics Canada, 2001). Similarly, for Canadians who speak English or French as their L1, approximately 20% report knowledge of both official languages, with numbers over 50% in cities such as Montreal (Statistics Canada, 2006). Clearly, understanding the processes by which we acquire multiple languages is of high social relevance and has many educational implications.

The focus of this dissertation is to better understand the basis of one of the most commonly observed phenomena in L2 acquisition – on average, those who begin learning an L2 as children tend to attain higher levels of L2 proficiency than adult learners. In fact, it has been argued that it is virtually impossible to acquire full native-like competence in an L2 if acquisition of the L2 begins after a certain age (e.g., Abrahamsson & Hyltenstam, 2009). The role of age of L2 acquisition (AoA) in determining level of ultimate L2 attainment is one of the most longstanding and controversial issues in the field (e.g., Birdsong, 2006; Singleton & Ryan, 2004). In particular, it is unclear whether the age-related decline in the success with which individuals master an L2 applies to all individuals, to all aspects of L2 acquisition, and, in particular, whether it is due to maturational changes in the brain that result in a fundamental differences in how we learn and process languages at different stages of life. The research for this dissertation examined the neuro-cognitive processes that late (i.e., post-puberty) L2 learners use when processing L2 morpho-syntax and phonology in order to investigate: (1) the extent to which these processes differ from those used by native speakers, and (2) how they change with learning. The purpose of this work was to provide evidence pertaining to the question of whether there are maturational changes in the brain that limit what adults can learn in their L2, how they learn, and the level of L2 attainment they can hope to achieve.

The Critical Period Hypothesis

Both in the laboratory and on the street, there is a common assumption that early is better as far as learning an L2 is concerned. In terms of how quickly children acquire an L2 compared to adults, this does not appear, in fact, to be true. Given the same amount of L2 exposure, adults advance through early stages of L2 acquisition faster than adolescents who, in turn, are often faster than younger children (e.g., Krashen, Scarcella & Long, 1982, for a review see Marinova-Todd, Marshall & Snow, 2000). However, in terms of ultimate level of L2 attainment, on average, children tend to achieve higher L2 proficiency than those who begin learning as adults (see Harley & Wang, 1997; Hyltenstam & Abrahamsson, 2003, for reviews). The question is, why?

At the heart of the debate is the “critical period hypothesis” (CPH) and, in particular, the issue of whether there are age-defined neuro-cognitive limits on when learning must occur in order to attain native-like levels of L2 proficiency. As early as 1959, Penfield and Roberts suggested that the brains of children have a specialized capacity for learning languages that decreases with age, such that "for the purpose of learning languages, the human brain becomes progressively stiff and rigid after the age of nine" (p.236).The idea that (first) language acquisition occurs within a particular developmental window was emphasized further by Lenneberg (1967) who argued that specific maturational stages in the brain's development constitute both the "prerequisites and limiting factors for language development" (p.169). He believed that lateralization of language to the left hemisphere sets a limit on the capacity for full language acquisition to no later than puberty. More than 50 years later, many of these views are still being discussed in the field today with respect to both L1 and L2 acquisition.

"Critical periods" are a special class of a broader concept in developmental biology known as "sensitive periods". The term "sensitive period" refers to a limited period in development during which certain aspects of behaviour and brain function are readily shaped or altered by experience (Knudsen, 2004). This concept has been used to explain many age-limited behaviours in humans and other animals, most notably ocular representation in the primary visual cortex, filial imprinting, and songbird memorization (see Knudsen, 2004, for a review). The term "sensitive" period is often used interchangeably with "critical" period in the field of L2 acquisition, an approach that will be taken here, although there are important distinctions between the terms. Applied to language acquisition, a "critical period" posits a short and sharply defined window-of-opportunity during which language input causes irreversible changes in brain function that allow young learners to comprehend and produce language in a "native-like" way. In contrast, a sensitive period suggests a more gradual decline in the capacity to acquire language with age (Lamendella, 1977; Oyama, 1979). This conceptualization of age-related effects on language learning leaves open the possibility that brain function can continue to be altered to some extent after the critical period as proposed by Lenneberg has ended and suggests further that L2 acquisition that begins after the period has ended may result in some level of language skill, even if that level is below native-like levels (Knudsen, 2004; Patkowski, 1982).

In most formulations of the critical/sensitive period hypothesis, the basic tenet is that the capacity for native-like language attainment begins to decline sometime in childhood and disappears at or around puberty because of maturational processes occurring in the brain (for reviews, see Harley & Wang, 1997; Hyltenstam & Abrahamsson, 2003). It is thought that there is “something special about the maturational state of the child's brain which makes children particularly adept at acquiring *any* language, first as well as second” (p. 64, Johnson & Newport, 1989). Although many researchers might agree that developmental changes in the brain may explain the age-related decline in the success with which individuals can acquire their *first* language (Mayberry & Eichen, 1991), it is far from clear whether L2 acquisition must also occur within the same window-of-opportunity, and whether it is maturation of the brain that sets the limits on this time period. In this dissertation, the term “maturational accounts of AoA effects in L2 acquisition” will be used to refer to theories of L2 acquisition (including the CPH) that propose that it is the brain’s maturational state at the time of L2 acquisition which can account for the finding that many children appear to attain higher levels of L2 attainment than adults. One extreme position, known as the "Fundamental Difference Hypothesis", is that the neuro-cognitive processes that are available to children for acquiring languages (L1 or L2) are not available to adults who, therefore, have to rely on fundamentally different mechanisms (Bley-Vroman, 1989; 2009). However, as discussed in the following sections, the neuro-cognitive evidence to support this hypothesis is highly controversial. Intense debate continues in the field surrounding the issues of whether L2 acquisition is somehow restricted by maturation of the brain and, more specifically, by a critical/sensitive period ending around puberty. It is also unclear which aspects of L2 acquisition are most affected and if native-like attainment might, in fact, be possible for some late L2 learners. Some of these issues are reviewed in the following sections as they pertain to behavioural and then neuro-cognitive evidence.

Multiple Critical Periods for Different Domains of Language

Many researchers refer to the critical period for language in a non-differentiated fashion. Others have argued that, because language is not a unitary phenomenon, it may be more accurate to consider the notion of multiple critical/sensitive periods. Long (1990) proposed that multiple critical periods, each with its own onset and offset, constrain different linguistic sub-domains independently. For example, L2 exposure between ages 6-12 years of age may be required for native-like L2 pronunciation, before age 15 for complete L2 morpho-syntactic acquisition, and by some intermediate age for the remaining linguistic domains (such as discourse and pragmatics). Eubank and Gregg (1999) suggested that whereas phonology and syntax may be determined by the input received during a specific time window (i.e., the critical period), vocabulary learning may not be subject to the same age-constraints, if at all. They argue that, evolutionarily speaking, it would be advantageous for humans to retain the ability to learn new words throughout life; however, as the morpho-syntactic structure of a language remains constant, the ability to comprehend and produce grammatical sentences would be most effective if its acquisition occurred once during early childhood and then was not open to change. Similarly, it could be argued that once learners have established the phonetic category representations that correspond to meaningful speech sounds in a particular language, it would be advantageous for further development of that language if these categories remain stable (see Werker & Curtin, 2005, for a discussion of how the development of phonetic discrimination in infants bootstraps word learning). Many researchers agree that, if one or more critical/sensitive periods exist, they exert particular restrictions on successful acquisition of L2 syntax and phonology by late learners, whereas the ability to acquire new lexical-semantic information may remain open across the life span (Clahsen & Felser, 2006; Lamendella, 1977; Neville & Bavelier, 2000; Sanders, Weber-Fox & Neville, 2008; Scovel, 1988).

The Shape of the Age-L2-Acquisition (AoA) Function

One of the contentious issues surrounding the CPH, and AoA effects in general, with respect to the acquisition of L2 morpho-syntax is the shape of the AoA function. Although most researchers would agree that, generally speaking, there is some sort of relationship between AoA and L2 attainment, there is little agreement as to the nature of this relationship and whether it can be taken as support for the existence of a critical/sensitive period. A critical/sensitive period predicts a discontinuity in AoA effects across the lifespan. More specifically, the relationship between AoA and L2 attainment, according to the CPH, should be different for those who begin learning their L2 within the postulated period compared to those who begin learning after the period has ended (see Birdsong, 2006, for a summary). In fact, any maturational account of AoA effects in L2 acquisition would predict that once maturation is complete, L2 attainment should be irregular and low overall; and the relationship between AoA and performance should be different for early and late learners.

Such a pattern was found by Johnson and Newport (1989). In this widely cited study, the English grammar skills attained by Korean and Chinese speakers who began learning English as an L2 between the ages of 3 and 39 (when they arrived in the U.S.) were examined and compared to native English speakers. Their results on the relationship between the AoA of the learners and their performance on an oral grammaticality judgement task were taken as strong evidence in favour of maturational effects on L2 acquisition and of the CPH in particular. To be more specific, the L2 learners in their study who began L2 learning between the ages of 3-7 performed within the range of native speakers, and native-like grammatical competency deteriorated among learners who began acquiring the L2 after this age. For those who learned their L2 between the ages of 8-15 years, performance declined linearly as a function of AoA, marked by a strong negative correlation between AoA and performance. For the adult learners (those who had begun L2 learning between the ages of 17-39), however, there was no systematic relationship between AoA and L2 performance. Overall, those who began L2 learning as adults obtained significantly lower scores than either the native English speakers or the child-L2 learners and, as a group, they displayed substantial inter-individual variability, which resulted in a non-significant relationship between AoA and performance. Moreover, these age-effects were apparently independent of other factors, such as length of residence in the U.S., motivation, or amount of formal instruction. This pattern of high L2 attainment during early childhood, followed by decreasing abilities with increasing age until the end of puberty, subsequently followed by a low-level plateau of L2 performance, was taken as strong support for a maturational account of age effects in L2 acquisition. That there was no relationship between AoA and performance for those who began L2 learning after age 17 was particularly compelling because "presumably there are not many important maturational differences between, for example, the brain of a 17 year old and the brain of a 27 year old" (pg. 79). The authors concluded that the capacity to learn languages (L1 or L2) to native-like levels of attainment is highest during childhood, declines and then eventually disappears after puberty, once the brain has matured.

The Johnson and Newport (1989) study has been extremely influential, in particular by demonstrating the role that AoA plays in influencing L2 morpho-syntactic attainment. However, it has also raised considerably discussion as to whether these results are indeed evidence that maturation of the brain is the cause of these AoA effects and that a critical period constrains L2 acquisition in the same way as L1 acquisition, as concluded by the authors. In fact, other researchers have provided counter evidence with respect to the shape of the AoA-L2 performance function (e.g., Bialystok & Hakuta, 1999; Bialystok & Miller, 1999; Birdsong & Molis, 2001). Moreover, they have suggested that, even within the postulated critical period, learners with different L1 backgrounds may show different levels of attainment (e.g., Bialystok & Miller, 1999; Birdsong & Molis, 2001) and that some late L2 learners can attain native-like levels of L2 attainment (Birdsong, 1992; Bongaerts, 1999; White & Genesee, 1996). It has also been argued that other factors related to L2 exposure and input may also account for the AoA-related decline in L2 performance (e.g., Marinova-Todd et al., 2000), all of which are incompatible with maturational accounts of L2 acquisition and of the CPH specifically. We review some of this evidence below.

In the Johnson and Newport (1989) study, the strongest evidence for maturational effects was the non-continuous function between AoA and L2 performance. However, not all studies have reported the same relationships between AoA and L2 performance across the lifespan. For example, in a re-examination of Johnson and Newport's data, Bialystok and Hakuta (1999) found a significant negative correlation between AoA and L2 performance among both early and late L2 learners when the cut-off between early and late learning was set at 20 years rather than 17, as in the original study, suggesting that there may not be an important change in performance at puberty. Birdsong and Molis (2001) tested Spanish-speaking L2 learners of English using the same stimulus materials and the same grammaticality judgement task as Johnson and Newport and reported a significant correlation between AoA and performance for late L2 learners (AoA ≥ 17 or AoA ≥ 20). In contrast, for early learners, the correlation was not significant because all participants performed at ceiling and within the range of native-speakers. Finally, Bialystok and Miller (1999) examined grammaticality judgement performance of Spanish- and Chinese-speaking L2 learners of English and found a significant negative correlation between AoA and performance for both older and younger learners in both language groups. In fact, in all of these studies, when all individual learners are treated together, AoA predicted performance over the entire span of AoA tested (i.e., not only for early learners; see also Hakuta, Bialystok & Wiley, 2003, for similar findings from a large scale study based on data taken from the U.S. census comprised of self-reported levels of proficiency among 2.3 million U.S. immigrants). As summarized by Birdsong (2006), behavioural data are, overall, inconsistent with the notion of a critical period for L2 acquisition in which there is either a "period of peak sensitivity whose end coincides with the end of maturation or with a levelling off of sensitivity whose beginning coincides with the end of maturation" (pg. 19). These studies suggest that there is a more general decline in level of L2 attainment across the lifespan and that puberty does not appear to be a very important maturational turning point for L2 acquisition – both of which are inconsistent with the CPH.

L1-L2 Similarity

In addition to demonstrating that AoA may correlate with L2 attainment across the lifespan, and not only before puberty, the studies reviewed in the previous section highlight another important finding – namely, that typological similarity between the L1 and L2 can influence L2 attainment in both early and late L2 learners. Birdsong and Molis (2001) tested native Spanish speakers who had acquired English as a L2 at various ages using the same test material that Johnson and Newport (1989) had used with Chinese- and Korean- L2 learners of English, allowing a direct comparison between studies. Similarly, Bialystok and Miller (1999) tested Spanish- and Chinese-speaking L2 learners of English in order to compare their performance on a grammaticality judgement task as a function of L1 background. For both early and late learners, these studies revealed higher overall L2 (English) performance for native Spanish- compared to native Chinese-speakers (Bialystok & Miller, 1999; Birdsong & Molis, 2001). Bialystok and Miller (1999) argued that finding differential L2 performance as a function of L1 background for younger learners, in particular, is difficult to align with the notion of critical/sensitive period that governs L2 acquisition in a uniform way for all learners. They reasoned that if the mechanisms underlying language learning are open and available during a critical/sensitive period, then children should be equally effective at acquiring any L2; a prediction which was not supported by their results.

Bialystok and Miller (1999) also compared the performance of the early and late L2 learners in their study as a function of whether the L2 morpho-syntactic structures tested in specific test items were used in the learners' L1. They reasoned that if there is a critical period for L2 acquisition, then as long as an individual begins acquisition within that period, he/she should be able to, in principle, acquire any L2 morpho-syntactic structure regardless of whether it is present in the L1. At the same time, if the mechanisms underlying language acquisition are no longer available for L2 acquisition after the critical period, then adult learners would need or might be prone to make use of L1-transfer to aid L2 acquisition. Under this logic, adult L2 learners would be expected to have relatively more difficulty than child L2 learners acquiring L2 morpho-syntactic structures that are not used in the L1. However, Bialystok and Miller (1999) did not find evidence for this. Chinese speakers, regardless of their AoA, had difficult acquiring L2 morpho-syntactic structures that are not present in Chinese. In contrast, Spanish speakers, regardless of their AoA, were able to acquire virtually all structures. Taken together, these results highlight two important findings: (1) L1 transfer was similar for L2 learners of a given language group regardless of whether they began learning before or after the postulated critical/sensitive period; and, (2) overall, L2 performance was influenced by the typological similarity between a learner's L1 and L2 (see also Bialystok & Hakuta, 1999). Bialystok and Miller (1999) argued that whatever effects L1-L2 similarities exert on L2 learning, they do not appear to change with the AoA of the learner, which is inconsistent with maturational accounts of L2 acquisition. The results of their study highlight the need for additional research to more fully understand the mechanisms by which L1 and L2 interact to shape L2 acquisition and processing, an issue that is discussed in more detail in both Study 1 and 2 of this dissertation.

Successful Late L2 learners

Many of the studies reviewed so far have examined the overall relationship between AoA and L2 attainment and have shown that, in general, younger learners are more likely than older learners to attain native-like levels of L2 proficiency in the long run. They were not designed to test whether native-like attainment is possible for late L2 learners or how prevalent this is – a more precise test of the CPH (Long, 1990). This raises the question of whether we are really testing what is ultimately possible for late L2 learners by examining what the average person is able to accomplish. White and Genesee (1996) took a different approach. They first recruited highly proficient L2 learners, selected those whose oral performance was judged to be virtually indistinguishable from that of native speakers (based on subjective ratings of L2 learners' language samples), and then compared their underlying grammatical competence in the L2 to that of native speakers using a grammaticality judgement task. They found that, although rare in the population at large, there were some late L2 learners who performed in an identical way to native speakers on the grammaticality judgement task designed to tap into aspects of Universal Grammar that are thought to be difficult for L2 learners to acquire. Thus, despite learning their L2 later in life (i.e., after the postulated critical period), the performance of these late L2 learners suggest that they had indeed obtained native-like levels of competence.

Together, the results of this study and other studies that have similarly identified late L2 learners whose L2 performance was virtually indistinguishable from that of native-speakers (e.g., Birdsong, 1992; Bongaerts, 1999) have been used to argue that native-like L2 attainment may be possible for some late L2 learners in some domains of L2 acquisition (although see Abrahamsson & Hyltenstam, 2009, for a different point of view). Some researchers have argued that successful L2 learners are exceptional outliers and not truly representative of what any and all late L2 learners can achieve (e.g., Bley-Vroman, 1989), whereas others (e.g., Birdsong, 1999) maintain that, although few in number perhaps, such learners call into question the view that universal, time-locked modifications of the brain result in uniform low levels of L2 attainment among all L2 learners. At the very least, if there are maturational constraints on L2 attainment, they do not appear to constrain all aspects of L2 acquisition in all late learners.

In Study 2, high proficiency participants were identified in a similar way as in White and Genesee (1996). Participants' L2 proficiency was first pre-screened with an informal telephone interview and then was assessed further based on an evaluation of their spontaneous speech samples by native speakers. Once high proficiency individuals had been identified, their neuro-cognitive processing mechanisms were compared to those of native speakers. Identifying and testing high proficiency L2 learners in this way, allows for a more precise test of whether native-like attainment and processing is possible for late L2 learners.

Amount of L2 Use, Motivation to Learn and Education

An alternative to the CPH and maturational accounts of age effects in L2 acquisition is that AoA effects may have very little to do with maturation of the brain and, instead, may reflect other factors that are correlated with the age. For example, for early and late L2 learners, performance has been found to be positively correlated with amount of L2 use at the time of testing (Birdsong & Molis, 2001) and negatively correlated with L1 use (Piske, MacKay & Flege, 2001), results which, taken together, highlight an important role for practice in determining level L2 attainment. Amount of formal education is also often a correlate of L2 attainment, suggesting a role for more general social-economic factors in influencing learning outcomes (Hakuta et al., 2003).

Amount of education in an L2 environment specifically may play an important role in contributing to the overall high and relatively uniform levels of attainment that are often observed for younger learners, contrasted with the generally low and more variable levels of attainment among older learners (Marinova-Todd et al., 2000). Experience in a public school system that uses the L2 as the primary language of instruction would provide daily opportunities for L2 use and allow for rich social interactions with peers. Developing friendships with native speakers of the target language would result in increased motivation to learn the target language and desire to identify with the target language culture, both of which are associated with higher levels of L2 attainment (e.g., Gardner, 1985; Moyer, 2007). Moreover, at school, young L2 learners must learn not only how to communicate in the L2, they must learn how to use the L2 in order to express complex and abstract ideas in order to succeed academically. This requires that they develop L2 proficiency skills that allow them to communicate in cognitively demanding situations without the aid of concurrent contextual cues (Cummins, 2000; Genesee, Paradis & Crago, 2004). For many adult learners who begin acquiring an L2 after they have been educated in their L1 (e.g., adult immigrants), this level of L2 proficiency may not be required or expected. Thus, age may well serve as a mediator in the relationship between the availability of enriched opportunities for L2 learning and level of L2 attainment, obviating the need to hypothesize a critical period governing the acquisition process.

The Role of L1 Experience

An alternative to maturational accounts of L2 acquisition is that it is the experience of L1 learning and neural commitment to patterns in the L1, rather than general maturation of the brain, that accounts for why adults tend to achieve lower levels of L2 attainment than children (e.g., Hernandez, Li & MacWhinney, 2005; Kuhl, 2004; Marchman, 1993). According to Doupe and Kuhl (1999) this is still somewhat consistent with the CPH, as prior L1 learning and experience may play a role in closing the critical/sensitive period. In contrast, Hernandez et al. (2005) argue that, compared to early L2 learners, late L2 learners experience stronger L1-entrenchment and competition from L1 representations and processing routines and that this could account for AoA effects in L2 acquisition, without need to invoke age-bounded biological capacities that might be restricted by a critical period. Both Kuhl and MacWhinney's experience-based accounts predict that early L2 learners face less L1 entrenchment and, thus, retain greater plasticity than adult learners. As a result they both predict lower, more variable levels of L2 attainment for late L2 learners compared to early learners, which is consistent with maturational accounts of L2 acquisition.

In contrast to maturational accounts however, these theories are better able to account for why similarities between the L1 and L2 can facilitate L2 attainment in both early and late L2 learners (see previous section on L1-L2 Similarities). During L2 processing, the L1 neuro-cognitive network is thought to be active; the more lexical, phonological, grammatical features the L1 and L2 share, the more this co-activation will be useful and will facilitate L2 processing and acquisition (MacWhinney, 2005). Importantly, MacWhinney's competition model predicts that it will be easier to acquire L2 structures that are completely absent in the L1, compared to L2 structures that exist in the L1 but are realized differently in the two languages and compete for processing. This view contrasts with Universal Grammar (UG) based theories, which would predict that late L2 learners will experience difficulty in acquiring L2 structures that are not used in the L1 at all (White, 2003). These L1-experience accounts are discussed in more detail in the context of Study 1 and 2.

Brain-based Studies of Age Effects

As noted earlier, central to all maturational accounts of L2 acquisition, including the CPH, is the idea that maturation of the brain and its ability to learn language is primarily responsible for the differential success rates of adult and child learners. For example, reduced neural plasticity (Penfield & Roberts, 1959) and changes in the degree of cortical myelination (Pulvermueller & Schumann,

1994) that occur around puberty have been postulated to account for age-related declines in the ability to learn and display high levels of L2 proficiency. It then follows that an ultimate test of the CPH, and maturational accounts more generally, is to examine the extent to which the neural bases of L2 learning and processing differ as a function of AoA. For example, are "fundamentally different" neuro-cognitive mechanisms used for processing early- and late-acquired languages? Formulating the issue of AoA-effects in L2 acquisition in neuro-cognitive terms could be fruitful because it is ultimately these neuro-cognitive systems that are both responsible for language learning and that are most directly affected by the hypothesized critical/sensitive period thought to constrain this learning.

Two of the most commonly used techniques for studying the neural basis of language processing today are functional magnetic resonance imaging (fMRI) and event-related-potentials (ERPs). fMRI studies of L2 processing examine changes in the ratio of oxygenated to deoxygenated blood (i.e., changes in the BOLD signal -- blood-oxygenation-level-dependent fMRI) in particular brain areas that occur as a result of engaging in particular linguistic tasks in the L1 compared to the L2 (see Sabourin & Stowe, 2005, for a review). This technique provides a high-spatial resolution image of the brain areas that are recruited during language processing and can reveal how factors such as AoA may influence the extent to which the same areas are used for L1 and L2 processing. However, despite its excellent spatial resolution, fMRI, with its moderate temporal resolution, has difficulty differentiating events that occur within a second or two of each other. Moreover, measuring changes in blood oxygen levels provides only an indirect measure of neural activity. In contrast, ERPs, despite their rather poor spatial resolution, have excellent temporal resolution and this makes them well suited for examining how specific linguistic structures in the L1 and L2 are processed, as this processing unfolds in time. ERPs are the focus of this dissertation and, thus, are the focus of the review that follows.

ERPs measure the electrical voltage changes that occur when large groups of neurons are activated in synchrony in response to a particular stimulus. Under the assumption that different cognitive processes manifest themselves in different patterns of neural activity, if reliable ERP differences are observed when a group of participants engage in different tasks or when two groups engage in the same task, it suggests that the cognitive processes employed also differ to some extent (Otten & Rugg, 2005). These differences can be both quantitative and qualitative. Quantitative differences include differences in the amplitude of a particular ERP response which, in turn, suggest differences in the degree or consistency with which a particular neuro-cognitive process is recruited. Differences in timing of specific ERP components – that is, differences in the onset, peak latency, or duration of a response, are interpreted to reflect differences in when the processes are recruited and how long those processes are engaged. Qualitative differences in ERP responses include differences in the topographical distribution of ERPs across the scalp or polarity differences which are interpreted to reflect the use of different neural generators or even completely different neuro-cognitive processes (Otten & Rugg, 2005). Thus, ERPs can provide sensitive information about both the neural and cognitive processes that participants may recruit when presented with particular stimuli or engaged in specific tasks.

ERPs are particularly useful for studying language processing because they have an extremely high temporal resolution (in the range of milliseconds) and, therefore, can differentiate between specific aspects of linguistic processing (e.g., phonological, orthographic, lexical-semantic, morpho-syntactic). For example, in native speakers, distinct ERP components (waveforms with positive or negative polarity) have been reliably associated with lexical-semantic and morpho-syntactic aspects of processing. This has been taken as strong evidence that "meaning-" and "structure-" related aspects of language processing recruit distinct neuro-cognitive networks (Hahne & Friederici, 1999). Lexical-semantic processing is associated with a negative-going ERP component, the N400, a widely replicated and reliable measure that is interpreted to reflect lexical representation, access, and activation (Lau, Phillips & Poeppel, 2008; see Study 2 for more information). In contrast, morpho-syntactic processing is typically associated with a positive-going ERP component, the P600, elicited at central-parietal electrodes approximately 600 ms after the critical word in a sentence (Hagoort, Brown & Groothusen, 1993; Osterhout & Holcomb, 1992). In many studies, the P600 is preceded by another ERP response, the left-anterior negativity ("LAN") approximately 300-500 ms after the critical word. There is evidence to suggest that the LAN is associated with automatic or implicit rule-based grammar processing, whereas the P600 is associated with later, controlled processing of the same grammatical structures (Hahne & Friederici, 1999; see Steinhauer, White &

Drury, 2009 for a review). Morpho-syntactic L2 processing was the focus of Study 1 (more information about the LAN/P600 components are given in Study 1).

A similar distinction between automatic and controlled, attention-driven aspects of processing has also been found during phonological processing. More specifically, the mismatch negativity (MMN) has been taken to reflect early, automatic and language-specific aspects of phonological processing, whereas the N2b and P3b components are thought to reflect later, domain general aspects of processing that involve attention (for reviews, see Näätänen, Ternaiemi, Sussman, Paavilainen & Winkler, 2001; Näätänen, Paavilainen, Rinne & Alho, 2007). Phonological L2 processing is the focus of Study 2 and more information is provided there about the MMN/N2b/P3b components. These distinctions highlight how ERPs can measure multiple neuro-cognitive processes that may underpin a particular behavioural response. They are, thus, ideal for examining whether late L2 learners who have attained high levels of L2 proficiency, as indexed by behavioural performance, display similar underlying neuro-cognitive processing profiles as native speakers. Moreover, changes in ERP responses have been observed in L2 learners before improvement can be measured on a behavioural level, highlighting how ERPs provide a sensitive measure of cognitive and language processing and learning (McLaughlin, Osterhout & Kim, 2004; Tremblay, Kraus & McGee, 1998).

ERP Studies of L2 Processing

Early ERP studies of L2 processing reported results that initially appeared to support the claim that AoA plays a critical role in determining access to the neuro-cognitive mechanisms that native speakers use to process their L1 by showing that late L2 learners use different neuro-cognitive mechanisms for L2 morpho-syntactic processing compared to both native speakers and early L2 learners. In the first ERP study to examine the neuro-cognitive basis of L2 semantic and syntactic processing, Weber-Fox & Neville (1996) compared native English speakers and five groups of Chinese-L2-learners of English who differed in their age of L2 acquisition (age 1-3, 4-6, 7-10, 11-13, 16+ years). Participants were presented with well-formed, meaningful sentences and sentences that contained semantic or syntactic violations. In response to semantically anomalous sentences, all of the groups exhibited qualitatively similar N400 responses, although the latency of this effect was somewhat delayed in the Chinese speakers who began L2 learning after the age of 11. These results were interpreted to suggest that the neuro-cognitive mechanisms underlying L2 semantic processing (e.g., lexical access, semantic integration) are largely similar to those used by native speakers and that AoA may exert its effects primarily by influencing the speed at which these processes are engaged.

Weber-Fox and Neville's results indicated that late L2 acquisition had a much more pronounced effect on syntactic processing as shown by differences in the distribution, latency, and amplitude of the LAN and P600 components in late learners in comparison to both early L2 learners and native speakers. When presented with sentences containing phrase structure violations (e.g., *The scientist criticized Max's* *__of__* *proof the theorem*), the early L2 learners (AoA < 11) in their study elicited a left-lateralized negativity between 300-500 ms after the onset of the violation that was similar to the LAN observed in the native speakers. The late L2 learners (AoA > 11) also exhibited a negativity in this time window; however, its topographical distribution was markedly different from that of the native speaker and early L2 learners subgroups. In the oldest L2 learners group (AoA >15), this component resembled a N400 more than a LAN. Similarly, the P600 exhibited by the early L2 learners (AoA < 11 years) was identical to that of native speakers, was delayed in the L2 learners who acquired English between 11-13 years, and was not present at all for the oldest group of L2 learners.

Together with subsequent work with Japanese-late-L2-learners of German (Hahne & Friederici, 2001), these results suggested two important conclusions. First, the neuro-cognitive basis of semantic processing may be relatively unaffected by AoA and, in fact, may rely on qualitatively similar mechanisms in native speakers and late L2 learners. Second, in contrast to semantic processing, grammar processing may be more sensitive to delays in L2 acquisition and may recruit fundamentally different neuro-cognitive systems in late L2 learners compared to both native speakers and early L2 learners. Theses results have been interpreted as important neuro-cognitive evidence for limits on L2 processing that are determined by AoA, and thus as support for the CPH in L2 morpho-syntax.

However, because the proficiency level of the L2 learners in these studies was not controlled or accounted for, it is not possible to conclude from these results that group differences in ERP effects reflect only, or even primarily, AoA effects. More specifically, in the Weber-Fox and Neville study, the L2 learners' AoA was negatively correlated with their level of L2 proficiency, as measured by standardized tests of English grammar, self-reported L2 proficiency, and performance on the grammaticality judgment task that was conducted during ERP recording. The early L2 learners also reported using English (their L2) more than Chinese, even at home, and being more comfortable in English than Chinese – that is they were likely more dominant in their L2 than their L1. In contrast, the late L2 learners reported higher levels of use and comfort in their L1 compared to their L2. Thus, it is impossible to determine whether the absence of LAN/P600 effects (native-like processing profiles) in the late L2 learners was due to their AoA or their lower levels of L2 proficiency (see also Hahne & Friederici, 2001; Kim, Relkin, Lee & Hirsh, 1997, for similar confounds).

In fact, findings from neuro-imaging (PET/fMRI) studies suggest that L2 proficiency level plays an important role in whether late L2 learners show similar patterns of cortical activation as native speakers during L2 processing (Abutalebi, Cappa & Perani, 2001; Perani et al., 1998). Moreover, ERP studies show that when adults are trained to high levels of proficiency in an artificial language and are presented with sentences that violate the grammatical constraints of that language, they exhibit the same LAN/P600 effects that are thought to be reserved to native speakers and early L2 learners (Friederici, Steinhauer & Pfeifer, 2002; Morgan-Short, Steinhauer, Sanz & Ullman, 2012). These findings suggest an important alternative conclusion to that drawn by Weber-Fox and Neville (1996);

namely, whether or not L2 learners recruit the same neuro-cognitive processes as native speakers may depend on their level of L2 proficiency, not their age of L2 acquisition. This alternative idea does not fit comfortably with maturational accounts of L2 acquisition which would predict that native-like neuro-cognitive processes should be unavailable to late L2 learners. In order to unequivocally examine the extent to which AoA influences the ability to engage in native-like language neuro-cognitive processing, it is essential to systematically disentangle the role played by L2 proficiency versus AoA.

A study by Steinhauer, White and Genesee (2012; see also Steinhauer, White, Cornell, Genesee & White, 2006) sought to do just that. They tested the processing of English phrase structure violations in English native speakers and four groups of late L2 learners of English (AoA > 15 years). The L2 learners had attained intermediate or high levels of English proficiency, overall, and spoke either French or Chinese as an L1. By comparing late L2 learners who differed in their level of L2 proficiency, it was possible to examine whether L2 proficiency or AoA was the more important factor in influencing the neuro-cognitive substrates for L2 morpho-syntactic processing. In addition, by comparing L2 learners who spoke Chinese or French as an L1 (and were matched on L2 proficiency), it was possible to examine the extent to which typological differences between the L1 and L2 influenced processing, another factor that could explain the absence of LAN/P600 effects in the Chinese-English and Japanese-German L2 learners tested in the Weber-Fox and Neville (1996) and Hahne and Friederici (2001) studies.

Participants in the Steinhauer et al. study read correct sentences and sentences that contained word category violations which were created by exchanging the position of the verbs and nouns in the well-formed sentences (e.g., Correct sentence: "*The man hoped to enjoy the meal with friends*."; violation sentence: *The man hoped to *meal the *enjoy with friends)*. Native speakers elicited three ERP components: a N400 (300-400 ms), a left-lateralized negativity (LAN, 400-500 ms), and a subsequent P600 (600-1100 ms). Identical "native-like" effects were observed in the native English speakers and both groups of high proficiency late L2 learners (French and Chinese native speakers) in all of the time windows investigated. Thus, despite acquiring their L2 after puberty, the late L2 learners who had attained high levels of L2 proficiency processed L2 phrase structures using the same neuro-cognitive mechanisms as native speakers. The presence of a LAN was particularly noteworthy insofar as previous studies have suggested that it may be the hallmark of rapid, automatic, (perhaps implicit) "native-like" syntax processing (Clahsen & Felser, 2006; Friederici, 2002; Ullman, 2001), and it is precisely automatic aspects of L2 acquisition that has been argued to be constrained by maturational effects since the original formulations of the CPH (Lenneberg, 1967).

The low proficiency L2 groups, in contrast, did not display left-lateralized negativities and, although they exhibited P600s, they were more broadly distributed than in the native speakers and high proficiency L2 groups. As well, whereas the low proficiency French-L1 group exhibited an N400 effect, the low proficiency Chinese-L1 group did not. Together, these results suggest that L2 proficiency plays an important role in late L2 learners' use of the neuro-cognitive mechanisms that underlie grammar processing in native speakers (i.e., the LAN/P600). Moreover, at high proficiency, the use of these processes may be relatively independent of L1-L2 similarities, at least when processing rather salient and straight forward word category violations.[1] In contrast, at low levels of L2 proficiency, L1-L2 similarities may play a role in influencing L2 syntactic processing. Thus, regardless of late AoA and different L1 backgrounds, the L2 learners with high levels of L2 proficiency engaged similar neuro-cognitive processing mechanisms as the native speakers when processing L2 morpho-syntax, and, more specifically, they exhibit the biphasic LAN/P600 response that is typically found in native speakers.

Native-like LAN/P600 effects have also been reported in other recent ERP studies of L2 morpho-syntactic processing in high proficiency late L2 learners (Bowden, Sanz, Steinhauer & Ullman, 2012; Dowens, Vergara, Barber & Carreiras, 2010; Rossi, Gugler, Friederici & Hahne, 2006). Moreover, the importance of proficiency in eliciting these neuro-cognitive processes is underlined by Pakulak and Neville's (2010) finding that adult native speakers who are less proficient in their L1 (as revealed by their performance on standardized tests of English oral language grammar skills) displayed less left-lateralized LANs (i.e., "AN" effects) and P600s with smaller amplitudes compared to more proficient adult native speakers. Thus, in contrast to the conclusions drawn by

---

[1] That L1-L2 similarities influence L2 processing to a greater extent in L2 learners who have attained relatively low levels of proficiency compared to those who have attained higher levels of L2 proficiency could also account for the findings of Bialystok and Hakuta (1999). In that study the Chinese speakers, who were more affected by L1-L2 similarities than the Spanish speakers, were also overall less proficient in their L2.

some previous ERP studies of L2 morpho-syntactic processing that attribute ERP differences between native speakers and late L2 learners to AoA (thus apparently supporting a maturational account of AoA effects in L2 acquisition; e.g., Hahne and Friederici, 2001; Weber-Fox & Neville, 1996), more recent studies that teased apart AoA and proficiency effects indicate that L2 proficiency may be a more important factor in eliciting native-like neuro-cognitive processing profiles. However, the issue is by no means settled. For example, Pakulak and Neville (2011) compared the ERP effects elicited by sentences containing a phrase structure violation in late L2 learners of English and native English speakers who were matched in their (relatively low) level of English grammatical proficiency, but who differed in the age when they acquired English (i.e., as an L1 or L2). They found that whereas the native speakers exhibited an early and long lasting early anterior negativity ("AN") followed by a P600, the late L2 learners exhibited only a P600 effect. Consistent with some of the early ERP studies discussed previously, the authors interpreted this to mean that the early and automatic processes that are thought to be indexed by LAN and AN effects in native speakers may be governed by maturational constraints in the brain that occur early in life (however, see Steinhauer et al., 2012, for a discussion of possible methodological confounds in this study that could account for the difference between the native speakers and late L2 learners). It is clear that more work is needed to obtain a deeper understanding of the relative contributions of AoA and level of proficiency to the neuro-cognitive basis of L2 processing.

In summary, although many researchers agree that AoA can play an important role in influencing the likelihood that individuals will attain high levels of L2 proficiency (Bialystok & Hakuta, 1999; Bialystok & Miller, 1999; Birdsong & Molis, 2001; Hakuta et al., 2003; Johnson & Newport, 1989), the conclusion that there are AoA-related changes in L2 attainment for learners before and after puberty is not consistently supported by the results of behavioural studies. Moreover, evidence from late L2 learners who display native-like performance (e.g., Birdsong, 1992; Bongaerts, 1999; White & Genesee, 1996) suggests that successful L2 acquisition may not be impossible for all L2 learners who begin L2 acquisition later in life and beyond the proposed critical/sensitive period. Moreover, some recent ERP studies investigating the neuro-cognitive bases of L2 morpho-syntactic processing have failed to demonstrate qualitative differences between native speakers and late L2 learners when the learners' level of L2 proficiency is controlled. These results suggest that late L2 learners may not necessarily use "fundamentally different" neuro-cognitive mechanisms to process L2 morpho-syntax. At the very least, general maturation accounts of L2 acquisition that predict large-scale differences in the way the brain processes a language learned during childhood or adulthood, are not supported. However, many questions remain. For example, Steinhauer et al.'s (2012) finding of native-like neuro-cognitive profiles in late L2 learners when processing salient grammatical violations that operate in a similar way in the L1 and L2, leaves open the question of whether age effects in L2 acquisition could reflect more nuanced differences in how the brain processes early- and late- acquired languages. For example, do late L2 learners show native-like neuro-cognitive processing profiles for grammatical and phonological aspects of their L2 that are different from their L1 and, thus, cannot be directly transferred?

The Present Studies

The ERP studies that comprise this dissertation attempt to extend our understanding of the neuro-cognitive bases of L2 processing for late L2 learners by addressing a number of outstanding questions: (1) How do late L2 learners process L2 morpho-syntactic structures that are not present in the L1 or that are expressed differently in the L1 and L2? (2) How do the neuro-cognitive processes underlying L2 morpho-syntactic processing change as late L2 learners become more proficient in their L2? (3) To what extent do late L2 learners use similar neuro-cognitive mechanisms to native speakers for processing L2 phonological contrasts that are not used in their L1? (4) Under which stimulus and task conditions can late L2 learners engage in native-like phonological processing? Questions 1 and 2 were the focus of Study 1 and questions 3 and 4 were the focus of Study 2.

Study 1 was a longitudinal ERP study that investigated the neuro-cognitive processes used by Chinese- and Korean- late L2 learners of English before and after participating in an intermediate level intensive English-as-a-second-language course. Specifically, it examined whether these late L2 learners, like native speakers, would exhibit a P600 in response to L2 morpho-syntactic structures that are not used in their L1 or that are expressed differently in the L1 and L2. The results of previous studies suggested that late L2 learners' ability to recruit the neuro-cognitive mechanisms underlying the P600 may be restricted to L2 morpho-syntactic structures that are expressed in a similar way in the L1 and L2 (Chen, Shu, Liu, Zhao, & Li, 2007; Ojima Nakata & Kakigi, 2005; Sabourin & Stowe, 2008). These previous studies support a maturational account of AoA effects in L2 acquisition because they argued that the consequence of L2 learning after puberty is the inability to recruit the same neuro-cognitive processes as native speakers. Study 1 examined the extent to which these processes become available as late L2 learners advance from relatively low to intermediate levels of L2 proficiency. The results of this study showed that the neuro-cognitive mechanisms underlying the P600 are not restricted by the grammatical inventory of the learners' L1 once late L2 learners have attained intermediate levels of L2 proficiency. The results are discussed further with respect to how increasing levels of L2 proficiency with respect to specific grammatical structures are linked to the recruitment of some of the neuro-cognitive processes used by native speakers.

Study 2 was a cross-sectional ERP study that investigated the extent to which high and low proficiency late L2 learners of English, with French as a native language, engage in similar neuro-cognitive processes as native speakers when processing L2 phonology. Specifically, it examined whether high proficiency late L2 learners of English engage in both attention-driven and automatic processing of a non-native phonetic contrast (i.e., speech sounds that are not used to contrast meaning in the L1) that is notoriously difficult for native French speakers to acquire (/h/versus /Ø/). Similar to the results reported by

Steinhauer et al. (2012) for L2 morpho-syntactic processing, the findings of Study 2 suggest that native-like, automatic aspects of L2 phonetic processing become increasingly available as late L2 learners advance in proficiency. This study is unique insofar as it is the first study that we are aware of to systematically evaluate how L2 proficiency influences the neuro-cognitive basis of L2 phonetic processing in late L2 learners in different task and stimulus conditions. It is also theoretically important insofar as L2 phonological attainment is thought to be one domain of language that is particularly affected by maturational factors (Lamendella, 1977; Long, 1990; Scovel, 1989). The results from Study 2 suggest that L2 phonetic discrimination involves the same neuro-cognitive processes in native speakers and high proficiency late L2 learners. Moreover, native-like L2 processing may be possible even in difficult task/stimulus conditions that require attention to semantic rather than phonological information, which previous research has found to be particularly difficult, if not impossible for late L2 learners (Strange, 2011; Strange & Shafter, 2008). This suggests that even automatic aspects of L2 phonological processes may occur via the same neuro-cognitive processes in native speakers and high proficiency late L2 learners.

Taken together, the results from Studies 1 and 2 have important implications for understanding AoA effects in L2 learners and, in particular, whether maturation of the brain is the cause of AoA effects that have been reported in previous behavioural studies. By studying the extent to which late L2 learners use similar neuro-cognitive processes as native speakers and how these processes change as learners become more proficient in their L2, we can examine whether AoA sets limits on the way the brain learns and represents a L2. These implications are discussed in detail at the end of each study and in the General Discussion section.

Study 1

Brain Responses Reveal Proficiency-Based Changes in Second Language Grammar Processing: An ERP Investigation of L1-L2 Differences in Late Second Language Learners[2]

Erin Jacquelyn White, Fred Genesee & Karsten Steinhauer

---

[2] Acknowledgements: This work was supported by grants awarded to Fred Genesee from the Social Sciences and Humanities Research Council of Canada (SSHRC) and les Fonds de Recherche Société et Culture Québec (FQRSC), to Karsten Steinhauer from the Canada Research Chair program (CRC), the Canada Foundation for Innovation (CFI) and Sciences and Engineering Research Council of Canada (NSERC), to Erin White from NSERC and to all authors from the Center for Research on Brain, Language and Music (CRBLM).

Abstract

This longitudinal study tracked the neuro-cognitive changes associated with second language (L2) grammar learning in adults in order to investigate how L2 processing is shaped by a learner's first language (L1) background and L2 proficiency. Previous studies have argued that late L2 learners cannot elicit a P600 in response to L2 grammatical structures that do not exist in the L1 or that are different in the L1 and L2. We tested whether the neuro-cognitive processes underlying this component become available after intensive L2 instruction. Korean- and Chinese late-L2-learners of English were tested at the beginning and end of a 9-week intensive English-L2 course. Event-related potentials (ERPs) were recorded while participants read English sentences containing violations of regular past tense (a grammatical structure that operates differently in Korean and does not exist in Chinese). Whereas no P600 effects were present at the start of instruction, by the end of instruction, significant P600s were observed for both L1 groups. Latency differences in the P600 exhibited by Chinese and Korean speakers may be attributed to differences in L1-L2 reading strategies. Across all participants, larger P600 effects at session 2 were associated with higher levels of behavioural performance on an online grammaticality judgment task. These findings suggest that the neuro-cognitive processes underlying the P600 (e.g., “grammaticalization”) are modulated by individual levels of L2 behavioural performance and learning.

Brain Responses Reveal Proficiency Based Changes in Second Language Processing: An ERP Investigation of L1-L2 Differences in Late Second Language Learners

When adults begin learning a second language (L2), they typically start with an already-established first language (L1) system. Depending on the similarity between the two languages, transferring knowledge from the L1 can provide a useful basis to begin communicating in the L2, be it through shared phonological, lexical-semantic, or grammatical forms. It has been argued that anything that can transfer from the L1, will, and that this can, in some cases, assist learning (MacWhinney, 2005). However, L1 transfer can also be problematic if the L1 and L2 systems do not map exactly onto one another and this can lead to difficulties acquiring some aspects of the L2. What is unclear, and highly debated, is the extent to which a learner's L1 continues to influence L2 acquisition and processing as he/she advances in L2 proficiency. Some researchers claim that late (i.e., post-puberty) L2 learners can *only* acquire grammatical structures that are present in the L1 (e.g., Franceschina, 2005), while others argue that these structures can be acquired, albeit more slowly than structures that are also instantiated instigated in the L1 (White, 2003). From a neuro-cognitive perspective, it has been argued that L2 acquisition in late learners is influenced by the neural networks that underpin L1 processing (Hernandez & Li, 2007; MacWhinney, 2005). However, it is unclear whether the L1 continues to influence (and potentially restrict) the neuro-cognitive mechanisms used for L2 processing as learners advance in proficiency. Using neuro-cognitive measures to longitudinally track the impact of learners' L1 on L2 grammar processing is an important step towards understanding the neuro-cognitive changes that are associated with late L2 acquisition and the extent to which processing is influenced by the L1 (Kotz, 2009).

In the present study, we report results from a 9-week longitudinal study that investigated the neuro-cognitive changes that are associated with L2 acquisition in adults participating in an intensive English-as-a-second-language course. This research sought to elucidate how learners' L1 influences the neuro-cognitive mechanisms that underlie L2 grammar processing at progressive stages of L2 proficiency. It also sought to examine how individual differences in L2 behavioural performance are associated with different profiles of L2 neuro-cognitive processing and plasticity. Specifically, we investigated: (1) to what extent L1 background influences the neuro-cognitive basis of L2 grammar processing; (2) how L2 processing changes with L2 learning; and (3) the relationship between behavioural measures of L2 grammatical performance and L2 neuro-cognitive processing.

Many previous studies investigating age of acquisition effects on the neural bases of L2 processing have compared native speakers and L2 learners using cross-sectional designs to examine the extent to which factors such as age of L2 acquisition, L1 background, and L2 proficiency constrain L2 processing. Using event-related potentials (ERPs), this research has demonstrated that the neural basis of L2 grammar processing may be particularly sensitive to the interplay between these factors, especially at lower levels of L2 proficiency (e.g., Steinhauer, White & Genesee, 2012). However, the relative role of each factor is unclear. In particular, it is unclear whether late L2 learners who have attained relatively high levels of L2 proficiency can engage the same neuro-cognitive processes as native speakers for processing grammatical structures that are not used in the L1 (Chen, Shu, Liu, Zhao, & Li, 2007; Ojima Nakata & Kakigi, 2005) or are expressed differently in the L1 and L2 (Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005).

In native speakers, grammar processing is reliably associated with the P600 ERP component. The P600 is a positive-going wave that is typically maximal at central-parietal electrodes approximately 600 ms after the onset of the critical word in a sentence (Hagoort, Brown & Groothusen, 1993; Osterhout & Holcomb, 1992). The P600 has been interpreted as an index of structural reanalysis (i.e., a controlled and attention-driven process occurring during a relatively late stage in sentence processing; Friederici, 1995; 2002), sentence repair (Hagoort et al., 1993; Osterhout & Holcomb, 1992), integration difficulty (Kaan, Harris, Gibson & Holcomb, 2000), or continued sentential analysis elicited by a mismatch between multiple levels of representation (Kuperberg, 2007). In many studies of L1 morpho-syntactic processing, the P600 is preceded by a "left anterior negativity" (LAN) – a negative-going wave that is often maximal at left anterior electrodes between 300-500 ms after stimulus onset[3]. The LAN has been

---

[3] Some studies of L1 phrase structure processing also report early LAN (ELAN) effects elicited 100-300 ms post-stimulus onset over left anterior electrode sites. Friederici (1995; 2002) proposed that ELAN and LAN effects reflect distinct stages of sentence processing, with the ELAN reflecting the parser's failure to assign a phrase structural representation at very early stages of processing and the LAN as reflecting processing difficulty with other classes of morpho-syntactic information (e.g., subject-verb agreement). Conversely, Steinhauer and Drury (2012) have argued that ELAN effects may be an artifact of stimulus design (e.g., unbalanced baseline intervals

linked to highly automatic rule-based parsing, thought to occur during early stages of morpho-syntactic analysis (Hahne & Friederici, 1999). A biphasic LAN/P600 response has been observed in response to many classes of grammatical violations, including phrase structure (e.g., Neville, Nicol, Barss, Forster & Garrett, 1991) and inflectional morphology (e.g., Gunter, Stowe & Mulder, 1997), although many other studies have reported P600 effects without a LAN (e.g., Osterhout & Mobley, 1995; Kuperberg et al., 2003, Sabourin & Stowe, 2008). In L2 learners, the presence of a P600 in response to L2 grammatical violations has been taken as evidence that they have "grammaticalized" the particular structure under investigation; that is, that they have incorporated the relevant rule-based grammatical knowledge into their online L2 processing system and engage in the same neuro-cognitive processes as native speakers when presented with a violation (Osterhout et al., 2008). A LAN effect in L2 speakers has been taken as an indication that they can access and apply this knowledge automatically (Dowens, Vergara, Barber & Carreiras, 2010; Steinhauer, White & Drury, 2009; Steinhauer et al., 2012). LAN effects have also been associated with implicit, as opposed to explicit, learning experiences (Morgan-Short, Steinhauer, Sanz & Ullman, 2012). In contrast, the absence of these components in L2 learners has been used to suggest that processing at least certain kinds of late-acquired L2 grammatical structures may not involve the same neuro-cognitive mechanisms that underlie grammatical processing in native speakers. In particular, it has been argued that late L2 learners may be unable to exhibit native-like P600 responses

---

between correct and violation sentences). Thus, the functional significance of ELAN effects is unclear.

when presented with L2 grammatical structures that are expressed differently in the L1 and L2 or that are not present in the L1 at all (Chen et al., 2007; Ojima et al., 2005; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005). It is thought that these L2 grammatical structures will not be salient enough to trigger the neuro-cognitive processes that are reflected by the P600 in native speakers (e.g., sentence reanalysis/repair as proposed by Hahne & Friederici, 1999).

For example, Ojima et al. (2005) compared the processing of English subject-verb agreement violations in native English speakers and late L2 learners of English who were Japanese native speakers. Because Japanese does not use grammatical morphology to encode number or person, native Japanese speakers cannot draw on relevant L1 grammatical knowledge when processing these structures in English. In response to violations, the native English speakers exhibited the aforementioned pattern of a LAN followed by a P600. This biphasic response was not observed in the L2 learners. Those with low levels of L2 proficiency did not exhibit any ERP responses, suggesting that they either did not recruit additional brain resources to process the violations or that their processing strategies varied too much to elicit a consistent ERP profile. In contrast, high proficiency L2 learners exhibited a left-lateralized negativity between 350-550 ms, as did the native speakers (i.e., a LAN); however, they did not display a P600. A lack of a P600 has also been observed in response to subject-verb agreement violations in high proficiency Chinese learners of English; Chinese grammar also does not use morphology to express number or person (Chen et al., 2007). Together, these studies suggest that late L2 learners may be unable to exhibit a

P600 in response to violations of L2 grammatical structures that they cannot transfer from their L1.

The complete absence of a P600 in these studies is striking. Previous research has documented a P600 in late L2 learners with low levels of L2 proficiency in response to morpho-syntactic violations when the structures under investigation are similar in the L1 (McLaughlin et al., 2010; Osterhout et al., 2008; Steinhauer et al., 2012; Tanner, Osterhout & Herschensohn, 2009). The L2 learners tested by Ojima et al. (2005) and Chen et al. (2007), however, had high levels of English proficiency (as determined by their scores on standardized tests of English proficiency) and, overall, they performed with high accuracy on a grammaticality judgment task that was administered either concurrently with ERP testing or directly following it. Moreover, the presence of a native-like LAN without a P600 in high proficiency L2 learners is surprising given the LAN is thought to index the recruitment of automatic morpho-syntactic processing mechanisms (Hahne & Friederici, 1999) and, thus, should be acquired at a later stage of L2 acquisition than the controlled processes reflected by the P600 (Hahne, 2001). To our knowledge, Ojima et al. (2005) is the only study to report a LAN in the absence of a P600 during grammar processing in either L1 or L2 speakers.

Ojima et al. (2005) suggest that the absence of a P600 in late L2 learners is a “true qualitative difference from native language processing” and that the cognitive processes reflected by the P600 “cannot be triggered by syntactic features acquired after a critical period” (p. 1223). Alternatively, the L2 learners in that study may have, in fact, displayed a P600, but it was out of the time range investigated. In both the Chen and Ojima studies (see also Weber-Fox & Neville, 1996; Tokowicz & MacWhinney, 2005), the L2 learners' ERP waveforms were not analysed after 1000 ms post-stimulus. It could very well be that late L2 learners who speak languages with different morpho-syntactic constraints are able to engage in the sentence reanalysis processes that are reflected by the P600, but are slower to initiate these processes and exhibit their effects only after 1000 ms. Indeed, P600s with peak latencies at around 1000 ms and later have been observed in previous studies of L2 grammar processing in low/intermediate proficiency L2 learners (e.g., Hahne, 2001; Rossi, Gugler, Friederici & Hahne, 2006).

P600s might also be delayed when L2 learners are required to read the experimental sentences, particularly if the L1 and L2 use different writing systems that require different strategies for efficient word reading, as was the case for the Chinese and Japanese participants in the Ojima and Chen studies (for a discussion of the neural basis of reading in different languages, see Perfetti, Nelson, Liu, Fiez & Tan, 2010). Indeed, Steinhauer et al. (2012) found that both high and low proficiency Chinese L2 learners of English exhibited a delayed P600 when reading English phrase structure violations compared to native English speakers. In contrast, native French speakers, even at rather low levels of English (L2) proficiency, exhibited a P600 with a similar onset and peak latency as found in native English speakers. Thus, for late L2 learners, the latency of the P600 may reflect an interaction between L1 grammatical knowledge, L1 reading experiences, and L2 proficiency level. Testing for late occurring ERP responses (i.e., after 1000 ms) may help clarify whether L2 learners fail to exhibit this component when presented with L2 grammatical features that are not used in their L1 or if they are merely slower to elicit it.

Others have argued that in order to exhibit a P600, the L2 grammatical feature under investigation must not only be present in the L1, but must operate in a similar way in the two languages. Sabourin and Stowe (2008) tested the processing of determiner-noun gender agreement in native Dutch speakers and two groups of Dutch L2 learners: those whose L1 was German or a Romance language (French, Italian or Spanish). While the concept of grammatical gender exists in the L1 of all participants, its expression is similar in Dutch and German and different in the Romance languages. Unlike German speakers, Romance speakers need to learn the gender of all Dutch nouns on a word-by-word basis and cannot transfer specific and surface level grammatical processing strategies from their L1. The ERP results seem to suggest that such transfer may be necessary for native-like sentence processing. Whereas the native Dutch speakers and the German-L2-learners-of-Dutch displayed similar P600 responses, no P600 was observed in the Romance speakers (even though the ERPs were analysed until 1500 ms). As highlighted by the authors, this lack of a P600 cannot be easily attributed to general L2 proficiency levels because both the Romance and German speakers displayed a native-like P600 in response to violations of past perfect tense, which operates in a similar manner in all three languages. Sabourin and Stowe (2008) concluded that, for late L2 learners, native-like recruitment of the mechanisms underlying the P600 may be limited to processing grammatical structures that are not only present in the L1 and L2, but expressed in a similar way in the two languages.

However, inspection of Sabourin and Stowe's Romance speakers' behavioural performance, compared to that of the Dutch and German native speakers, suggests an alternative interpretation. Performance on the grammaticality judgement task conducted concurrently with ERP testing was significantly higher for the native speakers and the German group than for the Romance group (who performed near chance level) in the gender agreement condition. These results contrast with those from the past tense condition where which all groups performed with high accuracy and exhibited significant P600s. The fact that the groups' behavioural performance was significantly different in the very condition in which the languages also differ raises the possibility that the ERP results may not reflect L1 background alone, but also proficiency in the target structure. Indeed, the Romance speakers were also significantly worse than the German speakers on an offline task that required participants to identify the gender of the nouns that were used in the ERP study. This is important because knowing the gender of a noun is critical for identifying a violation of gender agreement (Sabourin, Stowe & de Haan, 2006) and recognizing a grammatical violation as such is necessary to elicit the P600 (Osterhout & Mobley, 1995). Perhaps the Romance speakers, despite their otherwise high levels of general L2 proficiency, had not attained sufficient knowledge of the Dutch grammatical gender system specifically in order to engage the sentence reanalysis processes that are reflected by the P600. Indeed, significant P600 effects have been reported in response to gender agreement violations during L2 processing in native English speakers who have no L1 experience with grammatical gender whatsoever after participants received intensive artificial language training (Morgan-Short, Sanz, Steinhauer & Ullman; 2010) and in native English speakers who are highly proficient in Spanish (Dowens et al., 2009).

Viewed from this perspective, the lack of the P600 in the Romance speakers reported by Sabourin and Stowe (2008) may simply reflect what the L2 learners had not yet acquired rather than what they were incapable of acquiring. When a grammatical structure does not exist in an L2 learner's L1 (or operates differently in their L1 and L2), it may take longer to acquire compared to structures that are similar in both languages. At low levels of proficiency, L2 learners may fail to notice that it is obligatory to use the particular grammatical structure in certain cases (Ellis, 1994) and, as a result, they will not use the same neuro-cognitive mechanisms to process it as native speakers. However, this does not preclude the possibility that L2 learning can continue to more proficient levels and that native-like neuro-cognitive processing can become realized once higher levels of proficiency have been achieved (Steinhauer et al., 2009). We do not yet have a clear understanding of how L1 knowledge and developing L2 knowledge interact at different stages of acquisition to shape L2 processing (Kotz, 2009). As highlighted by Li and Green (2007, p. 119), the field is in need of "longitudinal research into the adaptive changes triggered in response to the acquisition of a new language". Rather than inferring developmental patterns by comparing different groups of learners who have attained high or low levels of proficiency, following a single group of learners as they acquire their L2 allows us to actually track this development directly. Moreover, examining learners' competence/proficiency with respect to specific grammatical structures, rather than globally, along with the neuro-cognitive mechanisms they use to process those structures could reveal to what extent it is important to assess L2 learners' competence/proficiency in more specific ways (Steinhauer et al., 2009).

Another important issue that is only beginning to be addressed is the extent to which individual differences in L2 proficiency and grammatical performance are associated with differences in neuro-cognitive processing profiles. A number of studies by Osterhout and colleagues (McLaughlin et al., 2010; Osterhout, McLaughlin, Kim, Greenwald & Inoue, 2004; Osterhout, McLaughlin, Pitkaenen, Frenck-Mestre, & Molinaro, 2006; Osterhout et al., 2008; Tanner et al., 2009; Tanner, McLaughlin, Herschensohn & Osterhout, 2012) demonstrate that the ERP waveforms of a group of L2 learners might not be representative of the neuro-cognitive processes available to subsets of learners who have attained either high or low levels of structure-specific proficiency. For example, Tanner et al., (2009; 2012) found that the amplitude of the P600 elicited in response to subject-verb agreement violations in English-learners of German correlated positively with their performance on an online grammaticality judgement task. This shows that learners who perform well behaviourally are more likely to recruit native-like processing strategies (or recruit them to a greater degree) than learners who perform poorly. In previous studies of L2 grammar processing (e.g., Sabourin & Stowe, 2008), native-like effects may have been elicited in a subset of participants although they were masked by the use of group ERP data. This is important because null effects in a group of L2 learners have often been taken as evidence that native-like processing is unavailable to all L2 learners (e.g., Ojima et al. 2005). Investigating the relationship between individual differences in behavioural performance with respect to specific L2 structures and the neuro-cognitive mechanisms used to process those structures may have important consequences for our understanding of whether it is possible for at least some late L2 learners to use native-like processing mechanisms.

The Present Study

The present study had three goals: (1) to investigate whether late L2 learners can exhibit a P600 in response to violations of grammatical structures that are either absent in their L1 or that are expressed differently in their L1 and L2; (2) to track how the neuro-cognitive basis of L2 grammar processing changes as a result of participating in an intensive L2 course; and (3) to investigate the relationship between behavioural measures of L2 grammatical performance and L2 neuro-cognitive processing. To address these questions, ERPs were recorded in late L2 learners at the beginning and end of an intensive 9-week English-as-a-second course. At each session, the learners read English sentences that were correct or that contained a violation of past tense regular verbs (**Table 1; Appendix 1**). Studying learners longitudinally allowed us to track any neuro-cognitive changes that might be associated with the acquisition of these structures and how L1 background and L2 grammatical proficiency influence L2 processing.

Moreover, by studying the same learners at progressive stages of proficiency we decreased some of the individual variability that is inherent to cross-sectional (between-subject designs) because each participant is compared to his/her own performance rather than to another individual.

The L2 learners were native Mandarin Chinese and Korean speakers, allowing us to examine how late learners process L2 grammatical structures that are not present in their L1 (Chinese) or that operate differently in their L1 (Korean). In contrast to English, Chinese does not use inflectional morphology to express tense, person, or number. Thus, Chinese learners of English can rely on little L1 transfer to process English past tense; rather, the grammatical knowledge they can use to process these structures reflects what they have acquired in the L2 as adults. Korean speakers, on the other hand, can rely on some form of L1 transfer, although the situations in which they can apply their knowledge of inflectional morphology for processing our particular stimuli are different in their L1 and L2. Korean expresses simple past tense through verbal morphology (as does English); however the distinction between simple past and past perfect that was used in the present experiment does not exist in Korean (e.g., the difference between *she did not start* vs. *she had not started*). As in English, Korean can express simple past tense with negation by inflecting an auxiliary verb rather than the main verb (e.g., *she did not start* literally translates into *she start did not*); however Korean can also express the same idea by inflecting the main verb (e.g., *he no started* is also acceptable in Korean). Thus, Korean L2 learners of English need to learn that to express the simple past with negation in English, they must inflect the auxiliary verb but not the main verb (e.g., *did not start* vs. *did not *started*), whereas to express past perfect they must learn to inflect both the auxiliary and the main verbs (e.g., *had not started*). See **Table 1** and **Appendix 1** for example sentences. In other words, although Korean speakers have knowledge of inflectional morphology from their L1 to process English past tense, they need to learn when to apply this knowledge in order to accurately process the stimuli used in the current experiment.

Korean- and Chinese-L2 learners of English also differ in the nature of their L1 reading experiences, which may influence the latency of ERP effects elicited during L2 sentence reading. Like English, Korean is an alphabetic language that uses letters to encode phonemes that are assembled to form syllables and words. In both languages, word reading is thought to occur in a similar way (Perfetti, Liu & Tan, 2005). Chinese, in contrast, is a logographic or morphosyllabic system - written characters correspond to spoken syllables, which in many cases are whole words. As a result of these writing system differences, Chinese speakers are thought to rely relatively more on orthographic processing and less on pre-lexical phonological processing during L1 reading than native English speakers (Perfetti et al., 2005). Importantly, behavioural evidence suggests that when reading in their L2, Chinese L2 learners of English are slower and less accurate than Koreans who are matched in English proficiency, particularly when they are required to differentiate between words that look alike (Wang, Koda & Perfetti, 2003). Thus, it is possible that word identification will take longer in the Chinese- compared to the Korean-speakers and that this may be reflected in P600 effects with delayed latencies. In order to observe effects that might occur with a delayed latency, we examine ERP responses until 1500 ms post-stimulus, rather than 1000 ms, as in some previous studies.

Insert Table 1 about here

A second issue explored in this study is the extent to which differences in L2 grammatical proficiency is associated with the use of different neuro-cognitive processing mechanisms. Following Tanner et al., (2009; 2012), we correlated behavioural measures of grammatical sensitivity (i.e., the ability to differentiate well-formed and violation sentences) and P600 effects at both sessions. This extends the work of Osterhout and colleagues by investigating whether the relationship between individual differences in performance and neuro-cognitive processing that has been reported for the acquisition of grammatical structures that are similar in the L1 and L2 also holds for the acquisition of L2 grammatical structures that are either not present or are dissimilar from those in the L1 (see McLaughlin et al., 2010, for a discussion).

The sentence structures used here have been found to elicit a LAN and a P600 in native English speakers (Drury, Steinhauer, Pancheva & Ullman, 2006; Drury, Steinhauer & Ullman, 2012). Based on previous work with low/intermediate proficiency L2 learners (e.g., Hahne, 2001; Ojima et al., 2005; Steinhauer et al., 2012), we did not expect the Chinese or Korean participants to exhibit a LAN. As noted earlier, the LAN is thought to reflect implicit rule-based processing that is automatically triggered in response to a violation of morpho-syntax and, thus, is usually associated with near-native levels of L2 proficiency. Thus, it is likely that these processes will become available only after years of L2 exposure, and not after 9 weeks of instruction (Steinhauer et al., 2009). Therefore, our focus of interest was on whether the L2 learners in the present study would exhibit proficiency-related changes in the P600 component, which would suggest the "grammaticalization" of L2 morpho-syntax (Osterhout et al., 2008) and the recruitment of sentence reanalysis processes that are used by native speakers during morpho-syntactic processing (Hahne & Friederici, 1999).

Different theoretical frameworks would make different predictions as to whether the Chinese or Korean participants would exhibit P600 effects. Following the claim that L2 learners cannot exhibit P600s in response to L2 grammatical structures that are not instantiated in the L1 (Chen et al, 2007; Ojima et al., 2005), we expected no P600 for the Chinese speakers; although P600s may be observed for the Korean speakers, as they could rely on at least some L1 transfer. In contrast, Tokowicz and MacWhinney (2005) have argued that native-like processing is unavailable for L2 grammatical structures that are *different* from those in the L1, but may be possible for structures that are absent from the L1. This is because when the L1 and L2 provide conflicting interpretations of a given grammatical structure, the stronger L1 interpretation will prevail. This on-line competition between the two languages is thought to continue to influence L2 processing even at higher levels of L2 proficiency. Thus, these authors would predict no P600 for the Koreans, whereas the Chinese speakers may exhibit P600s by the end of the L2 course. Finally, Sabourin and Stowe (2008) propose that L2 learners will exhibit a P600 only when they can transfer surface-level similarities between their two languages. In this case, we would expect to see no P600 for either group at either testing session.

Alternatively, if (as proposed in Steinhauer et al., 2009) it is learners' L2 proficiency level that is an important predictor of neuro-cognitive processing patterns, then we would expect to see P600s at session 2 for both groups, if the learners succeed at "grammaticalizing" the target structures (i.e., incorporate the relevant grammatical knowledge into their online language processing system; Osterhout et al., 2008). Moreover, P600 amplitudes should correspond to behavioural performance, as measured by grammatical sensitivity. If the Korean and Chinese speakers display P600 effects after intensive L2 instruction, it would provide evidence against the notion that the L1 grammatical system continues to limit L2 neuro-cognitive processing once intermediate levels of L2 proficiency have been attained. By examining ERP responses as a function of L1 background and L2 grammatical performance both before and after participating in an intensive L2 course, we were able to investigate how learners' L1-background and their level of L2 proficiency modulated learning-induced changes in L2 processing at early and later stages of proficiency.

Methods

Participants

Thirty-two late L2 learners of English participated in this study. Sixteen spoke Korean as an L1 (20-28 years old, M = 22.6, 13 female) and 16 spoke

Mandarin Chinese as an L1 (18-38 years old, M =23.9, 7 female). There was no age difference between the Korean and Chinese participants at the time of testing [**Table 2**, *t*(30) = 0.77, *p* >.10]. An additional 9 participants (5 Korean) were tested but excluded from the analyses because of excessive movement, eye-blink or alpha artifacts contaminating the EEG signal (in at least one of the sentence conditions during one of the testing sessions), and 3 were excluded because they did not return for the second testing session. Participants gave written informed consent, were paid for their participation, had normal or corrected-to-normal vision, and reported no history of hearing, language, speech or neurological disorders. All were right-handed (assessed using self-report and the Edinburgh handedness inventory; Oldfield, 1971) and reported comparable educational backgrounds (i.e., most were currently undergraduate students or had recently graduated). Two cohorts of participants were recruited over two consecutive summer language programs in order to increase the sample size. Recruitment of additional participants was not possible because it ran the risk of introducing confounds due to significant changes in the course itself (e.g., course materials, content, instructors etc.).

Insert Table 2 about here

All participants were foreign students living temporarily in Canada for the purpose of studying English in an intensive 9-week English-as-a-second language course at McGill University. They were enrolled in an intermediate level class (as determined by the school's placement test). At the first testing session, we administered a cloze test of English proficiency that has been used previously as a general indicator of L2 proficiency in previous studies (e.g., Goad & White, 2008). The test consisted of a one-page passage with approximately every seventh word missing, 30 in total. They were required to read the text and fill in the missing words by selecting a word from among 4 multiple-choice options. Both groups performed around chance level on the test, indicating similar low levels of general L2 proficiency at the start of the study and there was no significant difference between the Korean (45.7%) and Chinese (51.9%) participants [$t(30) = 0.98$, $p > .10$]. No significant difference was observed between the Korean (67.9 %) and Chinese (69.1%) in their final marks in the English course at the end of the study either [$t(28) = 0.36$, $p > .10$], indicating similar levels of L2 proficiency at the second testing session as well (**Table 2**).

At each session, participants self-rated their abilities in English on 6 dimensions (listening, reading, pronunciation, fluency, vocabulary and grammar) using a 7-point scale (1= no proficiency at all, 7 = like a native speaker; **Table 2**). Potential L1 or session differences were analyzed with a repeated measures ANOVA with the 6 dimensions and session as within-subjects variables and L1 as a between-subjects variable. Overall, participants rated their English abilities higher at session 2 ($M = 4.2$) than session 1 [$M = 3.9$; Sess: $F(1,30) = 9.07$, $p \leq .005$]. No significant main effect or interactions with L1 were observed, indicating that the Korean and Chinese participants perceived their own English abilities as similar.

The participants completed language background questionnaires that provided information about their previous and current English experiences. Previous English exposure was assessed by asking participants to report how much English they used at home and at school (as a percentage of total language use) between the ages of 0-4, 5-11, 12-14, 15-16, 17-18 and 19+. Neither the Koreans nor the Chinese reported substantial exposure to English before the age of 12. Thus, according to Birdsong (2006), both groups can be classified as late L2 learners. To test whether the Korean and Chinese groups differed in lifetime (and in particular childhood) English exposure, a repeated measures ANOVA was run using L1 group as a between-subjects variable; age (0-4, 5-11, 12-14, 15-16, 17-18, 19+) and location (home, school) were within-subjects variables. The Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to analyses involving the age factor (as it involves more than one degree of freedom). This revealed no significant main effect or interaction with the factor L1 group, indicating a similar amount of early English exposure for the Korean and Chinese groups. As seen in **Table 2**, both groups reported limited English use throughout their lives, particularly as children.

Current English exposure was assessed at both testing sessions by asking participants to report their current use of English and their L1 (as a percentage of their total daily language use within the week of the testing session; **Table 2**). This was analyzed using a repeated-measures ANOVA with session as a within-subjects variable and L1 as a between-subjects variable. This revealed no main effect or interaction involving L1 or session (ps > .10), indicating similar English use by both groups at both sessions.

Stimuli

At each session, participants read 72 experimental sentences (36 correct, 36 containing a violation) that tested their processing of the inflection rules governing the past tense of regular verbs in English. The stimuli were simple active-voice sentences consisting of 5-9 high frequency words. They were based on stimuli used in previous studies with English native speakers (Drury et al., 2006; Drury et al., 2012), with the vocabulary adapted for low proficiency L2 learners of English. See **Table 1** for examples of the sentences (asterisks mark violations, critical verbs are underlined). These sentences were randomized among 152 filler sentences containing other types of morpho-syntactic anomalies (subject-verb agreement and phrase structure), which will be described in another paper.

Sentences were designed to avoid ERP artifacts that can arise when the critical word and preceding baseline interval differ between the correct and violation sentences. Thus, in both conditions, 4 versions of each test sentence were created to ensure a balanced experimental design: the correct and violation contrast involved the identical verb form and preceding sentence context (see Drury et al., 2012; Steinhauer & Drury, 2012 for more discussion of stimulus design issues). Half of the sentences were grammatically correct and half contained a violation of English past tense (simple past or past perfect) involving a regular verb. The correct versus violation contrasts were created by manipulating the pre-target auxiliary verb, allowing us to compare ERP responses to correct and violation sentences involving identical verb forms. Half of the critical verbs used bare stem forms (e.g., *didn't start* and **hadn't start*) and half used *–ed* suffixed participles (e.g., *hadn't started* and **didn't started*); half were preceded by the auxiliary *do* and half by the auxiliary *have.* All of the items were negated since negation was needed to license *do.* In order to vary the position of the critical verb in the sentence, half of the items contained the contracted form of the auxiliary and negation (didn't/hadn't) and half contained full forms (did not/had not); in half the subject was a pronoun (he/she), in half it was a lexical noun phrase (e.g., the customer).

Participants were presented with different lists of sentences at each testing session. To create the lists, we first developed 72 sentences (each containing a different critical verb). Four versions of each sentence were then created according to the manipulations described above (see **Table 1**) and were evenly assigned to the four presentation lists (1A, 1B, 2A, 2B). No verb was repeated in a given list. Participants saw different forms of the critical verbs at each testing session; if a given verb was presented with inflection at the first session (e.g., *hadn't started*), it was presented without inflection at the second session (e.g., *didn't start*). The sentences were also counter-balanced across A and B lists so that a given verb form was presented in a correct sentence in one list and as a violation in the other (e.g., *didn't start* vs. * *didn't started*). Half of the participants were presented with a "1" list at the first session and a "2" list at the second session (e.g., 1A and 2A) and, vice versa, for the remaining participants.

As a result of this procedure, when the ERPs were averaged across participants, the same critical word and preceding context appeared in both correct and violation sentences. This design ensures that ERP effects are a result of the violation per se and not confounded with lexical differences between the critical words or the contexts preceding the target word (see Steinhauer & Drury, 2012, for a discussion of baseline problems in many other studies).

Procedure

Participants were tested twice: once after the first week of the intensive English course and then during the last two weeks of the course. At each testing session, they were seated comfortably in a sound-attenuated room, approximately 70 cm in front of a computer screen that displayed the stimuli. They were given specific instructions in English (both verbal and written) about the task and were asked not to blink or move while the stimuli were being presented. They were instructed to read each sentence carefully and to judge it for grammatical correctness by pressing one of two mouse buttons in response to a visual prompt at the end of each trial. The experiment began with the presentation of 8 practice sentences, followed by a short break in which they could ask questions. Each test trial began with the presentation of a fixation cross (500 ms) in the centre of the screen followed by sentences that were presented word-by-word in the centre of the screen (300 ms per word at an inter-stimulus interval of 200 ms). The response prompt (“*good?”)* was presented 1000 ms after offset of the last word and remained on the screen until the participants responded with a button press or 5 seconds had elapsed. After a subsequent ‘eye blinking’ interval of 1500 ms, the next trial began. Prior to the ERP session, the participants completed the language-background questionnaires. Each testing session lasted for 2.5-3.0 hours, including short breaks.

ERP Recordings and Analysis

Continuous EEG was recorded from 19 cap-mounted tin electrodes according to the international 10-20 system and digitized online at 500 Hz. The recordings were referenced to the left ear lobe and re-referenced off-line to averaged left-/right-mastoids. Eye movements were monitored by additional electrodes placed at the outer canthus of each eye (EOGH) and above and below the left eye (EOGV). Electrode impedances were kept below 5 kΩ. For approximately half of the participants, we used Compumedics/NeuroScan NuAmps amplifiers to amplify the EEG and EOG signals at the first session, whereas for the remaining participants and sessions we used Compumedics/NeuroScan SynAmps2 amplifiers. As we found no difference in the data obtained from the two amplification systems, we collapsed all data together for subsequent analyses.

Offline, the EEG was filtered with a phase-true 0.3-30 Hz band-pass filter using the EEProbe software package (Advanced Neuro-Technology, ANT; Enschede, the Netherlands). Data were screened for eye movements, muscle, and other noise artifacts. Participants were included in further analyses if they contributed a minimum of 20 artifact-free trials for the correct and violation sentences at each session. On average, participants contributed 75 % artifact-free trials. A repeated measures ANOVA using Session (1 or 2) and Condition (correct or violation sentences) as within-subjects variables and L1 as a between-subjects variable revealed that each Session, Condition and L1 group contributed the same number of trials (*p*s >.10). After pre-processing the data, artifact-free ERP responses were averaged for each participant for each condition (i.e., correct and violation sentences) and testing session. This was done for a 1600 ms interval, time-locked to the onset of the critical verb, including a 100 ms pre-stimulus baseline interval.

Single-subject ERP averages can be based on trials that correspond to correct behavioural responses only (i.e., "response contingent" analyses) or on "all trials" irrespective of behavioural accuracy. Each approach has advantages and disadvantages. For example, it is possible that participants exhibit larger P600 effects in response to violation sentences that they deem unacceptable (i.e., correct rejections) compared to violation sentences that they perceive as correct (i.e., misses; Osterhout & Mobley, 1995). Let us assume this is the case and assume that, for the present study, behavioural performance may also improve between sessions. Under this scenario, it would be unclear from the analysis of "all trials" whether any change in ERP effects reflects a *quantitative* change in neuro-cognitive processing due to a larger proportion of trials with correct responses at session 2 versus session 1 or *qualitative* change in how trials corresponding to correct responses were processed (i.e., P600 effects emerged at session 2 that were not present at session 1 whatsoever). By analyzing the ERP effects corresponding to "correct only trials", we can distinguish between these possibilities. If significant P600 effects at session 2 are observed that are not present at session 1 when only correctly-answered trials are entered into the analyses, then it would suggest a true qualitative change in processing and the recruitment of neuro-cognitive processes that were not available to the L2 learners at session 1. In this respect, the analysis of “correct only trials” is advantageous.

On the other hand, significant changes in ERP components have been observed in L2 participants even before corresponding changes in behavioural measures of language processing occurred, suggesting that ERP measures may be more sensitive to learning progress than behavioral measures (McLaughlin, Osterhout & Kim, 2004; Tremblay, Kraus & McGee, 1998). Consequently, trials that correspond to incorrect behavioural responses may nevertheless elicit ERP effects in L2 learners. By discarding ERP trails based on behavioural responses, we may lose valuable information about the neuro-cognitive changes that co-occur with increasing L2 proficiency. Response-contingent analyses might also result in the exclusion of participants with an inadequate number of correctly answered trials at both testing sessions – an issue that is particularly problematic for longitudinal research with low proficiency participants. In this respect, analysis of “all trials” (i.e., trials corresponding to both correct and incorrect behavioural responses) would be advantageous.

Most ERP studies of L2 grammar processing have conducted either response-contingent analyses (e.g., Dowens et al., 2010; Pakulak & Neville, 2011) or analyses of all trials (Ojima et al., 2005; Tanner et al., 2012; Tokowicz & MacWhinney, 2005; Weber-Fox & Neville, 1996). However, in the present study, we conducted analyses of ERP data corresponding to all artifact-free trials as well as ERP data corresponding to correctly-answered trials only (response contingent analyses). In this way, we hoped to better understand whether any change we observe between sessions reflects a quantitative or qualitative difference in neuro-cognitive processing. Moreover, by comparing the ERP effects elicited by "correct trials only" and "all trials" (i.e., trials corresponding to both correct and incorrect responses) we can infer whether L2 learners engaged different neuro-cognitive processes for sentences that they responded to correctly or incorrectly.

For the response contingent analyses, participants were included in the analysis if they contributed at least 12 correctly-answered artifact-free trials for each condition (i.e., 1/3 of total sentences). This resulted in the exclusion of 11 participants (6 Koreans). The 21 remaining participants who were included in these analyses contributed, on average, 22 trials for each condition and session. For both sets of analyses, the mean amplitude of ERP waves was analyzed within two time windows (early: 500-700 ms and late: 750-950) based on previous studies of L2 P600 effects (e.g., Weber-Fox & Neville, 1996) and visual inspection of the grand averages for each L1 group.

Repeated-measures ANOVAs were performed for each time window on 12 lateral (F3, F4, C3, C4, P3, P4, F7, F8, T3, T4, T5, T6) and 3 midline (Fz, Cz, Pz) electrodes. For the lateral electrodes, L1 (Chinese, Korean) was a between-subjects factor and the within-subjects factors were: Condition (correct or violation), Session (1 or 2), Hemisphere (Left or Right), Anterior-Posterior (anterior, central, parietal), and Laterality (lateral-lateral, medial). For the midline sites, the factors were: L1 (Chinese, Korean), Condition (correction or violation), Session (1 or 2) and Anterior-Posterior (anterior, central, parietal). Results are reported for main effects and interactions that involve at least one condition factor. The results of the midline analyses are reported only when they yielded results that were not revealed in the analyses of the lateral electrodes. The Greenhouse-Geisser correction was applied to all analyses involving the AP factor (as it involves more than one degree of freedom) and corrected *p* values are reported.

Analysis of Behavioural Data

Following Tanner et al. (2012) and Morgan-Short et al. (2012), behavioural results (i.e., grammaticality judgments obtained in the EEG experiment) were quantified using d-prime scores (Macmillan & Creelman, 2005). D-prime scores provide an unbiased measure of grammatical sensitivity – participants' ability to discriminate the correct and violation sentences. Scores were calculated based on performance on the grammaticality judgment task for each participant at each session using the following formula: $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$ These scores were analyzed using a repeated measures ANOVA with Session (1, 2) as a within-subjects factor and L1 (Korean, Chinese) as a between-subjects factor.

Results

Behavioural Results

Mean grammatical sensitivity (d-prime) scores for the L1 groups at each session are presented in **Table 3**. Overall, grammatical sensitivity improved substantially from session 1 to session 2 [Sess: $F(1,30) = 20.04$, $p < .001$]. No effects or interactions with L1 background were observed (*p*s > .10), indicating similar performance by the Chinese and Korean participants at both sessions and a similar improvement throughout the duration of the course.

Insert Table 3 about here

ERP Results: Analysis by L1 Groups (All Trials)

Grand average ERP waveforms for the Korean and Chinese participants at each session are presented in **Figure 1** and **2,** respectively, and topographical maps are shown in **Figure 3**. Both groups of participants exhibited a positivity at session 2 in response to the tense violations that was not present at session 1. This positivity occurred earlier for the Koreans than the Chinese. Results from the global ANOVA are presented in **Table 4**.

Insert Figures 1, 2 and 3 about here

Insert Table 4 about here

Analysis of the ERP data elicited between 500-700 ms revealed a significant Sess x Con x Lat interaction ($p \leq .01$), pointing to a significant change in ERP effects between sessions. However, significant Sess x Con x Lat x L1 and Sess x Con x Hem x L1 interactions (*ps* < .05) suggest a different pattern of effects for the two L1 groups. Indeed, the Korean participants displayed a Sess x Con interaction that approached significance [$F(1, 15) = 2.51$, $p = .051$] as well as significant Sess x Con x Lat [$F(1, 30) = 12.29$, $p < .005$] and Sess x Con x Hem [$F(1, 30) = 4.49$, $p < .05$] interactions. While no significant effect of Con was observed at session 1, the Koreans exhibited a highly significant P600 at session 2 [Con: $F(1, 15) = 24.48$, $p < .001$], as seen in **Figures 1** and **3**. Significant Con x Lat [$F(1, 30) = 10.23$, $p < .01$], Con x Hemi [$F(1, 30) = 6.51$, $p < .05$], and Con x Lat x Hem [$F(1, 30) = 5.91$, $p < .05$] interactions revealed that this positivity was largest at medial right [$F(1, 15) = 25.09$, $p < .001$], medial left [$F(1, 15) = 23.88$, $p < .001$], and lateral left sites [$F(1, 15) = 19.17$, $p \leq .001$]. Similarly, at midline electrodes, a significant positivity was observed at session 2 [$F(1, 15) = 16.24$, $p \leq . 001$] that was not present at session 1 [*p*s > .10], resulting in a significant Sess x Con interaction [$F(1, 15) = 4.71$, $p < .05$]. For the Chinese participants, in contrast, no significant ERP effects or change in effects between sessions was observed in this time window (*p*s > .10).

Between 750-950 ms, a number of significant interactions involving the factors Con, Lat, Hem and AP were observed (see **Table 4**). Most importantly, a Sess x Con x Lat x Hem x AP x L1 interaction ($p < .05$), again, suggests different patterns of ERP effects for the two L1 groups at each session. Analyses of each group separately revealed a significant Sess x Con interaction for the Chinese participants [$F(1, 15) = 7.13$, $p < .05$], indicating a change in ERP effects between sessions in this later time window. Specifically, a significant positivity was observed at session 2 [Con: $F(1, 15) = 10.22$, $p < .01$] that was not seen at session 1[4]. Significant Con x Lat [$F(1, 15) = 8.97$, $p < .01$] and Con x Lat x AP [$F(1, 15)$

---

[4] At session 1 a significant Con x Lat x Hem interaction was observed for the Chinese participants [$F(1, 15) = 5.24$, $p < .05$], however it did not lead to a significant main effect of Con.

= 3.61, $p < .05$] interactions in session 2 revealed that this positivity was largest at medial central [$F(1, 15) = 11.59$, $p < .005$], medial posterior [$F(1, 15) = 13.22$, $p < .005$] and lateral posterior [$F(1, 15) = 12.55$, $p < .005$] electrodes, consistent with P600 effects reported in previous studies. Similarly, at midline electrodes, a significant positivity was observed at session 2 [$F(1, 15) = 15.77$, $p \leq .001$] that was not present at session 1 [*p*s > .10; Sess x Con: $F(1, 15) =13.07$, $p < .005$].

For the Korean participants, two interactions approached significance in this time window: Sess x Con x Lat x Hem [$F(1, 15) = 3.84$, $p = .069$] and Sess x Con x Lat x Hem x AP [$F(1, 15) = 3.44$, $p = .051$]. Analysis of each session revealed no significant effects at session 1 and significant Con x Hem [$F(1, 15) = 6.57$, $p < .05$] and Con x Lat x Hem [$F(1, 15) = 6.81$, $p < .05$] interactions at session 2. However, unlike the effect exhibited by this group in the earlier time window, these interactions point to only a marginally significant positivity at lateral left electrodes [$F(1, 15) = 3.8$, $p = .07$].

ERP Results: Analysis by L1 Groups (Correct Only Trials)

The main results from the analysis of correctly-answered trials are summarized in **Table 5.** Overall, these data mirror the findings reported for the analyses across all trials. Again, we confirmed the emergence of a significant positivity at session 2 for both the Korean and Chinese participants that was not present at session 1. As in the analysis of all trials, this positivity began later in the Chinese than in the Korean participants, although in response-contingent analyses the Koreans' positivity extended into the later time window.

Insert Table 5 about here

Between 500-700 ms, significant Sess x Con x L1 and Sess x Con x Hemi x L1 interactions again point to L1 group differences (**Table 5**). The Korean participants showed significant Sess x Con [$F(1, 10) = 5.54, p < .05$], Sess x Con x Lat [$F(1, 10) = 10.21, p \leq .01$], Sess x Con x Hem [$F(1, 10) = 7.43, p < .05$] and Sess x Con x Lat x Hem x AP [$F(1, 10) = 4.53, p < .05$] interactions, reflecting the emergence of an early P600 in session 2 that was absent in session 1 (all *p*s involving Con > .10). The positivity in session 2 [Con: $F(1, 10) = 8.53, p < .05$] was larger over medial sites [$F(1, 10) = 12.09, p < .01$; Con x Lat: $F(1, 10) = 12.17, p < .01$] and the left hemisphere [$F(1, 10) = 12.15, p < .01$; Con x Hem: $F(1, 10) = 3.26, p = .10$]. As in the analysis of all trials, no significant effects were observed for the Chinese participants in this time window.

Between 750-950 ms, the ANOVA for response-contingent ERP data revealed the following pattern: a main effect of Con ($p < .05$) was qualified by significant Sess x Con ($p < .01$) and Sess x Con x Hem x L1 ($p < .01$) interactions (see **Table 5**). As in the analyses of all trials, these interactions reflect a late P600 in session 2 for the Chinese participants [Con: $F(1,9) = 11.03, p < .01$] that was not present at session 1 [*p*s > .10; Sess x Con: $F(1, 9) = 16.79, p < .005$]. The positivity at session 2 was largest at medial [$F(1,9) =$, $p < .01$; Con x Lat: $F(1, 9) = 8.65, p < .05$] and posterior [$F(1,9) = 24.37, p \leq .001$; Con x AP: $F(2, 8) = 2.59, p < .05$] electrodes of both hemispheres. For the Korean participants, we observed a pattern that differed from what analyses across all trials had suggested. That is, we found evidence that – for correct trials only – the P600 in session 2 extended into this late time window. First, there was significant main effect of condition at session 2 [$F(1,10) = 6.11$, $p < .05$]. Second, and consistent with the P600 scalp distribution in the earlier (500-700 ms) time window reported above, the late part of this positivity was also maximal over medial sites [$F(1,10) = 6.78$, $p < .05$; Con x Lat: $F(1,10) = 5.82$, $p < .05$] and the left hemisphere [$F(1,10) = 9.84$, $p < .05$; Con x Hem: $F(1,10) = 3.74$, $p = .082$].

In summary, analysis of "correct trials only" largely confirmed the emergence of P600s in session 2 observed when "all trials" were analyzed. In addition, it revealed a prolonged P600 for the Koreans. Importantly, no P600 effects were observed at session 1 for either group. This suggests that the change in ERP effects observed between sessions reflects a qualitative change in processing and the emergence of neuro-cognitive processes that were not present at the beginning of the L2 course, rather than a quantitative change in the proportion of trials that participants engaged these processes.

ERP Results: Comparison of "All Trials" vs. "Correct Only Trials"

Both the analysis of all trials and the response contingent analyses revealed significant P600 effects at session 2 that were not present at session 1. This is compatible with two distinct underlying patterns: (1) similar P600 effects were elicited in response to trials corresponding to correct and incorrect behavioural responses (i.e., hits and misses). In other words, although ERPs already consistently distinguished between correct sentences and sentences containing tense violations, subsequent behavioural judgments were less consistent and reflected this discrimination only in a subset of trials (behavioural performance for a given trial was not related to P600 amplitudes). Alternatively, (2) P600 amplitudes may have been larger when participants correctly categorized sentences containing tense violations as ungrammatical. If this is the case, then it would suggest a close relationship between P600 amplitudes and subsequent behavioural performance. By directly comparing the analysis of "all trials" and "correct trials only," we can infer whether trials corresponding to correct and incorrect behavioural responses elicited similar ERP effects.

To compare the two analyses directly, we conducted a repeated measures ANOVA on the mean amplitude of the P600 difference wave (i.e., response to violation sentences minus correct sentences) using Analysis Type (all trials, correct only), Sess (1, 2), and time window (500-700, 750-950) as within-subjects variables and L1 background as a between-subjects variable. The 21 participants who were included in the response contingent analyses were included here. Unsurprisingly, this analysis again found the expected Sess effect [$F(1,19) = 8.11$, $p \leq .01$], reflecting larger P600 effects at session 2 than session 1. Importantly, it also revealed a significant Sess x Analysis Type interaction [$F(1,19) = 4.89$, $p < .05$] and, at session 2, a highly significant main effect of Analysis Type [$F(1,19) = 11.82$, $p < .005$], pointing to a larger P600 effect for the analysis of correctly answered trials ($M = 1.73$ uV) compared to the analysis of all trials ($M = 1.34$ uV). No difference between Analysis Type was observed at session 1($F < 1$). This suggests that the L2 learners exhibited larger and/or less variable P600 effects for trials they responded to correctly, compared to incorrect trials. **Table 6** displays the mean amplitude of the P600 effects for each analysis type, session, time window, and group.

Insert Table 6 about here

Relationship between ERP Results and Behavioural Performance

In the analyses of the behavioural data, we found that measures of grammatical sensitivity (d-prime scores) were significantly higher after the intensive English-as-a-second-language course compared to those at the start of the course for both the Korean and Chinese speakers. ERP measures of neuro-cognitive processing were also different at session 1 and 2, with both groups displaying significant P600 effects at session 2 that were not present at session 1. Taken together, this suggests that higher levels of behavioural performance were associated with P600 effects at session 2. However, previous studies have shown that late L2 learners demonstrate considerable variability, both in terms of their behavioural performance and ERP effects (e.g., Tanner et al., 2012). Therefore, an important question is whether an individual's ability to discriminate grammatical from ungrammatical sentences is directly reflected by ERP measures.

To investigate the relationship between individual learners' behavioural performance and P600 effects, we ran bivariate correlations between the participants' grammatical sensitivity (d-prime) scores at session 2 and the mean amplitude of session 2 P600 difference waves (violation minus correct) for "all trials" at a representative electrode (Pz) between 500-950 ms. This revealed a significant positive correlation between session 2 P600 effects and grammatical sensitivity at session 2 ($r = .378$, $p < .05$), indicating that participants with higher d-prime scores exhibited larger P600 effects[5]. This provides further evidence for a link between higher levels of behavioural performance and larger P600 effects. This also suggests that the session 2 P600 effects reported earlier for the Korean and Chinese participants were likely driven by the participants with the highest behavioural performance. The relationship between behavioural performance and P600 amplitude is presented as a scatter plot in **Figure 4**.

Insert Figure 4 about here

Discussion

Four main findings emerged from this study. First, P600s effects that were absent at the start of the L2 course emerged after intensive L2 exposure for both Korean- and Chinese- L2 learners of English, and thus regardless of L1-L2 differences. Second, generally speaking, the amplitude of the P600 effect at session 2 was largest in the L2 learners who displayed the highest levels of performance during the on-line grammaticality judgement task. These findings argue that it is a learner's level of L2 proficiency and, in particular, proficiency

---

[5] Correlations were also run to compare the behavioural performance and the amplitude of P600 effects based on correctly-answered trials only (i.e., response contingent P600 effects). As expected, the correlation no longer reached significance ($r = .302$, $p = .091$), because P600 effects were less variable across participants and because of reduced power resulting from excluding participants.

with respect to the particular grammatical structures under investigation, that determines access to the P600's underlying neuro-cognitive processes, rather than the grammatical structures in his/her L1. Third, a significant change in ERP effects was observed in the analysis “all trials” and “correct trials only,” which is evidence for a qualitative change in the neuro-cognitive processes used by the L2 learners at the start and end of the L2 course. Finally, although all participants exhibited a P600, its onset latency was later in the Chinese compared to the Korean speakers. In what follows, we discuss how L2 proficiency influences the presence and magnitude of the P600 and how, in some cases, L1 background may influence its latency.

The presence and magnitude of the P600 reflects L2 proficiency

Previous ERP studies of L2 grammar processing in late L2 learners have argued that a learner’s L1 background will determine which neuro-cognitive mechanisms are available for L2 processing (e.g., Chen et al., 2007; Ojima et al., 2005; Sabroutin & Stowe, 2008; Tokowicz & MacWhinney, 2005). In particular, these studies concluded that late L2 learners will not exhibit a P600 when processing L2 morpho-syntactic structures that are expressed differently in the L1 and L2 or that are not present in the L1 at all. In contrast, the results of the present study showed that P600s can, in fact, be elicited when processing L2 grammatical features that are acquired later in life and that do not exist in the L1 (Chinese speakers) or operate differently in the L1 and L2 (Korean speakers). Moreover, P600 effects were observed after (but not before) the L2 learners participated in intensive intermediate-level L2 instruction, suggesting that the neuro-cognitive processes that are thought to underpin the P600 in native speakers become available to late L2 learners as their L2 proficiency improves. Applying Osterhout et al.'s (2008) interpretation of P600 effects in L2 speakers, this means that by session 2 both the Chinese and Korean speakers had “grammaticalized” the morpho-syntactic rules that differentiated the correct and violation sentences in the current experiment and had incorporated these rules into their on-line language processing system. Thus, the ability to engage these processes does not appear to be limited to L2 grammatical structures that are similar to those in the L1, and appears to become available to L2 learners at intermediate levels of L2 proficiency.

The emergence of P600 effects after intensive L2 instruction corresponded with overall improvement in behavioural measures of grammatical sensitivity (as measured by d-prime scores on the grammaticality judgement task), suggesting that the processes underlying the P600 became available to L2 learners as their L2 proficiency with respect to the specific grammatical structures tested (i.e., regular past tense verbal morphology) increased. Two lines of evidence support this claim. First, there was a significant correlation between the P600 amplitude and d-prime scores at session 2, demonstrating that individuals who displayed larger P600 effects were those who displayed higher levels of behavioural performance. This corroborates the findings of Tanner et al., (2009, 2012) who reported significant correlations between P600 amplitude and grammatical sensitivity in their English-speaking learners of German-L2 when processing a grammatical structure (subject-verb agreement) that is present in the L1 and L2. Our results show that the same relationship between individual differences in P600 amplitude and L2 grammatical sensitivity holds for processing L2 grammatical structures that cannot be directly transferred from the L1.

Second, in an additional exploratory analysis, we grouped participants according to their grammaticality sensitivity scores at session 2, to examine whether sub-groups of participants who displayed relatively high levels of behavioural performance after the intensive course showed a larger change in ERP effects between sessions and a larger P600 at session 2, compared to those with relatively low levels of behavioural performance. To do this, we used a median split of session 2 d-prime scores for the Chinese and Korean participants separately to create 4 groups, each with 8 participants: high Chinese, low Chinese, high Korean, and low Korean. A repeated measures ANOVA was then conducted on ERP effects exhibited within the 500-950 ms time window data at midline electrodes[6].

The results are consistent with the idea that higher levels of behavioural performance at session 2 and learning (as measured by improvements in behavioural performance from session 1 to 2) is associated with the development of the P600. The "high" performance groups showed a significant change in ERP effects between sessions [$F(1, 14) = 11.01, p \leq .005$], and a highly significant P600 at session 2 [$F(1, 14) = 10.96, p \leq .005$], whereas the "low" groups showed neither significant change nor session 2 P600 effects (all ps > .10). Importantly, this mirrors their behavioural performance as well. The "high" groups showed a

---

[6] Analyses were conducted on all trials (regardless of behavioural responses) as the majority of participants who were excluded from the analyses of correct-only trials (due to an insufficient number of artifact-free trials with correct responses) were "low" performers.

significant improvement in their d-prime scores [$t(15)= 5.32$, $p < .001$] whereas the “low” groups did not [$t(15)= 1.65$, $p > .10$]. Together, this is evidence for a direct link between neuro-cognitive processing and the behavioural performance that is associated with this processing. The difference between the mean amplitude of P600 effects exhibited by the “high” and “low” groups at each session can be seen in the bar graphs in **Figure 5**.

Insert Figure 5 about here

The results of this study underscore how indices of proficiency with respect to the specific structures used in the ERP experiment (i.e., d-prime scores) may be more appropriate indicators of the neuro-cognitive mechanisms used during L2 grammar processing than general measures of L2 proficiency, as has been used in some previous studies (see also Steinhauer et al., 2009, for a discussion). Despite the fact that all participants were enrolled in an intermediate level English-as-a-L2 course, differences in structure-specific L2 proficiency were found over time and these corresponded to different profiles of ERP effects. Combining L2 learners who display varying levels of structure-specific L2 proficiency, despite otherwise similar levels of general L2 proficiency, may have contributed to the absence of P600 effects in some previous studies (e.g., Sabourin & Stowe, 2008). In line with the findings of Osterhout and colleagues (McLaughlin et al., 2010; Osterhout et al., 2004; 2006; Tanner et al., 2009; 2012), our results indicate that, although a group of L2 learners may appear to be homogeneously proficient overall, substantial differences can exist among individuals with respect to their proficiency in particular grammatical structures and, moreover, the neuro-cognitive mechanisms they use to process those structures. Future research will benefit from exploring this source of individual variation further.

An important question we also sought to address is whether the emergence of P600 effects at session 2 reflected a qualitative or quantitative change in neuro-cognitive processing between sessions. Had we only conducted analysis of “all trials” (i.e., correct and incorrectly answered trials combined) we would not have been able to tease these possibilities apart. It could have been argued that P600 effects may have been exhibited at both sessions in response to correctly answered trials but that these effects did not reach significance at session 1 because they were observed by the relatively large number of trials with incorrect responses. We would not have been able to rule out the possibility that the L2 learners engaged similar neuro-cognitive processes at the start and end of the intensive L2 course, but differed in how consistently they used them – a quantitative change in neuro-cognitive processing. However, this explanation cannot account for the results of the “response contingent analyses”, in which no significant P600 effects were observed at session 1, even when correctly answered trials were analyzed on their own. P600s were, however, observed at session 2. This suggests, instead, that the emergence of the P600 at session 2 reflects a qualitative change in processing and the recruitment of neuro-cognitive mechanisms that were not available to the L2 learners only 9 weeks previously.

This corroborates the findings of artificial language learning experiments that report the emergence of P600 effects when processing grammatical structures that are not present in participants' L1 after relatively short training periods (Morgan-Short et al., 2010; 2012).

By directly comparing the two types of analyses, we could also speculate as to whether the L2 learners engaged different neuro-cognitive processes for sentences that they responded to correctly or incorrectly. Previous studies have reported significant changes in the ERP components exhibited by L2 learners even before corresponding changes in behavioural measures of language processing occurred, suggesting that ERP measures are more sensitive to learning progress than behavioral measures (McLaughlin et al., 2004; Tremblay et al., 1998). Thus, it is possible that trials with both correct and incorrect behavioural responses elicited similar P600 effects in the present study as well, a conclusion that could be drawn if similar P600 effects were observed for the analysis of "all trials" (i.e., correct and incorrectly answered trials combined) and the analysis of "correct trials only" (i.e., response contingent analyses)[7]. However, this comparison revealed *smaller* P600 effects for the analysis of "all trials" compared to "correct trials only". Additionally, for the Korean participants, the duration of the P600 was longer for trials they responded to correctly. We can infer from this that trials that corresponded to incorrect behavioural responses either did not elicit

---

[7] Not enough participants had a sufficient number of artifact-free trials with incorrect behavioural responses to analyze these trials on their own. However, the ERP effects that corresponded to trials with incorrect behavioural responses can be inferred by comparing the effects for "all trials" (i.e., correct and incorrect trials combined) and "correct trials only".

P600 effects or the effects were smaller and less consistent than those elicited by correctly answered trials. This corroborates the results of Osterhout and Mobley (1995) who found that sentences containing gender agreement violations involving personal pronouns elicited P600 effects only in a set-set of native speaker participants who deemed the sentences to be unacceptable. This reinforces the tight link between behavioural performance and P600 effects and suggests that P600s may be a marker of morpho-syntactic processing only for trials and in participants who detect the violations as such.

L1 reading strategies may influence the latency of the P600

Although both the Chinese and Korean speakers exhibited significant P600 effects at session 2, for the Chinese speakers this effect started approximately 250 ms later than for the Korean speakers. In studies of L2 grammar processing, delayed P600s have often been attributed to low levels of L2 proficiency (e.g., Hahne, Mueller & Clahsen, 2006; Steinhauer et al., 2009). For example, Rossi et al. (2005) reported P600 effects that began approximately 300 ms later for low- compared to high-proficiency L2 learners in all three sentence conditions tested (subject-verb agreement, word category, and combinations of both types of violations). In the present study, however, differences in L2 proficiency is unlikely to account for the latency differences observed between the Chinese and Korean speakers because both groups displayed similar levels of behavioural performance on the grammaticality judgement task (as measured by d-prime scores), rated their English abilities as similar, and received comparable marks at the end of their intensive English-as-a-L2 course.

It is possible that the latency difference could reflect differences between the Chinese and Korean speakers in terms of their L1 grammatical knowledge, independent of L2 proficiency. Under this account, the Chinese speakers may have been slower to initiate the processes underlying the P600 because they could not transfer any relevant grammatical knowledge from their L1, whereas the Korean speakers might have been faster because they could transfer at least some L1 knowledge of verbal inflection to aid L2 sentence processing (even though the specific nature of their knowledge was different). However, this explanation has difficulty accounting for the results of Dowens et al. (2009) who compared the processing of Spanish number and gender agreement violations in native English speakers who were highly proficient late L2 learners of Spanish. P600s with similar onset latencies were observed in response to violations of both number (a grammatical structure used in both English and Spanish) and gender agreement (a grammatical structure that is not used in English), although the amplitude of the P600 was smaller in response to the gender violations[8]. Thus, it is unclear whether L1-L2 grammatical similarities can account for the delayed latency of the P600 observed here by the Chinese speakers.

Another possibility is that late L2 learners transfer reading strategies from their L1, rather than grammatical knowledge. Slower and less accurate word reading by the Chinese speakers, compared to the Koreans, may have delayed

---

[8] Both violation types also elicited LAN effects in the L2 speakers in this study, although the LAN began approximately 50 ms earlier in response to the number- compared to the gender- agreement condition. This may be evidence that L1-L2 similarities in terms of grammatical rules influenced the timing of this effect. However, accuracy in the grammaticality judgement task was also lower for this condition compared to for the number agreement condition. Thus, it is difficult to tease apart whether the latency difference reflects proficiency or L1-L2 similarity differences.

subsequent morpho-syntactic analysis and led to a later P600. Behavioural studies show that Chinese speakers are less accurate English word readers than Koreans who are matched in their level of English proficiency, particularly when differentiating between words that look alike (Wang et al., 2003). This difficulty is thought to reflect differences in how Chinese speakers read words in their L1 and L2, compared to both Korean and English native speakers. Word reading is thought to occur in a similar way in English and Korean because they are both alphabetic languages – the phonological and orthographic representations of words are activated in a cascade style and feed forward to activate the meaning of the word being read (see Perfetti et al., 2005). Chinese, in contrast, has a logographic or morphosyllabic system; written characters correspond to spoken syllables and morphemes which, in many cases, are whole words (Perfetti et al., 2005). This means that orthographic processing plays a central role when reading Chinese and that it must reach a certain threshold before the corresponding phonological and semantic representations of the word can be activated (Perfetti & Tan, 1998).

Neural imaging studies report that the network of brain areas that are activated during Chinese word reading include areas that are not consistently activated by native English speakers when reading English (Tan, 2005). These areas include the left middle frontal gyrus (lMFG), which Perfetti *et al.,* (2010) have suggested may be involved in maintaining the orthographic form of the character in working memory while the phonological and semantic representations of the word are retrieved and integrated. Interestingly, these same areas are activated when Chinese learners of English read in English, suggesting that they use the brain network they developed for reading their L1 when reading their L2 (Tan et al., 2003). In other words, Chinese speakers appear to rely heavily on orthographic processing during word reading in both Chinese and English, and this may make them slower and less accurate than Korean speakers at reading individual English words and, in particular, differentiating between words that look alike.

In short, when presented with critical words in the current experiment, that differ by only two letters (e.g., *didn't start vs. didn't *started*), the Chinese speakers may have been slower at differentiating between correct and violation sentences because of the way they read, rather than the way they process the sentence's grammaticality. Gouvea, Phillips, Kazanina and Poeppel (2010) have suggested that when processing morpho-syntactic violations, the onset latency of the P600 reflects the time needed to recognize and retrieve the incoming verb, access the relevant features of the noun phrase from working memory, detect a mismatch, and begin sentence reanalysis. Thus, if the system is slowed down at the word reading level, morpho-syntactic processing may also be delayed. For the Chinese speakers, reading English words as they would read Chinese characters is a less efficient strategy that may have resulted in a P600 with a delayed onset. For the Koreans, in contrast, this would not have been a problem because they are accustomed to using both phonological and orthographic information during word identification. As a result, the Koreans exhibited a P600 whose latency was similar to that observed previously in English native speakers (Drury et al., 2012;

Drury et al., 2006). Thus, the delayed latency of the P600 may reflect of the transfer of L1 reading strategies rather than grammatical features.

This explanation may also account for the lack of a P600 before 1000 ms in previous reading studies of morpho-syntactic processing by Chinese L2 learners of English (Chen et al., 2007; Steinhauer et al., 2006; Weber-Fox & Neville, 1996). It may also explain why a P600 was not observed prior to 1000 ms in the high proficiency Japanese L2 learners of English reported by Ojima et al. (2005), as Japanese Kana is a syllabic writing system that also relies heavily on orthographic processes during word reading (Perfetti et al., 2005). Rather than exhibiting no P600 whatsoever, the L2 learners in these studies may have displayed a delayed P600 that was not evident in the 1000 ms time window that was used to analyze the waveforms. If it is indeed the transfer of L1 reading strategies that influenced the latency of the P600, rather than L1 grammatical knowledge, this would suggest that Chinese learners of English may display P600s with earlier onset latencies if they are required to listen to experimental sentences rather than read them. Comparing the latency of the P600 in Chinese speakers as they either read or listen to English sentences containing morpho-syntactic violations would elucidate this issue further.

It is important to highlight that, viewed from this perspective, differences in latency of the P600 is a not a “fundamental difference” between groups. In the current study, both the Korean and Chinese L2 learners of English displayed a P600 at the second testing session that differed primarily with respect to *when* it was exhibited. This demonstrates that after intensive L2 instruction, both groups were able to apply the same neuro-cognitive mechanisms during L2 morpho-syntactic processing that are thought to be reflected by the P600 in native English speakers. Moreover, it is possible that the difference between the Chinese and Koreans in terms of the onset latency of their P600 effects would disappear if the Chinese speakers participated in intensive instruction that focused specifically on improving their English word reading skills (e.g., by learning how to segment the phonological information contained in printed English words). If such a change in the P600's latency was found, it would further underscore how deployment of processes underlying the P600 are not restricted to grammatical rules acquired early in life, but rather can be applied to new structures, even in adult learners.

Future Research Directions

Before concluding, we would like to highlight three more potential avenues for future research. First, behavioural studies indicate that L2 learners are more likely to make errors of omission in production (i.e. omitting required inflections) than errors of commission (i.e. adding inflections where none is called for; Jia & Fuse, 2007). Thus, processing of these types of errors may also engage different neuro-cognitive processes, especially at early stages of L2 proficiency. In the present study, these violation types were combined in order to avoid ERP artefacts that arise from using different word forms in the correct and violation conditions. Tracking how processing of these types of violations change as L2 proficiency increases would provide a better understanding of how knowledge of L2 verbal inflection develops during L2 acquisition in late learners.

Second, a number of studies by Osterhout and colleagues (Osterhout et al., 2004; 2006; McLaughlin et al., 2010) suggest that L2 learners may pass through an intermediate stage of neuro-cognitive processing during which they exhibit N400 effects in response to morpho-syntactic violations, before they advance in proficiency and display P600s as found in native speakers. The N400 stage has been interpreted to mean that L2 learners may initially memorize inflected words as unanalyzed "chunks" rather than decomposing them into stem and affix sequences. However, the relationship between theses processes and behavioural measures of L2 performance are unclear. For example, in their study of L2 morpho-syntactic processing, Tanner et al. (2012) reported a significant correlation between grammatical sensitivity performance and P600 amplitudes, but no correlation was found between behavioural performance and N400 amplitudes, suggesting the neuro-cognitive processes underlying the N400 effect may not be as closely linked to behavioural performance as P600 effects. However, these findings have been based primarily on analyses of ERP effects corresponding to "all trials", irrespective of behavioural responses. Thus, it is unknown to what extent N400s are exhibited when processing sentences that participants respond to correctly or incorrectly. To this end, directly compare the ERP effects elicited in response to "all trials" versus "correct trials only" (as in the present study) may help reveal the relationship between different kinds of neuro-cognitive processes and behavioural performance.

Finally, one of the goals of the present study was to examine the relationship between behavioural measures of L2 performance and underlying neuro-cognitive processing, as indexed by P600 effects. The results of the analyses comparing “all trials” and “correct only trials” suggested larger and more consistent P600s were elicited for trials that participants responded to correctly compared to incorrect trials. However, a limitation of the current study, as with most ERP studies of L2 morpho-syntactic processing, is that performance was assessed with a two-choice grammaticality judgement task; participants were required to respond even if they were not sure of a sentence’s grammaticality. Thus, it is likely that a portion of correct responses occurred as a result of chance, rather than the learners’ L2 grammatical knowledge. One possibility for future studies would be to assess performance with a three-choice grammaticality judgement task, in which participants could categorize each sentence as “grammatical”, “ungrammatical”, or “I don’t know”. Such an approach would decrease the likelihood that correctly-answered trials were due to chance and, therefore, provide a more sensitive measure of the relationship between behavioural performance and ERP effects. It is also possible that ERP effects might be observed for sentences that corresponded to “I don’t know” responses. If so, this would indicate that participants could, on some level, distinguish the sentence’s grammaticality although they were not confident in their subsequent behavioural judgements. Comparing ERPs as a function of these three response types might provide sensitive information about the neuro-cognitive processes that L2 learners engage as they advance in proficiency.

Conclusion

The results of this study reveal the capacity for qualitative neural changes that can co-occur with L2 learning in adults, even when processing L2 morpho-syntactic structures that either operate differently in the L1 and L2 or that are not present in the L1 at all and, thus, cannot be directly transferred from their L1. Moreover, the finding that P600s were: (1) observed after, but not before, participating in an intensive L2 course; (2) largest in L2 learners who displayed the highest levels of grammatical sensitivity; and (3) larger in response to trials that corresponded to correct behavioural responses, strongly suggests that L2 proficiency plays a critical role in determining access to the P600's underlying neuro-cognitive processes, rather than L1-L2 similarities.

References

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*, 9-49.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*(2), 161-174.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2009). Morphosyntactic processing in late second language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870-1887.

Drury, J. E., Steinhauer, K. & Ullman, M. T. (2012). Left anterior negativities, verbal inflectional morphology and (ir)regularity: An ERP study. Unpublished manuscript. Department of Linguistics, Stony Brook University, Stony Brook, New York

Drury, J. E., K. Steinhauer, R. Pancheva, & M. T. Ullman (2006). The definiteness effect: An ERP study. Journal of Cognitive Neuroscience Supplement, 225.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Franceschina, F. (2005). Fossilized second language grammars: The acquisition of grammatical gender. Amsterdam: John Benjamins.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language, 50*, 259-281.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing.*Trends in Cognitive Sciences 6*, 78–84.

Goad, H. & L. White. (2008). Prosodic structure and the representation of L2 functional morphology: a nativist approach. *Lingua 118*: 577-594.

Gouvea, A. C., Phillips, C., Kazanina, N. & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*(2), 149-188.

Greenhouse, S. W., & Geisser, S. (1959). On the methods in the analysis of profile data. *Pryschometrika, 24*(2), 95-112.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34,* 660-676.

Hagoort, P., Brown, C. & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes 8*, 439–83.

Hahne, A. 2001. What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistics research 30*, 251–66.

Hahne, A. & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience 11*, 194–205.

Hahne, A., Mueller, J., & Clahsen, H. (2006). Morphological processing in a second language: behavioural and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*(1), 121-134.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638-650.

Jia, G. & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: age related differences. *Journal of Speech, Langauge and Hearing Research, 50*, 1280-1299.

Kaan, E., Harris, A. Gibson, E. & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes 15*, 159–201.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language, 109*, 86-74.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research, 1146*, 23-49.

Kupberberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M & Caplan, D. (2003). Distinct patters of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272-293.

Li, P., & Green, D. W. (2007). Neurocognitive approaches to bilingualism: Asian languages. *Bilingualism: Language and Cognition, 10*(2), 117-119.

Macmillan, N. A. & Creelman, C. D. (2005). Detection Theory: A User's Guide. 2nd Edition. New York, NY: Cambridge University Press.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp.49-67). Oxford: Oxford University Press.

McLaughin, J., Tanner, D., Pitkaenen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(2), 123-150.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*, 703–704.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning, 60,* 154-193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933-947.

Neville, H. J., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151-165.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Ojima, S., Nakata, H. & Kakigi, R. (2005). An ERP study on second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience 17*, 1212–28.

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language 31*, 785–806.

Osterhout, L. & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language 34*, 739–73.

Osterhout, L., McLaughlin, J., Kim, A., Greewald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking , ERPs, and Beyond* (pp. 271 – 308). New York: Psychology Press.

Osterhout, L., McLaughlin, J., Pitkanen, I, Frenck-Mestre, C. & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A paradigm for exploring the neurocognition of second-language processing. *Language Learning, 56,* 199-230.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*, 509-521.

Pakulak, E & Neville, H. (2011). Maturational Constraints on the Recruitment of Early Processes for Syntactic Processing. *Journal of Cognitive Neuroscience, 23*(10), 2752-2765.

Perfetti, C. A., Nelson, J., Liu, Y., Fiez, J., & Tan, L. H. (2010). The neural bases of reading: Universals and writing system variations. In P. Cornelissen, M. Kringelbach, & P. Hansen (Eds.), *The neural basis of reading* (pp. 147-172). Oxford: Oxford University Press.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model : some implications of research on Chinese for general theories of reading. *Psychological Review, 12*(11), 43-59.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24* (1), 101-118.

Rossi, S., Gugler, M. F., Friederici, A. D. & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience 18*, 2030–48.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24* (3), 397-430.

Sabourin, L., Stowe, L. A. & de Haan, G. J. (2006). Transfer effects in learning an L2 grammatical gender system. *Second Language Research, 27*, 1–29.

Steinhauer, K. & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language, 120*(2), 135-162.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13-41.

Steinhauer, K., White, E. J. & Genesee, F. (2012). Proficiency and first language transfer effects in late second language acquisition: ERP data from a balanced word category violation design. Unpublished manuscript. McGill University, Montreal.

Tan, L. H. (2005). The neuroanatomical system underlying Chinese reading and its constraints on second language learning. *Hong Kong Journal of Paediatrics, 10,* 131-139.

Tan, L. H. Spinks, J. A., Feng, C., Siok, W. T., Perfetti, C. A., Xiong, J., et al. (2003). Neural systems of second language reading are shaped by native language. *Human Brain Mapping, 18*, 158-166.

Tanner, D., Osterhout, L., & Herschensohn, J. (2009). Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. In J. Chandlee, M. Franchini, S. Lord & G.-M. Rheiner (Eds*.), Proceedings of the 33rd Boston University Conference on Language Development* (pp. 528-539). Somerville, MA: Cascadilla.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2012). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. Unpublished manuscript. University of Washington, Seattle.

Tokowicz, N. & MacWhinney, B. (2005). Implicit and Explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition 27*, 173–204.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport, 9*(16), 3556-3560.

Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners. *Cognition, 87*, 129-149.

Weber-Fox, C. M. & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience 8*, 231–56.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Table 1. *Sample Stimuli used at each Testing Session.*

1 a. The teacher didn't/did not start the lesson.
2 a. *The teacher didn't/did not started the lesson.
1 b. The teacher hadn't/had not started the lesson.
2 b. *The teacher hadn't/had not start the lesson.

Numbers and letters refer to the four presentation lists. Half of the participants saw list 1 at session 1 and list 2 at session 2, and vice versa for the other participants. Lists a and b were counterbalanced across participants. Asterisks mark violations, critical verbs are underlined.

Table 2. *Participant Information.*

	Koreans (n=16)		Chinese (n=16)	
Age (years)	22.6 (2.9)		23.9 (5.9)	
Average English use age 0-4 (%):				
at school	1.3 (3.4)		0 (0)	
at home	0.1 (0.3)		0 (0)	
Average English use age 5-11 (%):				
at school	4.1 (7.4)		3.3 (5.7)	
at home	2.3 (4.4)		0.6 (2.5)	
Average English use age 12-14 (%):				
at school	14.9 (14.0)		8.6 (7.8)	
at home	5.1 (8.9)		0.6 (2.5)	
Average English use age 15-16 (%):				
at school	22.7 (14.2)		11.5 (11.7)	
at home	3.2 (6.0)		0.6 (2.5)	
Average English use age 17-18 (%):				
at school	24.8 (16.4)		19.6 (25.6)	
at home	3.6 (5.9)		1.6 (5.1)	
Average English use age 19+ (%):				
at school	26.9 (22.4)		36.7 (37.6)	
at home	6.6 (10.7)		15.9 (31.7)	
Cloze test of English proficiency at session 1 (%)	45.7 (17.1)		51.9 (18.3)	
Final mark in course (%)	67.9 (8.8)		69.1 (10.4)	
L2 self-rating test (7 point scale) M (SD)	**Session 1**		**Session 2**	
Listening	3.8 (1.0)	3.9 (0.9)	**Session 1**	**Session 2**
Reading	4.0 (1.1)	4.2 (1.0)	3.5 (1.0)	4.3 (1.1)
Pronunciation	3.7 (1.2)	3.8 (0.9)	4.1 (1.0)	4.6 (0.7)
Fluency	3.6 (1.2)	3.5 (1.0)	4.2 (1.4)	4.4 (0.7)
Vocabulary	3.6 (1.0)	4.0 (0.9)	3.8 (1.0)	4.3 (0.7)
Grammar	3.9 (0.9)	4.4 (1.1)	3.8 (1.0)	4.3 (0.8)
Total *	3.8 (0.8)	4.0 (0.8)	4.4 (1.0)	5.1 (0.0)
Daily use of English as % of total language use	66.2 (22.7)	67.3 (13.5)	4.0 (0.8)	4.5 (0.4)

* Session 2 > Session 1 $p \leq .005$

Table 3. *Grammatical sensitivity (d-prime) scores at each session for the Korean and Chinese participants. Mean values are reported with standard deviations in parentheses.*

		Session 1	**Session 2**	**Overall**
L1 Groups	**Korean**	1.41 (1.12)	2.15 (1.32)	1.78 (1.26)
	Chinese	1.27 (0.99)	1.96 (1.11)	1.62 (1.09)
	Overall	1.34 (1.04)	2.06 (1.20)	1.70 (1.17)

Note that a complete inability to discriminate (i.e., chance level performance) would yield a d-prime score of 0 and that d-prime scores above 2.5 correspond to very high levels of sensitivity (i.e., proportion correct over 0.90; Macmillan & Creelman, 2005).

Table 4. *Summary of ANOVA F-values and degrees of freedom for comparison of the correct and violation sentences in the analysis of all trials using L1 as a between-subjects variable.*

	df	500-700 ms	750-950 ms
Effects Shared Across Sessions			
Con	1, 30	> 1	3.69$^+$
Con (mid)	1, 30	> 1	4.97*
Con x Lat	1, 30	2.08	6.85*
Con x Hem	1, 30	1.50	5.04*
Con x AP	2, 29	2.85*	3.75*
Con x AP (mid)	2, 29	3.41*	2.04$^+$
Con x Lat x Hem	1, 30	3.75$^+$	5.84*
Con x Lat x AP	2, 29	> 1	2.52$^+$
Con x Lat x L1	1, 30	> 1	5.44*
Changes Between Sessions			
Sess x Con	1, 30	2.62	3.01$^+$
Sess x Con (mid)	1, 30	3.04$^+$	3.80$^+$
Sess x Con x Lat	1, 30	7.57**	4.82*
Sess x Con x Lat x L1	1, 30	4.99*	> 1
Sess x Con x Hem x L1	1, 30	4.90*	1.58
Sess x Con x Lat x Hem x L1	1, 30	1.89	4.07$^+$
Sess x Con x Lat x Hem x AP x L1	2, 29	2.60$^+$	6.37*

$^+p \leq .10$ $*p \leq .05$ $**p \leq .01$ $***p \leq .001$
Con = Condition, AP = Anterior-Parietal, Hem = Hemisphere Sess = Session, L1=L1-background; Mid=midline
Group differences are highlighted in gray.

Table 5. *Summary of ANOVA F-values and degrees of freedom for comparison of the correct and violation sentences in the analysis of correctly answered trials only using L1 as a between-subjects variable.*

	df	500-700 ms	750-950 ms
Effects Shared Across Sessions			
Con	1, 19	> 1	6.41*
Con (mid)	1, 19	1.16	5.46*
Con x Lat	1, 19	1.67	8.77**
Con x AP	2, 18	2.49*	8.74***
Con x AP (mid)	2, 19	4.51*	4.31**
Con x Hem x AP x L1	2, 18	3.74*	2.25
Changes Between Sessions			
Sess x Con	1, 19	1.19	9.49**
Sess x Con (mid)	1, 19	2.08	9.30**
Sess x Con x Lat	1, 19	6.504*	6.67*
Sess x Con x L1	1, 19	4.66*	> 1
Sess x Con x L1 (mid)	2, 18	3.30^{+}	> 1
Sess x Con x Lat x L1	1, 19	> 1	> 1
Sess x Con x Hem x L1	1, 19	8.48**	8.53**
Sess x Con x Lat x Hem x L1	1, 19	2.19	3.54^{+}
Sess x Con x Lat x AP x L1	2, 18	2.38^{+}	1.15
Sess x Con x Lat x Hem x AP x L1	2, 18	> 1	1.40

$^{+}p \leq .10$ $*p \leq .05$ $**p \leq .01$ $***p \leq .001$
Con = Condition, AP = Anterior-Parietal, Hem = Hemisphere Sess = Session, L1=L1-background; Mid=midline
Group differences are highlighted in gray.

Table 6. P600 amplitude for "all trials" and "correct trials only".
Mean amplitude (uV) of the P600 difference wave (violation minus correct sentence) at electrode Pz for each group, session and time window based on "all trials" and "correct trials only". Standard Errors are presented in parentheses. P600 effects that reached significance are shaded in gray.

	Session 1				**Session 2**			
	500-700 ms		750-950 ms		500-700 ms		750-950 ms	
	All Trials	Correct Trials	All Trials	Correct Trials	All Trials	Correct Trials	All Trials	Correct Trials
Korean	-.08 (.83)	-.21 (1.09)	-.30 (.58)	.-30 (.80)	1.92 (.79)	2.46 (.95)	1.61 (.72)	1.86 (.73)
Chinese	.57 (.87)	.72 (1.15)	.55 (.61)	.26 (.83)	.40 (.82)	1.19 (1.0)	2.55 (.76)	3.17 (7.7)

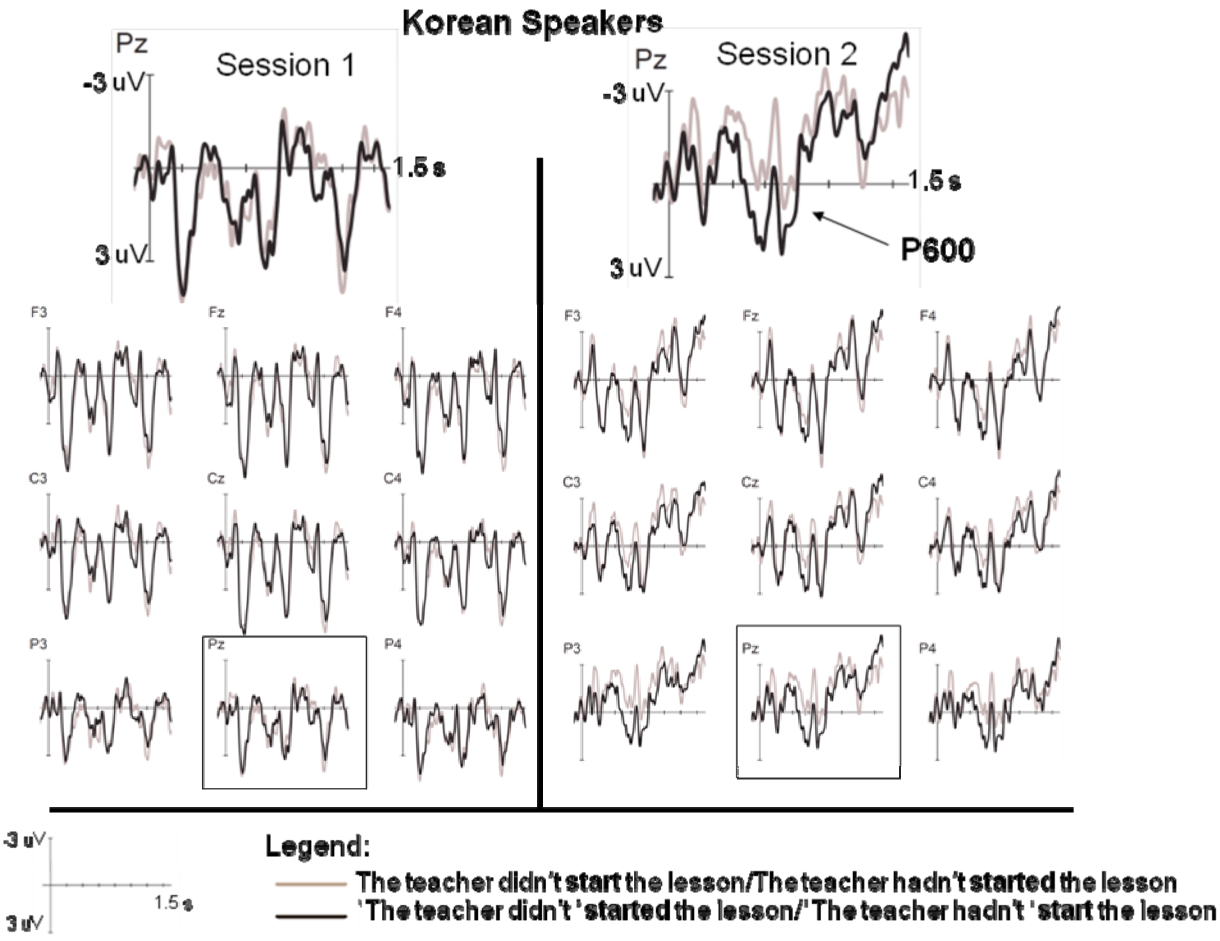


Figure 1. ERP data for Korean participants. Averaged ERPs for the Korean participants at session 1 and 2 for analysis of all trials. All time specifications are relative to the onset of the critical word. The Koreans exhibited a significant P600 at session 2 that was not present at session 1.

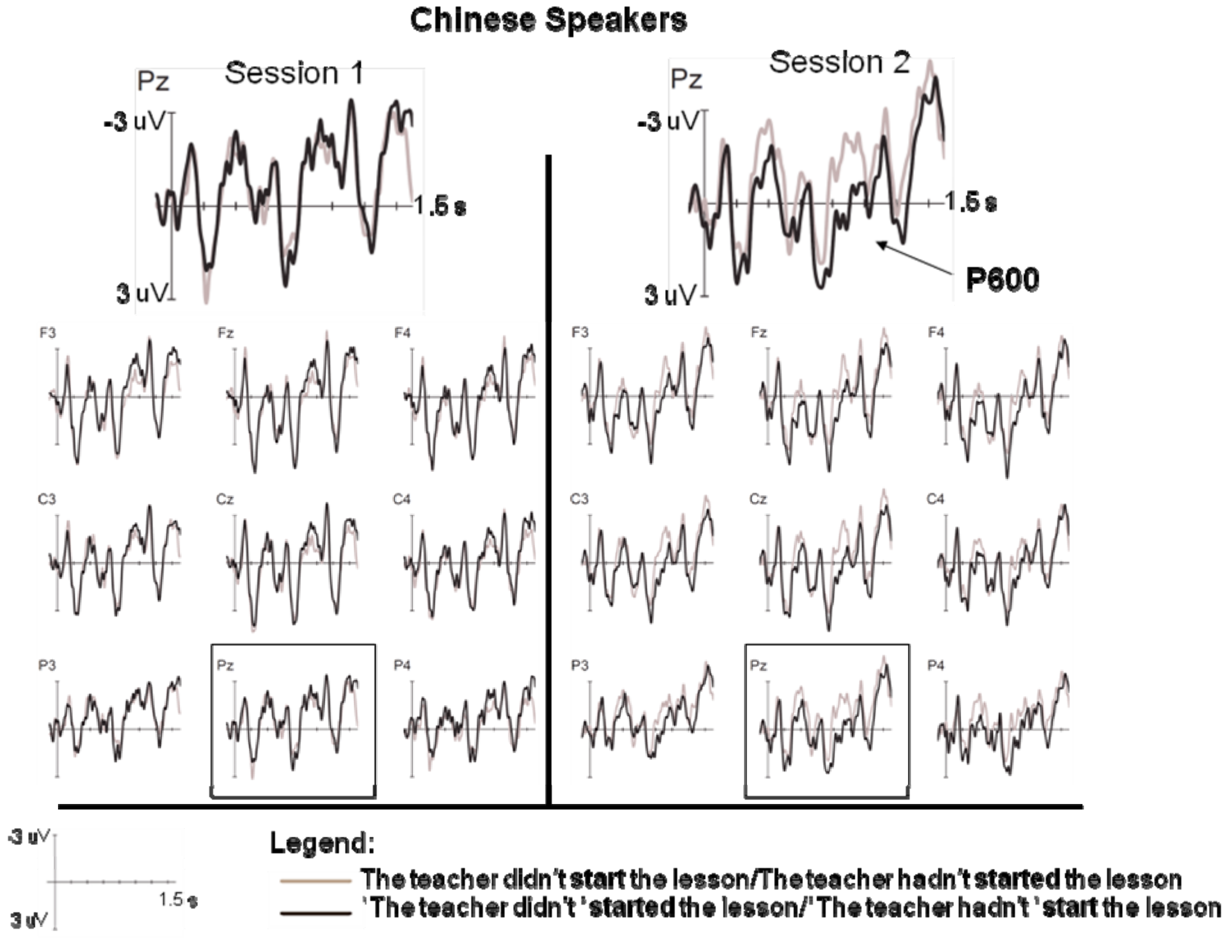


Figure 2. ERP data for Chinese participants. Averaged ERPs for the Chinese participants at session 1 and 2 for analysis of all trials. All time specifications are relative to the onset of the critical word. The Chinese participants exhibited a significant P600 at session 2 that was absent at session 1.

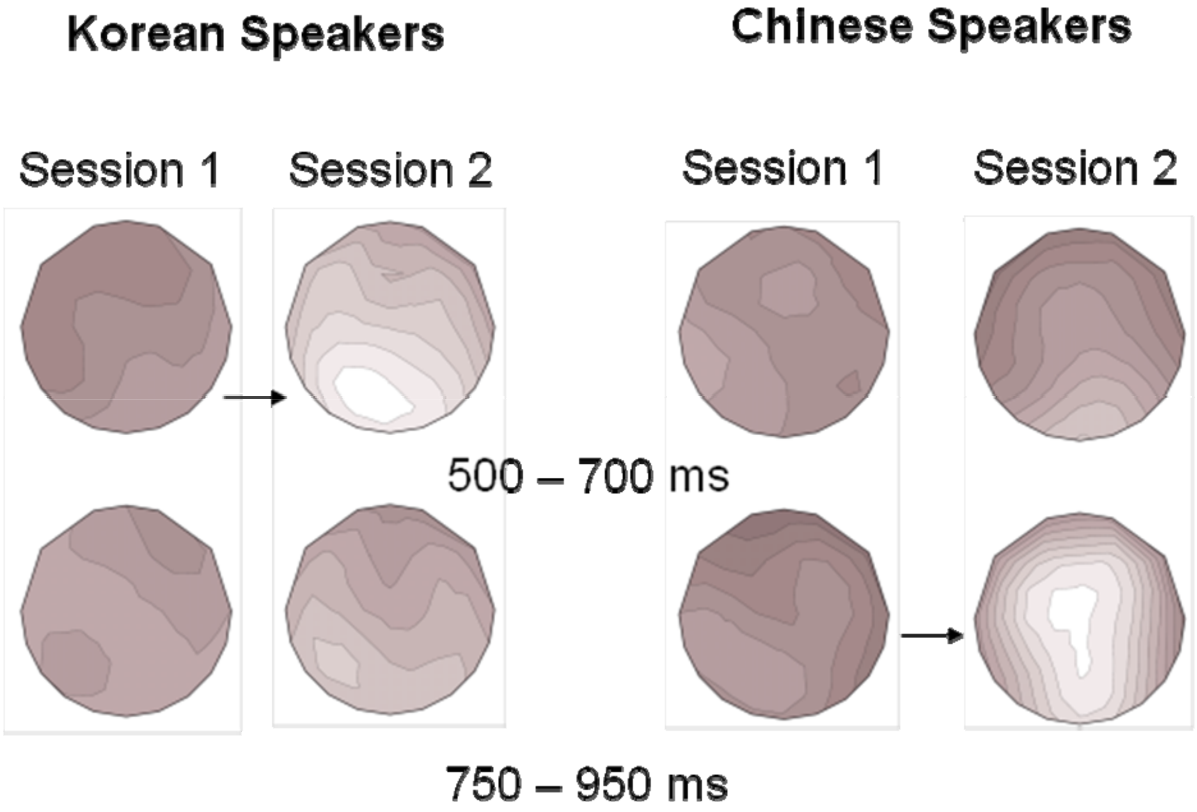


Figure 3. Topographical Maps. Voltage maps for the Korean and Chinese participants at session 1 and 2 in the 500-700 ms and 750-950 ms time windows. Both L1 groups exhibited significant P600 effects at session 2, although they were earlier in the Korean than the Chinese speakers.

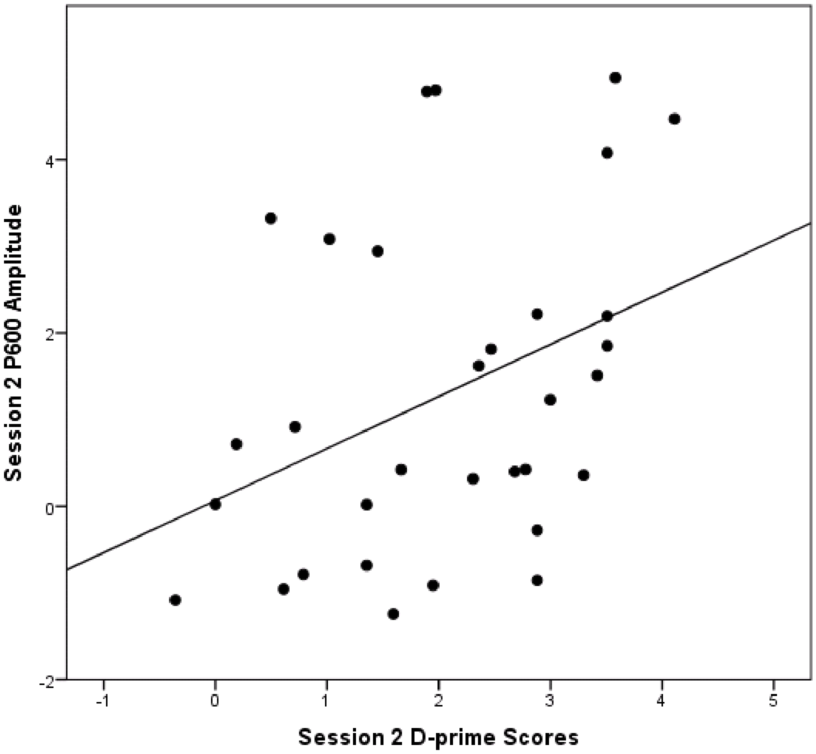


Figure 4. Relationship between ERP Results and Behavioural Performance. Scatter plot showing the correlation between behavioural measures of grammatical sensitivity (d-prime scores) at session 2 and P600 effects (mean amplitude of P600 difference wave at electrode Pz between 500-950 ms).

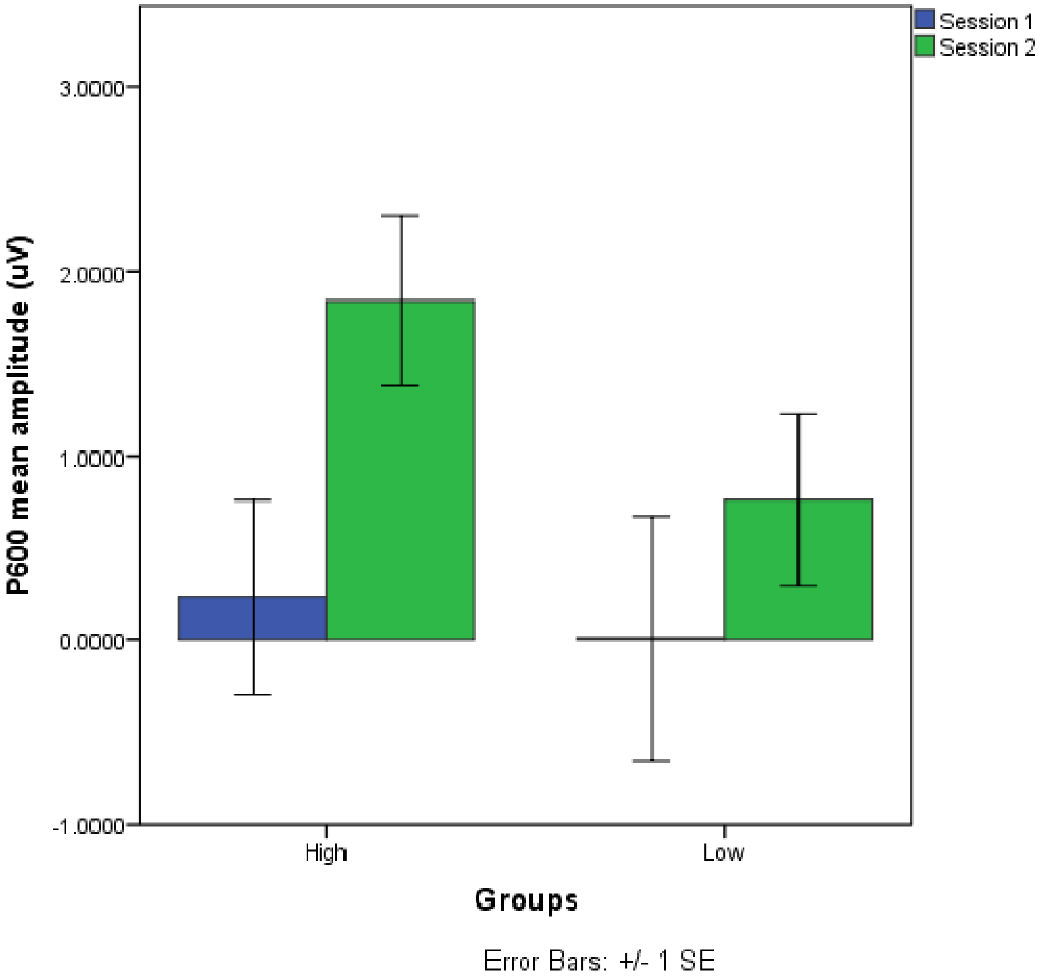


Figure 5. P600 Amplitude in “High” and “Low” Performance Groups. Mean amplitude of the P600 difference wave elicited at electrode Pz between 500-950 ms at session 1 and session 2. Only the “high” group showed a significant change in ERPs between sessions and a significant P600 effect at session 2.

Appendix 1: Sample Stimuli

Condition	**Sample Sentences**
Simple Past	The warning did not scare the travelers. He did not cry at that movie. He did not talk to the reporters. He didn’t walk to the store. The waiter did not serve the food. The girl did not smile at the photographer. She did not lock the doors. She didn’t stay in the hotel. The inmates didn’t watch much television. He did not use any of the tape. He did not play for several weeks. She didn’t turn the key. The teacher didn’t start the lesson. He didn’t stop the fight. The boy didn’t sail for years. The secretary did not mail the package. The clerk didn’t weigh the fruit. The owner did not yell at the managers.
Past Perfect	She hadn’t packed the bags. The student hadn’t learned the lessons. The sailor had not cleaned the kitchen. She hadn’t helped the old man. He hadn’t begged for money. She hadn’t named the baby. The cook had not boiled the lobsters. She hadn’t filed the report. She had not shopped in that neighbourhood. The groom hadn’t kissed the bride. She hadn’t dropped her purse. The guard had not closed the gate. The pupil hadn’t guessed the answer. The driver had not crashed the truck. She had not fixed the car. The salesman had not knocked on every door. He hadn’t filmed the event. The actress hadn’t brushed her hair.

Connecting Text – Study 1 to Study 2

In Study 1, it was found that L2 proficiency played an important role in influencing the neuro-cognitive mechanisms available to late L2 learners during L2 morpho-syntactic processing. Specifically, as L2 learners advanced from relatively low to intermediate levels of L2 proficiency, they began to engage at least one aspect of neuro-cognitive processing used by native speakers (i.e., controlled, aspects of sentence processing, as indexed by P600 effects). An important implication of these findings is that advancing to higher levels of L2 proficiency is associated with a qualitative change in neuro-cognitive processing for late L2 learners. Moreover, these changes occurred even when processing L2 grammatical structures could not be directly transferred from the learners' L1 because these structures either operate differently in the L1 and L2 or they are not present in the L1 at all.

Study 2 further examined the role of L2 proficiency by investigating the extent to which *high* proficiency late L2 learners engage the same neuro-cognitive mechanisms as native speakers during L2 processing. In contrast to Study 1, which examined L2 morpho-syntactic processing at low and intermediate proficiency, Study 2 focused on L2 phonological processing at intermediate and high levels of proficiency. Specifically, native French speakers who had attained either intermediate or high levels of English (L2) proficiency were tested on their processing of an English phonetic contrast that is not used in spoken French and notoriously difficult for native French speakers to acquire: /h/-Ø/. Participants listened to the /h/-Ø/ contrast in two experimental conditions. In Experiment 1, the contrast was presented as simple syllables in an attended discrimination paradigm; in Experiment 2 it was presented in words and pseudo-words in a semantic monitoring task. This allowed for an examination of how stimulus and task factors may influence L2 processing, which has been a focus of relatively few previous ERP studies of L2 processing. A group of monolingual native English speakers was also tested in Study 2, to examine the extent to which high proficiency L2 learners recruit the neuro-cognitive mechanisms underlying native speakers' phonological processing.

Study 1 and 2 complement each other by both examining the role of L2 proficiency (low, intermediate and high) on two domains of language (morpho-syntax and phonology) that are thought to be difficult for L2 learners, in general, according to maturational accounts of L2 acquisition (Johnson & Newport, 1989; Lamendella, 1977; Neville & Bavelier, 2000; Sanders, Weber-Fox & Neville, 2008; Scovel, 1989). Moreover, in both studies, the particular structures examined are thought to be difficult for the specific groups of L2 learners tested because they are not expressed in a similar way in the L1 and, therefore, cannot be directly transferred (i.e., regular past tense verbal morphology for native Korean and Chinese speakers, and the English /h/-/Ø/ phonetic contrast for native French speakers). Thus, both examine the extent to which L2 learners can apply neuro-cognitive processing mechanisms to the processing of linguistic structures that are unique to their L2.

Study 2

Native-like Phonological Processing by Late Second Language Learners: ERP Evidence from Syllable- and Word-Level Processing[9]

Erin Jacquelyn White, Fred Genesee, Debra Titone and Karsten Steinhauer

The work has been submitted publication and is currently under review at *Journal of Memory and Language*

---

[9] Acknowledgements: This work was supported by grants awarded to Dr. Genesee from the Social Sciences and Humanities Research Council (SSHRC), to Dr Steinhauer from the Canada Foundation for Innovation (CRC/CFI) and the Natural Sciences and Engineering Research Council of Canada (NSERC), to Debra Titone from CRC/CFI and NSERC, to Erin White from NSERC and to all authors from the Center for Research on Brain, Language and Music (CRBLM).

Abstract

This Study examined the neuro-cognitive processes that late second language (L2) learners use to discriminate non-native phonetic contrasts (i.e., speech sounds that are not used contrastively in their first language; L1). We tested whether highly proficient late L2 learners, like native speakers, engage in both attention-driven and automatic phonetic processing using event related brain potentials (ERPs). Specifically, we compared processing of a difficult English-specific contrast (/h/versus /Ø/) by native English speakers and late L2 learners of English (French L1) as a function of L2 oral proficiency and stimulus/task demands. In Experiment 1, phonetic contrasts were presented as syllables in an attended categorization task (oddball paradigm). Three ERP components (MMN, N2b and P3b) were examined to investigate automatic and attention-driven stages of phonetic processing. In Experiment 2 contrasts were presented as words and pseudo-word pairs in a semantic monitoring task. N400 effects were examined to investigate phonetic discrimination in a lexical context when attention was directed away from phonetic analysis. High proficiency L2 learners displayed ERP effects similar to native speakers, indicating native-like (and automatic) processing, whereas intermediate proficiency L2 learners showed evidence of attention-driven but not automatic neuro-cognitive processing. We argue that native-like automatic processing becomes increasingly available as late L2 learners advance in proficiency.

Phonological Processing by Late Second Language Learners: ERP Evidence from Syllable- and Word-Level Processing

The effect of age of acquisition on language learning is one of the most longstanding and still controversial issues in the field of second language (L2) acquisition. The age one begins L2 learning is often considered a critical, if not the most critical, factor in determining the level of proficiency attainable in an L2, particularly with respect to syntax and phonology (e.g., Johnson & Newport, 1989; Long, 1990). The controversy is what causes these age effects. According to the widely held Critical Period Hypothesis, child-adult differences in ultimate L2 attainment are due to biologically-based maturational changes in the brain, and native-like levels of L2 attainment are possible only if learning occurs within a sensitive or critical period(s) in neuro-cognitive development (Lenneberg, 1967; Penfield & Roberts, 1959). Consequently, processing of L2 phonology and morpho-syntax is thought to recruit different neural systems in late (i.e., post-puberty) L2 learners compared to native speakers and early L2 learners.

However, recent neuro-cognitive studies indicate that, at least for L2 morpho-syntax, processing is affected primarily by L2 proficiency level, rather than age of L2 acquisition (e.g., Friederici, Steinhauer & Pfeifer, 2002; Morgan-Short, Steinhauer, Sanz & Ullman, 2012; Rossi, Gugler, Friederici & Hahne, 2006; Steinhauer, White & Drury, 2009; Steinhauer, White & Genesee, 2012). More specifically, and contrary to the critical period hypothesis, these studies have shown that late L2 learners who have attained high levels of L2 proficiency display the same patterns of neuro-cognitive processing as native speakers. This includes automatic aspects of processing, which are thought to be particularly difficult for late L2 learners due to involvement of procedural memory systems (Ullman, 2001). In contrast, low proficiency L2 learners appear to engage qualitatively different neuro-cognitive systems and/or recruit native-like processes to a lesser extent than L2 learners with relatively high levels of L2 proficiency. If indeed native-like neuro-cognitive processing is proficiency- rather than age-dependent, then arguments pointing to irreversible maturational-based neurological changes as the cause of child-adult differences in L2 attainment must be reconsidered. The goal of the present Study was to investigate whether native-like L2 phonological processing is possible for late L2 learners who have attained high L2 proficiency. By comparing the neuro-cognitive substrates for phonological processing as a function of L2 proficiency in different stimulus and task contexts, we also sought to shed light on the specific neuro-cognitive processes that are implicated in L2 learning among late L2 learners and the circumstances under which these processes are engaged.

Behavioural research on L2 speech perception in adults has demonstrated that the first language (L1) phonological system plays an important role in the perception of speech sounds in other languages. It is thought that once the L1 phonological system has been established and optimized for detecting and differentiating sounds that convey meaning in the L1, adults will have difficulty discriminating sounds in the L2 that are not used contrastively in the L1 (i.e., non-native phonetic contrasts) because the L1 phonological system assimilates them into a single L1 category or filters out features that are not phonologically relevant in the L1 (Best, 1995; Brown, 2000; Flege, 1995; Kuhl & Iverson, 1995). In fact, a number of current theoretical models of L2 speech perception (e.g., Native Language Magnet Model, Kuhl & Iverson, 1995; Perceptual Assimilation Model, Best, 1995; Speech learning Model, Flege, 1995) argue that most, if not all, difficulties with perceiving (and producing) L2 speech sounds can be explained by how those sounds are categorized by the L1 system. The question is whether late L2 learners, who have already established a functional L1 phonological system, can establish new L2 phonetic category representations and discriminate phonetic contrasts that are phonemic in their L2 but not their L1 using similar processing mechanisms as native speakers.

The evidence for this is unclear. On the one hand, highly proficient late L2 learners show superior discrimination of L2 phonetic contrasts in comparison to less proficient learners (Tees & Werker, 1984). Moreover, laboratory training studies with adult listeners show lasting improvements in non-native speech perception (Logan & Pruitt, 1995). However, findings from other studies indicate that, in contrast to native speakers, successful L2 discrimination by late learners depends on characteristics of the stimulus or task context, suggesting that native speakers and L2 learners use different processing mechanisms. For example, Yoshida (2004) found that while late L2 learners could discriminate L2 phonetic contrasts (i.e., phonemic in L2, allophonic in L1) when listening to vowels during a discrimination task, the same learners failed to discriminate the same contrasts when listening to minimal pairs during a lexical decision task. This difference in discrimination success was not observed for native speakers or for the L2 learners when presented with contrasts that were phonemic in both the L1 and L2. This suggests that for late learners L2 discrimination success may depend not only on similarities between the L1 and L2 phonological systems and on the L2 speakers' level of proficiency, but also on stimulus and task characteristics. Because most previous research has investigated the discrimination of L2 phonetic contrasts presented as syllable pairs, we know relatively little about how stimulus and task factors influence L2 phonetic perception and at which level of processing (see Beddor & Gottfied, 1995, for a discussion). Such an understanding may help elucidate the underlying processes that lead to successful L2 phonological processing in real-life learning situations.

Many current models of L2 speech perception have difficulty accounting for stimulus/task factors because they describe L2 speech perception as an all-or-none phenomenon – that is, L2 sounds are or are not assimilated into L1 categories, and new L2 categories are/are not formed. An alternative view, captured by the Automatic Selective Perception model (ASP; Strange, 2011; Strange & Shafer, 2008) argues that success at L2 phonological discrimination falls along a continuum depending on the stimulus context in which a contrast is presented and on the task demands required during discrimination. According to the ASP model, L2 phonetic discrimination can be successful when contrasts are presented in a simple context and when task demands are minimized so that listeners can attend to the relevant phonetic features; for example, in simple discrimination tasks involving syllables. However, L2 learners are predicted to have progressively more difficulty discriminating the same contrasts as the complexity of the stimuli and task increase; for example, during phonetic categorization tasks involving multiple tokens of each phonetic segment where listeners must simultaneously ignore irrelevant within-category variations and respond only to phonetically-relevant between-category acoustic differences. L2 phonetic discrimination is predicted to be most difficult for late L2 learners when they are engaged in a task that directs their attention away from phonetic analysis in favour of other levels of processing (e.g., when contrasts are presented as words and listeners are attending to meaning). This contrasts with L1 phonetic discrimination which is often successful regardless of stimulus complexity or task demands.

According to the ASP model, the difference between L1 and L2 phonetic processing is that L1 processing occurs automatically by recruiting highly over-learned "Selective Perception Routines" that are attuned to patterns of acoustic-phonetic cues that reliably differentiate L1 phonetic categories. In contrast, L2 phonetic processing is argued to be less automatic and to require late L2 learners use attentional resources to discriminate L2 phonetic contrasts that are not phonemic in their L1. Thus, L2 phonetic discrimination by late L2 learners is predicted to be successful in situations in which phonetic information is within their focus of attention by recruiting attention-driven neuro-cognitive processing mechanisms. However, when L2 learners are attending to other information, such as word meaning, it is predicted that they will fail to discriminate non-native contrasts because they are unable to engage automatic neuro-cognitive processes.

An important claim of the ASP model is that while late L2 learners may develop L2 Selective Processing Routines after considerable L2 experience, these routines can never become as fully automatic as those used for L1 processing because the underlying phonetic category representations will be based on sub-optimal weightings of acoustic-phonetic parameters (Strange & Shafer, 2008). Thus, even highly proficient L2 learners are predicted to perform poorly on L2 discrimination tasks under challenging listening conditions. However, research on the ASP model has been based largely on L2 learners with relatively low proficiency, particularly with respect to pronunciation skills (Strange, 2011). Behavioural research suggests that L2 learners with high L2 oral proficiency (as measured by foreign accent ratings during a sentence elicitation task) display more native-like patterns of speech perception compared to those with low proficiency (Flege & Schmidt, 1995). Moreover, studies of L2 morpho-syntactic processing using event related brain potentials (ERPs) reveal that it is precisely automatic aspects of processing that become available once high L2 proficiency has been attained (Steinhauer et al., 2009; Steinhauer et al., 2012; Rossi et al., 2006). Thus, it is quite possible that L2 phonetic processing becomes increasingly automatic and requires fewer attentional resources as late L2 learners become more proficient. However, to our knowledge, little brain-based evidence exists that directly tests the extent to which highly proficient L2 learners use attention-driven versus automatic phonetic processes (see Hisagi, Shafer, Strange & Sussman, 2010, for work with non-native listeners).

The goal of the present study was to investigate the neuro-cognitive mechanisms used during L2 phonetic processing in different experimental conditions by late L2 learners with different levels of L2 proficiency. Using ERPs, we investigated: (1) whether late L2 learners can exhibit native-like phonetic processing in their L2; (2) how this relates to their level of L2 proficiency; (3) whether late L2 learners can engage both automatic and attention-driven neuro-cognitive processes during L2 phonetic discrimination; and finally (4) under what stimulus and task conditions these processes can be engaged? To address these questions, native English speakers and late L2 learners of English (native speakers of French) who had attained intermediate or high levels of oral L2 proficiency were presented with native and non-native (L2 specific) phonetic contrasts in two experimental conditions. In Experiment 1, phonetic contrasts were presented as individual syllables in an attended categorization task (i.e., participants' attention was drawn exclusively to phonetic processing). Three ERP components (MMN, N2b and P3b) were examined to investigate automatic and attention-driven stages of phonetic processing. In Experiment 2, the contrasts were used to create minimal pairs of words (e.g., happy) and pseudo-words (e.g., "appy") and were presented in a semantic monitoring task. N400 effects were examined to investigate phonetic discrimination in a lexical context when attention was oriented to word meaning rather than phonetic form (see **Table 1** for example stimuli). By comparing results of the same participants when processing the same contrasts in different stimulus and task contexts, we were able to assess whether native-like L2 phonological processing depends on stimulus/task conditions, as predicted by the ASP model. Comparing ERP responses as a function of L2 proficiency allowed us to investigate whether automatic phonetic processing is possible for highly proficiency late L2 learners. More information about the specific ERP components is provided in the introduction to each Experiment.

Insert Table 1 about here

Experiment 1

As with behavioural studies described earlier, most ERP research into the neuro-cognitive basis of L2 phonological processing has examined the discrimination of phonetic contrasts presented as syllables. This work has identified an ERP component, the Mismatch Negativity (MMN), as a sensitive measure of automatic and language-specific phonetic discrimination (for reviews see, for example, Näätänen, Ternaiemi, Sussman, Paavilainen & Winkler, 2001; Näätänen, Paavilainen, Rinne & Alho, 2007). The MMN is a fronto-central negativity elicited by an infrequent “deviant” sound within a sequence of repeating “standard” sounds or sound patterns (i.e., the auditory odd-ball paradigm). The MMN reflects an early stage in detecting a change in auditory presentation. It is elicited between 100-250 ms after the onset of the deviant and largely by generators in the auditory cortex, although other cortical areas are also implicated (Näätänen et al., 1997). Two properties of the MMN are particularly relevant for the present study. First, it can be elicited even when participants are not attending to the stimuli and, therefore, it is thought to reflect automatic, pre-attentive stages of sound discrimination (Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993). Second, for phonetic contrasts, its amplitude and latency appear to depend on the status of the contrast in the listener's L1, in addition to the physical (acoustic) differences between the deviant and standard stimuli. The MMN elicited by phonetic contrasts is enhanced when the contrasting sounds are highly familiar to the listener and correspond to members of distinct phonetic categories in the L1 (e.g., Kazanina, Phillips & Idsardi, 2006; Näätänen et al., 1997; Sharma & Dorman, 2000; see however, Sharma, Kraus, Carrell & Nicol, 1993). These findings support the notion that L1 phonetic discrimination is highly automatic and involves the activation of phonemic representations stored in long term memory (Näätänen et al., 1997).

There is some evidence that for highly proficient L2 learners, L2 phonetic discrimination is also automatic and elicits a MMN response. Winkler et al. (1999a) presented native Finnish speakers, Hungarians who were fluent L2 speakers of Finnish, and monolingual Hungarians with a vowel contrast (/e/- /æ/) that is phonemic in Finnish and allophonic in Hungarian. One /e/ token was the standard stimulus and one /æ/ was the deviant. As expected, the monolingual Hungarians identified both sounds as belonging to the Hungarian /E/ category and displayed no significant MMN effects. In contrast, both the native Finnish speakers and the fluent L2 learners differentiated the sounds and exhibited identical MMN responses. Thus, the proficient L2 learners had learned to discriminate the L2 phonetic contrasts by automatically detecting acoustic-phonetic variation that is allophonic in the L1 and phonemic in the L2 and elicited a MMN response.

However, not all studies have reported native-like MMN effects in late L2 learners in response to L2 phonetic contrasts. Mah (2011; see also Mah, Goad & Steinhauer, 2006) investigated the perception of the English phoneme /h/ in late L2 learners of English who spoke French as a L1. /h/ does not exist in spoken French as a phonemic or allophonic variation of other sounds, and it is notoriously difficult for native French speakers to both perceive (LaCharité & Prévost, 1999) and produce (Janda & Auger, 1992). Many native French speakers have difficulty differentiating /h/ from silence, even after many years of English exposure. During perception, they tend to confuse English minimal pairs that involve the /h/-/Ø/ contrast (e.g., heat-eat) and, during production, they may delete /h/ from h-initial English words (e.g., they may say "ockey" instead of "hockey") or hypercorrect, inserting /h/ to vowel-initial words in English where it does not belong (e.g., "hovertime" instead of overtime). Mah presented native English speakers and native French late L2 learners of English with the English-specific /h/-/Ø/ contrast ("hum" vs "um") using an adapted "categorical oddball" paradigm, to test whether French-speakers' difficulties with /h/ reflect difficulties in automatically classifying it as a distinct phonetic category. This paradigm differs from that used in the Winkler study in two ways. First, participants hear multiple tokens of each contrasting phonetic category. Thus, to detect the difference between the standard and deviant stimuli, listeners must ignore irrelevant acoustic differences between tokens that belong to the same phonetic category (e.g., stimuli length) and respond only to the phonologically-relevant acoustic cues that differentiate members of each category (see also Dehaene-Lambertz, Dupoux & Gout, 2000; Hisagi et al., 2010; Kazanina et al., 2006). Second, tokens of each contrasting phonetic segment are presented as standards in one stimulus block and as deviants in another; the MMN is obtained by comparing the ERP response generated by the exact same sounds when presented as a deviant versus standard. This procedure rules out any confound between the relevant mismatch effect and other irrelevant ERP effects than could arise from acoustic differences between the standard and deviant stimuli (Näätänen et al. 2007). When listening to phonetic contrasts in this categorical design, L2 learners are less likely to base their discrimination on readily available acoustic cues and, instead, must classify the speech sounds into two distinct phonetic categories. Using this paradigm, Mah (2011) found that while native English speakers exhibited a clear MMN in response to the /h/-/Ø/ contrast, the L2 learners did not.

There are a number of ways to account for the discrepancy in results reported by Winkler et al. and Mah. First, the experimental designs differed in the extent to which acoustic differences between the standard and deviant stimuli could be used as reliable cues for discrimination. According to the ASP model, L2 learners with relatively little L2 experience may successfully discriminate L2 phonetic contrasts in Winkler et al.'s paradigm by responding to differences in the acoustic properties of the stimuli; however discriminating the same contrasts may be more difficult when acoustic differences are controlled for or when the critical between-category differences occur with other irrelevant within-category differences[10]. Thus, Mah's categorical oddball paradigm is a better test of whether L2 learners, like native speakers, classify stimuli into two distinct phonetic categories. Second, the English /h/-/Ø/ contrast may be more challenging for French speakers to acquire than the Finnish /e/- /æ/ contrast is for Hungarians. The English /h/ is thought to be specified by a distinctive feature (laryngeal: spread-glottis) that is not used to distinguish any French phonemes (Mah, 2011). According to Brown (2000), late L2 learners will have persistent difficulty acquiring L2 phonemes that are specified by distinctive features that are not used contrastively in their L1 at all. In contrast, the Finnish /e/ and /æ/ differ in vowel height, which both Hungarian and Finnish use contrastively. The English /h/ is also acoustically non-salient – it is a voiceless, non-strident fricative with low overall intensity (Mah et al., 2006), making it difficult to perceive (Polka, 1992). Third, the L2 learners tested by Mah may have been less proficient than those tested by Winkler and, thus, the lack of MMN may mean they had not had enough experience with English to automatically discriminate the /h/-/Ø/ contrast. As neither study provided detailed information about their L2 learners' level of L2 proficiency, this possibility is difficult to assess.

The present Experiment attempted to disentangle these possibilities by presenting the /h/-/Ø/ contrast in a similar categorical oddball paradigm as used by Mah to native English speakers and two groups of native French-speaking late L2 learners of English who differed in their level of L2 proficiency. We examined

---

[10] In fact, Mah (2011) found that the same French speakers displayed a significant MMN in response to /h/ when it was presented in the context of fricative noise bursts (e.g., "hf" vs. "f") rather than in the context of syllables (e.g., "hum" vs. "um"). Mah interpreted this to mean that native French speakers can perceive the acoustic properties of /h/ but have difficulty identifying it as a distinct phonetic category in a linguistic context.

whether, in the absence of reliable acoustic cues to signal the phonetic contrast, high proficiency L2 learners could engage automatic processing of L2 phonetic contrasts presented in syllables, as evidenced by a MMN response.

We also investigated the extent to which L2 learners make use of controlled, attentional processes to categorize L2 phonetic segments, as predicted by the ASP model. To this end, we examined two additional ERP components (N2b, P3b), which are thought to be associated with later attentional stages of phonetic discrimination in native speakers. The N2b is a central-posterior negativity that directly follows the MMN (and partially overlaps it) in time, typically between 200-300 ms after the deviant stimuli. The P3b (sometimes referred to as the "P300") is a posterior positivity that follows both negativities, typically between 300-500 ms. Unlike the MMN, the N2b and P3b are not specific to auditory processing (Ritter, Simson & Vaughan, 1983) and are most prominent when participants attend to test stimuli (Näätänen et al., 1993). The N2b is thought to reflect a domain-general attention orienting response (Näätänen & Gaillard, 1983). The P3b is widely associated with updating working memory after the presentation of an infrequent deviant that is highly relevant to the task at hand (Donchin, 1981). In native speakers, the N2b and P3b are typically elicited in attended oddball paradigms. Thus, to examine these components, participants in the present study performed a discrimination task in which they pressed a button when they heard a sound that differed from the others (e.g., an attended odd-ball paradigm). However, in contrast to some previous ERP studies of non-native phonetic discrimination (e.g., Hisagi et al., 2010), participants were not explicitly told which contrasts they would hear; nor were they given specific instructions as to what to listen for. This was to avoid teaching L2 learners the contrast before testing and to examine the extent to which they could spontaneously classify tokens into L2 phonetic categories (see Hisagi & Strange, 2011).

By examining all three ERP components, Experiment 1 sought to investigate the extent to which high proficiency late L2 learners, like native speakers, engage in multiple stages of phonetic discrimination when presented with simple syllables, and when their attention is oriented to phonetic information. These stages include early, automatic stages (indexed by the MMN), and later attention-based stages (indexed by the N2b/P3b). We also examined the final outcome of these processes: participants' accuracy in the categorical discrimination task. We predicted that in response to a control condition involving a phonetic contrast that is phonemic in both the L1 and L2 (/f/-/Ø/), both intermediate and high proficiency late L2 learners would display relatively accurate behavioural performance and exhibit similar MMN, N2b and P3b responses as native speakers. Following the ASP model, we also expected both late L2 learner groups would engage attention-based processing of the L2-specific /h/-/Ø/ contrast as evidenced by significant N2b/P3b responses. Of particular interest was whether they would engage in early automatic stages of L2 phonetic categorization (indexed by a MMN for the /h/-/Ø/ contrast) and whether this would be more evident in those who had attained high levels of L2 proficiency.

Method

Participants. Fifteen monolingual English speakers (19-35 years old, M = 22.6, 9 female) and 27 native French speakers who spoke English as a L2 (20-35 years old, M = 26.4, 15 female) participated in the study. An additional 16 participants were tested but were excluded because they reported speaking an additional language during childhood (n=4), because of technical problems during data acquisition (n=2), or because of excessive movement, eye-blink or alpha artifacts contaminating the EEG signal in Experiment 1 or 2 (n=10). All were living in the Montreal area, a French-English bilingual city, at the time of testing. Participants gave written informed consent, were paid for their participation, and reported no history of hearing, language, speech or neurological disorders. All were right-handed (assessed using self-report and the Edinburgh handedness inventory; Oldfield, 1971) and reported comparable educational backgrounds (i.e., most were currently undergraduate students).

L2 Proficiency. The L2 learners were divided into intermediate and high proficiency groups based on ratings of their English pronunciation given by two native English speakers who were not involved in data collection (see also Piske, MacKay & Flege, 2001). Ratings were based on speech samples obtained from the L2 learners and approximately half of the native speakers following the EEG session; participants were asked to speak freely on a chosen topic (e.g., “what makes a person healthy and happy?”) while being recorded; approximately 30 seconds of speech was selected for assessment. Raters were asked judge each sample on 6 dimensions: overall pronunciation, pronunciation of /h/-initial words, vocabulary, grammar, fluency, and overall nativeness, using a 14-point scale (0= not at all like a native English speaker, 13= definitely a native English speaker). They were first trained together with speech samples from participants who were not included in the ERP analyses and then independently rated participants reported here. Bivariate correlations of the ratings given by each rater for the 6 dimensions revealed r values that ranged from 0.74 to 0.91 ($p < .001$), indicating that they rated the samples in a similar way.

The L2 learners were divided into two approximately equal groups based on their "overall pronunciation" ratings because this was the focus of the present study. The high proficiency (HP) group had pronunciation ratings of 9/13 or higher by both raters (n=12; 20-35 years old, mean age= 27.3 years; 5 female), and the intermediate proficiency (IP) group had ratings of 8 or lower by at least one judge (n=15; 21-34 years old, mean age= 25.7 years; 10 female). Mean pronunciation rating for the HP group was 10.6 versus 6.6 for the IP group. To test whether groups differed on other aspects of oral proficiency (in addition to pronunciation), a repeated measures ANOVA was run on the speech sample ratings with the 6 dimensions as a within-subjects variable and group as a between-subjects variable. This revealed a significant main effect of group [$F(1,25) = 55.90, p <.001$], indicating that the HP group was rated more "native-like" than the IP group on all dimensions of proficiency; there was no significant group by scale interaction [$F(5,21) = 1.25, p > .10$]. The average ratings of the IP and HP groups on each of the 6 scales and overall are presented in **Table 2**. The rating results were confirmed by participants' self-reports and their performance on a cloze test of English proficiency, which revealed higher L2 proficiency for the HP compared to the IP group (see **Table 2**).

Insert Table 2 about here

Previous and Current English Exposure. Age of L2 Acquisition (AoA) was defined as participants' first significant and intense exposure to English; that is, when they were called upon to use English regularly for communicating with native-speakers, as in White and Genesee (1996). Thus, to ensure that L2 learners were in fact late learners of English, several inclusionary criteria were used: (1) acquired English as a L2 after the age of 14 (i.e., after the close of the hypothesized critical period; Birdsong, 2006); (2) reported very limited exposure to English or other languages at home before this age; and (3) the English-as-second language instruction received at school as children was regular curriculum instruction. Mean AoA for HP group was 19.0 years and for the IP group was 18.9 years, a difference which was not statistically significant [$t(25) = .05$, $p > .10$]. Life-time English exposure was also assessed by means of a questionnaire in which participants reported how much English they used at home and at school (as a percentage of total language use) between the ages of 0-4, 5-11, 12-14, 15-16, 17-18, and 19+. To examine whether the HP and IP groups differed in early English exposure, a repeated measures ANOVA was run using group as a between subjects variable; age range (0-4, 5-11, 12-14 and 15-16) and location (home, school) were within subjects variables. This revealed no significant main effect or interaction with the factor group (*p*s > .10), indicating a similar amount of early English exposure for the IP and HP groups. As seen in **Table 2**, both groups reported limited English use as children and adolescents.

The questionnaire also asked about current English exposure as a percentage of daily language use in six contexts: with friends, family, at work/school, during their spare time, watching TV/listening to the radio, and reading for pleasure (see **Table 2** for group means). To test for differences between the IP and HP groups, a repeated measures ANOVA was conducted using context as a within-subjects variable and group (HP, IP) as a between subjects variable. This revealed that the HP group used English significantly more than the IP group in multiple aspects of their adult lives [M= 43.9 % vs. 25.2 %, $F(1, 25) = 10.37$ $p < .005$].

Stimuli. The stimuli consisted of two tokens each of three different speech sounds, “ha”, “a” and “fa” (**Table 1)**. The /f/-/Ø / contrast was chosen as a control because it is phonemic in both French and English, and like /h/, /f/ is a low intensity fricative. The stimuli were recorded by a female native English speaker. One of the “a” tokens was longer in duration (384 ms) and roughly matched the overall length of the two “ha” (402, 413 ms) and two “fa” (395 and 429 ms) tokens; the other “a” token was shorter (332 ms) and comparable to the vowel length of one of the “ha” (331 ms, 349 ms) and “fa” tokens (341 and 354 ms). This ensured that participants could not use length of the token as a reliable cue for detecting the deviant phonetic categories. Any ERP effects are, thus, more likely to reflect phonetic categorization rather than detection of pure acoustic differences (Kazanina et al., 2006).

The stimuli were presented in 4 five-minute blocks so that the “ha”, “fa” and “a” sounds could each serve as a standard and deviant stimulus. The 4 blocks consisted of: “ha” deviant among “a” standards, “a” deviant among “ha” standards, “fa” deviant among “a” standards, and “a” deviant with “fa” standards; the order of blocks was counter-balanced across participants. In each block, 50 “deviant” stimuli were presented intermixed with 235 standard stimuli in an 82.5 to 17.5 % standard-to-deviant ratio. The order of standards and deviants was pseudo-randomized to ensure that each deviant was preceded by a minimum of 3 standards. Both tokens of each sound were used equally and randomly within a block.

Procedure. Participants were seated comfortably in a sound-attenuated room, approximately 70 cm in front of a computer screen. They were given instructions, both orally and in writing, that they would hear a series of speech sounds, many of which were the same, but some of which were different. They were asked to listen carefully and to identify any speech sound that they judged to be different from the others in the series by clicking a mouse button upon hearing the deviant sound. No explicit instructions were given about the /h/-/Ø/ or /f/-/Ø/ contrasts. They were also asked to move and blink as little as possible and only between instances of the stimuli. The experiment began with the presentation of 7 practice sounds, followed by a short break in which they could ask questions. To minimize eye movements, each block began with the presentation of a fixation cross that remained in the centre of the computer screen for the duration of the block, and participants were asked to fixate on the cross for each trial. Stimuli were presented binaurally at 70 dB via inserted headphones (Etymotic Reseach) and were set to a comfortable volume by each participant. They were presented with a variable stimulus onset asynchrony (SOA) of 1000, 1020, 1040, 1060, 1080, 1100, 1120, 1140, 1160, 1180, or 1200 ms that was randomly distributed across stimuli. A variable SOA ("jitter") prevented participants who were unable to perceive /h/ from using any perceived delay between the consonant and vowel stimuli as a reliable cue for detecting deviants (Phillips et al., 2000). A relatively long SOA was used to increase the likelihood that discrimination occurred by referencing phonetic representations stored in long-term memory rather than the detection of acoustic differences using short-term sensory memory templates (Molnar, Baum, Polka & Steinhauer, 2009; Werker & Logan, 1985). Prior to the EEG session, the participants completed the language-background questionnaires. Experiment 1 lasted for approximately 20-25 minutes, including short breaks.

EEG Recordings and ERP Analyses. Continuous EEG was recorded with a BioSemi ActiveTwo amplifier system (http://www.biosemi.com) from 64 cap-mounted Ag-AgCl electrodes according to the international 10-20 system, digitized online at 250 Hz and referenced to an additional active electrode (common mode sense; CMS). Eye movements were monitored by additional electrodes placed at the outer canthus of each eye (EOGH) and above and below the left eye (EOGV). Electrode offsets were kept below 25 µV.

Offline, the EEG was re-referenced to the averaged left/right mastoids and filtered with a 0.5-40 Hz band-pass filter using the EEProbe software package (Advanced Neuro-Technology, ANT; Enschede, the Netherlands). An epoch was rejected whenever the standard deviation for a 200 ms moving window exceeded 25 mV in EOG or EEG channels included in ERP analysis (see below). The EEG data for individual participants were also thoroughly examined for eye movements, and muscle or other noise artifacts; contaminated epochs were excluded. Participants were included in further analyses if they contributed a minimum of 50 artifact-free trials in each of the Deviant and Standard conditions.

To ensure that there was an approximately equal number of trials contributing to the grand averages for each condition, ERPs were computed for each participant in response to the 50 deviant stimuli within each block and the 50 standard stimuli that directly preceded the deviants (rather than all 235 Standards). This procedure ensured that ERPs for the standard condition were not smaller in amplitude simply because their averages were based on more trials. Four conditions were created for each participant: ***Deviant H*** (average response to "ha" deviants among "a" standards and "a" deviants among "ha" standards), ***Standard H*** (average response to the corresponding standards), ***Deviant F*** (average response to the "fa" deviants among "a" standards and "a" deviants among /fa/ standards), ***Standard F*** (average response to the corresponding standards). Collapsing across sub-conditions in this way increased the signal-to-noise ratio and decreased any potential block effects related to the participants' physical or mental state. Moreover, because the same stimuli were presented as both standards and deviants, we could be more confident that any ERP effects were due to the detection of a phonemic contrast, rather than the particular acoustic or phonemic properties of the individual stimuli themselves (Näätänen et al., 2007; Sharma & Dorman, 2000). Analyses were also conducted on the sub-conditions to investigate whether there were differences in ERP responses depending on Deviant Type (e.g., deviant “ha” with standard “a” versus deviant “a” with standard “ha”). This revealed similar effects to the analyses of the collapsed conditions; differences between analyses, when present, are reported below.

For each participant, artifact-free ERP responses were averaged for each of the 4 conditions. This was done for a 650 ms interval, time-locked to the onset of the deviant/standard stimuli, including a 100 ms pre-stimulus baseline interval. Statistical analyses were conducted on the mean amplitude of the ERP waveform within time windows that were selected based on previous literature and visual inspection of the grand averages: 100-175 ms (MMN), 175-250 ms (N2b) and 300-500 ms for the subsequent positivity (P3b). For each time window, repeated-measures ANOVAs were performed on 12 representative lateral (F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8) and 3 midline (Fz, Cz, Pz) electrodes, as effects were most prominent at these electrodes and no additional components of interest emerged at other sites. For the lateral electrodes, this analysis involved Group (Gp) as a between-subjects factor (NS, HP and IP) and the following within-subjects factors: Contrast (/h/-/Ø/, /f/-/Ø/), Deviant (Deviant, Standard), Anterior-Posterior (Anterior, Central, Parietal) and Hemisphere (Left, Right). For the midline sites, the factors were: Gp, Con, Dev and Anterior-Posterior (Anterior, Central, Parietal). Results are reported for main effects and interactions that involve at least one Dev factor. To simplify presentation of the results, midline analyses are reported only when they yielded results that were not revealed by analyses of the lateral electrodes. The Greenhouse-Geisser correction was applied to all analyses involving the Ant-Post factor (as it involves more than one degree of freedom), and corrected *p* values are reported.

Results

Accuracy Results. Accuracy in the discrimination task (i.e., correct identification of a deviant) was analysed with a repeated measures ANOVA using Con (/h/-/Ø/ vs. /f-Ø/), Dev Type ("a" deviant among "ha"/ "fa" standards vs. "ha"/"fa" deviants among "a" standards) and Token (1 vs. 2) as within-subjects variables and Gp (NS, HP, IP) as a between subjects variable. This revealed a main effect of Con [$F(1, 39) = 10.46, p < .005$], indicating that all groups were more accurate detecting the /f/-/Ø/ control contrast compared to the /h/-/Ø/ contrast. A significant effect of Dev Type [$F(1, 39) = 38.14, p < .001$], revealed that, across all groups, deviants "ha" and "fa" were more difficult to identify than were "a" deviants. An additional Dev Type x Con interaction [$F(1, 39) = 14.17, p \leq .001$], showed that, (again across all groups), the most difficult deviant to identify was "ha" (presented with "a" as the standard; M = 65.7 %), followed by "fa" (with "a" standard; M = 87.5 %), a difference which was significant [$F(1, 39) = 12.53, p < .001$]. By contrast, accuracy identifying the deviant "a" was similar when "ha" was the standard (M= 96.9 %) compared to when "fa" was the standard (M=95.3 %), a numeric difference which was not significant [$F(1, 39) = 3.14$, $p = .084$].

There was no main effect of Gp or Gp x Con interaction, indicating similar discrimination for both contrasts for all three groups. However, a Dev Type x Token x Gp interaction [$F(2, 39) = 5.39$, $p < .01$] revealed that the IP group responded differently to the two “ha” and “fa” tokens [$F(1, 14) = 4.94$, $p < .05$], such that accuracy was higher for the longer compared to the shorter tokens (M = 73 vs 69 %). This difference was not observed for the NS or HP groups ($p > .10$) and suggests that whereas the IP group may have used stimulus length as a cue for detecting deviance (i.e., an irrelevant acoustic cue in English), the HP and NS groups ignored this within-phonetic category difference and responded instead to the presence/absence of the /h/ and /f/ sounds (see **Table 3** for group averages).

Insert Table 3 about here

ERP Results. The ERP grand mean waveforms for the NS, HP and IP groups for the /h/-/Ø/ and /f/-/Ø/ contrasts are presented in **Figures 1-3**. These figures show a central-posterior negativity followed by a posterior positivity in response to the deviant stimuli for all groups. For the NS and HP groups, two negative peaks (one between 100-175 ms and the other between 175-250 ms) can be seen in the ERP difference waves for both the /h/-/Ø/ and /f/-/Ø/ contrasts. For the IP group, a similar pattern is observed for the /f/-/Ø/ contrast; however no clear negative peak can be seen within the early time window for /h/-/Ø/. Results of the global ANOVAs for the three time windows are presented in **Table 4** and resulting effects are described below.

Insert Figures 1-3 and Table 4 about here

In the MMN time window, between 100-175 ms, a significant main effect for Dev and a significant Dev x AP interaction revealed a highly significant negativity for deviants at central [$F(1, 39) = 26.32, p < .001$] and posterior [$F(1, 39) = 50.62, p < .001$] electrodes across all groups and both contrasts. A Dev x Con interaction indicated that the effect was larger and more widely distributed for the /f/-/Ø/ compared the /h/-/Ø/ contrast, as revealed by a significant main effect for /f/-/Ø/ [$F(1, 39) = 25.43, p < .001$] but not /h/-/Ø/ [$F(1, 39) = 3.47, p = .07$]. No interaction with group was observed. However, as can be see in **Figure 3**, the IP groups' response to the /h/-/Ø/ contrast is smaller than their response to the /f/-/Ø/ contrast. To investigate whether each group in fact exhibited significant effects for both contrasts, analyses were conducted for each group separately. For the IP group, a significant Dev x Con interaction [$F(1,14) = 10.13, p < .01$] revealed a significant negativity for /f/-/Ø/ [Dev: $F(1,14) = 10.32, p < .01$] but not /h/-/Ø/ [Dev: $F < 1$]. In contrast, for the NS and HP groups, no significant interactions with the factor Con were observed. For these groups, significant Dev x AP interactions [NS: $F(2,13) = 4.88, p \leq .05$; HP: $F(2,10) = 12.20, p < .001$] revealed significant negativities across all deviants that were largest at posterior electrodes for both contrasts [NS: $F(1,14) = 10.15$, $p < .01$; HP: $F(1,11) = 20.98$, $p < .001$][11], indicating statistically similar effects for both contrasts.

In the N2b interval, between 175-250 ms, a main effect of Dev and a Dev x AP interaction, again, indicated a highly significant negativity at central [$F(1, 39) = 24.45$, $p < .001$] and posterior [$F(1, 39) = 78.03$, $p < .001$] electrodes across all groups and both conditions. A number of interactions with the factors Dev and Con were also observed (**Table 4**). Follow-up analyses revealed a significant negativity for both /h/-/Ø/ [Dev: $F(1, 39) = 4.76$, $p < .05$] and /f/-/Ø/ [Dev: $F(1, 39) = 21.25$, $p < .001$], although this effect was larger for /f-Ø/, particularly over the right hemisphere [Right: Dev x Con $F(1, 39) = 6.85$, $p < .05$]. To be consistent with analysis of the earlier time window, analyses were conducted for each group separately. This revealed significant interactions with the factor Con for the IP group only, including Dev x Con x AP [$F(2, 13) = 2.88$, $p < .05$] and Dev x Con x AP x Lat [$F(2, 13) = 3.89$, $p < .05$] interactions. However, unlike the earlier time window, the IP group exhibited a significant posterior negativity for both /h/-/Ø/ [post: $F(1, 14) = 10.25$, $p < .01$; Dev x AP: $F(2, 13) = 2.78$, $p < .05$] and /f/-/Ø/ [post: $F(1, 14) = 17.39$, $p \leq .001$; Dev x AP: $F(2, 13) = 9.90$, $p < .001$], albeit larger for /f-Ø/, particularly at lateral posterior electrodes [Dev x Con: $F(1, 14) = 4.91$, $p < .05$]. For the NS and HP groups, no significant interactions with contrast were observed (*p*s > .10). Thus, whereas statistically similar effects were observed

---

[11] To be thorough, analyses were also conducted for the /h/-/Ø/ and /f/-/Ø/ contrasts separately. Unlike the IP group, the HP group exhibited a significant main effect of deviant between 100-175 ms for both the /h/-/Ø/ [$F(1,11) = 7.25$, $p < .05$] and /f/-/Ø/ [$F(1,11) = 8.03$, $p < .05$] conditions.

for the NS and HP groups for the /h/-/Ø/ and /f/-/Ø/ contrasts, the IP group exhibited smaller effects for the /h/-/Ø/ contrast in this time window[12].

In the P3b interval, between 300-500 ms, a large and highly significant positivity was observed for all groups and both contrasts. Consistent with P3b effects, this positivity was largest at posterior medial electrodes, [$F(1,39) = 84.99$, $p < .001$; see **Table 4** for values from the Global ANOVA]. Although significant effects were observed for both contrasts, a Dev x Con x AP x Hem interaction suggests somewhat different topographical distributions. For /f-Ø/, a larger positivity was observed at left posterior [$F(1,39) = 84.38$, $p < .001$] compared to right posterior [$F(1,39) = 62.66$, $p < .001$] electrodes, as revealed by a significant Dev x Hem interaction at posterior electrodes [$F(1,39) = 5.91$, $p < .05$]. For /h-Ø/, the Dev x Hem interaction was not significant ($p > .10$), indicating a similar effect at both left posterior [$F(1,39) = 41.69$, $p < .001$] and right posterior [$F(1,39) = 50.43$, $p < .001$] sites. One marginal interaction with the factor Gp was observed in this time window [Dev x Con x Hem x Gp: $F(2,39) = 2.55$, $p = .091$]; however, this effect did not lead to further significant differences between groups or contrasts. As can be seen in **Figures 1-3,** all groups exhibited a large positivity that was highly significant at posterior medial electrodes in both /h/-/Ø/ and /f/-/Ø/ contrasts (*ps* < .005).

---

[12] Between 100-175 ms, analyses of the sub-conditions revealed significant Dev x Dev Type [$F(1, 39) = 7.33$ $p \leq .01$] and Dev x Dev Type x Con x AP [$F(2, 38) = 8.33$ $p < .005$]. Follow up analyses were conducted for each Dev Type to investigate whether ERP responses were significantly different for that stimulus when it was presented as a standard or deviant (e.g., "ha" presented as a deviant versus as a standard). This revealed significant main effects of Dev for "ha" [$F(1,39) = 9.07$, $p \leq .005$], "fa" [$F(1,139) = 10.83$, $p < .005$] and "a" when it was presented with "fa" [$F(1, 39) = 10.83$, $p < .005$]. The main effect of Dev was not significant for "a" when it was presented with "ha", however significant effects were observed at posterior electrodes [Dev: $F(1,11) = 5.23$, $p < .05$; Dev x AP: $F(2, 38) = 14.83$ $p < .005$].

Analysis of Mastoid electrodes. In order to differentiate MMN and N2b effects, additional analyses were conducted at mastoid electrodes within the 100-175 ms and 175-250 ms time windows (using the nose as the reference electrode), because the MMN, but not the N2b, is known to reverse in polarity along the sylvian fissure (Alho, Paavilainen, Reinikainen, Sams & Näätänen, 1986; Novak, Ritter, Vaughan & Wiznitzer, 1990). Thus, a positivity at mastoid sites, would indicate effects due to the MMN alone. Repeated measures ANOVAs were conducted to compare effects at mastoid electrodes with effects at electrode Pz (which showed the largest negativity in the analyses reported earlier). The factors were <u>Dev</u>iant (Deviant, Standard), <u>Ele</u>ctrode (Pz, left mastoid, right mastoid), <u>Con</u>trast (/h-Ø/, /f-Ø/) and <u>Gr</u>oup (NS, HP, IP). Significant main effects and interactions with the factor Dev are reported. Greenhouse-Geisser corrected *p* values are reported for interactions involving the factor Electrode. One IP participant was excluded due to missing nose electrode data.

Between 100-175 ms, this revealed significant Dev x Ele [$F(2,37) = 16.73$, $p < .001$] and Dev x Ele x Con [$F(2,37) = 2.34$, $p < .05$] interactions. Analysis at electrode Pz revealed a significant main effect of Dev [$F(1,38) = 10.29$, $p < .005$], indicating a negativity, as reported earlier. At left and right mastoids, the response to the deviant stimuli was more positive than to the standard, although this was not significant (Fs < 1). Between 175-250 ms, a main effect of Dev [$F(1,38) = 17.04$, $p < .005$] and a Dev x Ele interaction [$F(2,37) = 33.73$, $p < .001$] revealed significant negativities at electrode Pz [$F(1,38) = 43.72$, $p < .001$] as well as the left [$F(1,38) = 4.22$, $p < .05$] and right [$F(1,38) = 4.80$, $p < .05$] mastoids. Thus, a polarity reversal at mastoid electrodes was observed between 100-175 ms but not between 175-250 ms, a finding we return to later. As in the analysis of scalp electrodes, no Gp x Dev interactions were observed.

Discussion

Experiment 1 demonstrated that HP late L2 learners accurately discriminated and categorized an L2 speech sound that is not used contrastively in their L1 (i.e., /h/). Moreover, their ERP results suggest they did so by recruiting similar neuro-cognitive processes as NSs, including both automatic and attention-driven processes. The IP group, in contrast, had difficulty ignoring within-category acoustic variation, and their ERP responses suggest difficulty engaging automatic stages of L2 phonetic processing. In what follows we discuss the behavioural and ERP results for the NS, HP and IP groups and speculate about the underlying neuro-cognitive processes.

We asked participants to perform an on-line discrimination task so that we could relate their behavioural categorization skills with neuro-cognitive processing. Native-like categorization of L2-specific phonetic contrasts requires that participants not only respond to phonetically-relevant acoustic cues that differentiate two contrasting phonetic categories, but also ignore phonetically-irrelevant acoustic cues that differentiate tokens of the same category. This is thought to be more difficult than simply differentiating two phonetic segments and indexes the extent to which L2 learners have organized sound patterns into L2-specific phonetic categories (Strange & Shafer, 2008). Consistent with this view, we observed no significant group effects, indicating that all groups could discriminate the /h/-/Ø/ and /f/-/Ø/ contrasts. However, a Group x Block x Token interaction revealed that the IP group had difficulty ignoring phonetically-irrelevant acoustic differences between tokens of the same phonetic category for both contrasts (e.g., the length difference between the two “ha” tokens), suggesting that they treated all acoustic variation, phonemic or not, as meaningful. This was not observed for the NS and HP groups, suggesting they perceived both tokens as belonging to the same phonetic category. Arguably, the IP group's discrimination of acoustic differences between speech sounds may be an intermediate step towards developing true L2 phonetic categories, as demonstrated by the HP group (Winkler et al, 1999b).

All groups were more accurate in their discrimination of the /f/-/Ø/ compared to the /h/-/Ø/ contrast. This was somewhat unexpected for the NSs because both contrasts are phonemic in English. However, it corroborates results from other studies showing that some native contrasts are more difficult to discriminate than others (Guion & Pederson, 2007). That /h/ is acoustically non-salient and difficult even for native speakers to perceive, reinforces why /h/-/Ø/ is a difficult contrast for L2 learners to acquire (Polka, 1992).

As for the ERP results, between 100-175 ms both the NS and HP groups exhibited a significant negativity for both the /h/-/Ø/ and /f/-/Ø/ contrasts. In contrast, no significant negativity was observed for the IP group in response to the /h/-/Ø/ contrast. These findings are similar to those reported by Winkler et al. (1999a) who argue that early automatic stages of L2 phonetic discrimination are available to L2 learners who have attained high L2 proficiency; although caution must be exercised in interpreting group differences in the present study because no interaction with the factor Group was observed in the global ANOVA. In the absence of such an interaction, all groups must, therefore, be considered as showing a discrimination effect (reflected by the shared main effects of Dev; see Nieuwenhuis, Forstmann & Wagenmakers, 2011, for a discussion). The findings here indicate that, even in the presence of within category variation, late L2 learners can discriminate non-native phonetic contrasts using automatic mechanisms. That the ERP pattern was clearer for the HP, compared to the IP group, may reflect the IP group's apparent difficulties in ignoring within-category acoustic variation, as shown by their behavioural responses.

The negativity observed between 100-175 ms across all groups and both contrasts shared similar topographical distributions as effects observed between 175-250 ms. This raises the question of whether these effects are in fact distinct ERP components and, in particular, whether the early negativity is truly a MMN, or if it reflects a combination of MMN and N2b effects. This is important because it raises questions whether the negativity observed for the HP L2 learners reflects automatic processing alone, or whether it includes contributions from later attention-driven stages of processing (i.e., attention orientation that is thought to be reflected by the N2b). Indeed, for both time windows, the negativity displayed by all groups and for both contrasts was largest at posterior rather than frontal electrodes, which is more typical for N2b than MMN effects (Novak, Ritter & Vaughan, 1992). Moreover, although we observed a polarity reversal at mastoid electrodes between 100-175 ms (which is thought to index the MMN, independent of N2b effects; Novak et al., 1990), this positivity was small and did not reach significance. Thus, at first glance, both negativities appear to resemble N2b rather than MMN effects.

However, this distribution is also consistent with MMN effects reported previously. Other studies that have used consonant-vowel (CV) syllables or CVC words/pseudo-words as stimuli, rather than individual vowels, have also reported central-posterior MMN effects (Kazanina et al., 2006; Mah, 2011; Shytrov, Hauk & Pulvermüller, 2004). Posteriorly-distributed MMNs may reflect automatic activation of lexical (in addition to phonetic) memory traces, which are not activated when vowels are used as stimuli (Shytrov et al.). Moreover, not all MMN effects occur concurrently with a mastoid positivity (e.g., Näätänen et al., 1993), and when they are observed together, they may reflect different levels of stimulus processing. Sussman and Winkler (2001) suggested that mastoid positivities reflect activity in the auditory cortex and the detection of purely acoustic aspects of a contrast, whereas activity at scalp electrodes may reflect other cognitive processes, such as phonetic categorization, for example (Sussman, Kujala, Halmetoja, Lyytinen, Alku & Näätänen 2004). Thus, the small mastoid positivity observed here may suggest that the auditory cortex is not exclusively responsible for detecting phonetic category deviants in paradigms that control for acoustic features of the standard and deviant stimuli. This explanation warrants further investigation, but may also account for the absence of a mastoid positivity in some other studies that have used a categorical odd-ball paradigm (e.g., Molnar et al., 2009; but see Dehaene-Lambertz et al., 2000). Finally, although topographical distributions of the negativities observed between 100-175 ms and 175- 250 ms are similar, their temporal characteristics clearly differ. Two distinct and successive negative peaks, separated by approximately 75 ms, can be seen in the ERP difference waves of the NS and HP groups for both contrasts. The timing of these peaks is highly consistent with the chronometry of MMN and N2b effects, respectively, reported in previous studies, in which the output of the MMN is thought to trigger attention-orienting N2b effects (Novak et al., 1992). In short, it seems unlikely that spatial and temporal overlap of MMN and N2b components accounts for the early negativity observed here.

That being said, it is still possible that attention could have contributed to the HP group's negativity between 100-175 ms by increasing the saliency of the stimuli’s phonetic features before they entered into the MMN’s automatic deviance detection process. Forming a sensory memory trace of the “standard” stimulus involves organizing and grouping sounds in order to establish a sound context. When multiple groupings are possible, such as in a categorical oddball paradigm in which physically different tokens can be grouped into one or more phonetic categories, then directing attention to these groupings can influence the sound context against which the deviant is compared, thereby modulating the MMN response. In other words, attention can modulate the automatic process of detecting a deviant by influencing the encoding of the standard stimulus (Sussman, 2007). Indeed, there is evidence from naïve subjects listening to non-native phonetic contrasts (Hisagi et al., 2010) and from children listening to difficult-to-discriminate tone contrasts (Gomes, Molholm, Ritter, Kurtzberg, Cowan & Vaughan, 2000) that the MMN amplitude is larger when participants attend to stimuli rather them ignore them. Thus, the negativity observed within the MMN time window in the present study might indicate that actively attending to L2 speech sounds facilitated the HP group's ability to classify them and detect a deviant category. Future studies should compare the MMN effect exhibited by HP L2 learners when they are actively or passively listening to L2 phonetic contrasts in a categorical odd-ball paradigm to further investigate how attention influences L2 phonetic categorization ability. To our knowledge no such MMN study currently exists. In Experiment 2, we address this by investigating whether HP L2 learners process L2 phonetic information automatically when engaged in a different level of linguistic analysis.

Between 175-250 ms and 300-500 ms (i.e., the N2b and P3b time windows, respectively) all three groups exhibited similar and significant effects for both phonetic contrasts. Thus, as expected, later stages of L2 phonetic processing that involve attention (i.e., attention orientation, working memory updating, response preparation) are similar in IP and HP L2 learners and in NSs. These effects are noteworthy, particularly for the IP group, given that we examined the processing of an English-specific phonetic contrast (/h-Ø/) that is notoriously difficult for native French speakers to acquire. Our results indicate that, when tested with simple syllable stimuli and when conducting a task that focuses attention on phonetic information, even relatively low proficiency L2 learners can accurately discriminate this non-native contrast by using attention-driven processes that are similar to those used by NSs.

In summary, Experiment 1 demonstrates that late IP and HP L2 learners accurately discriminate native and non-native phonetic contrasts using the same attention-driven processing mechanisms as NSs. HP L2 learners also engaged automatic processes; however, this may have been facilitated by directing attention to the phonetic features of the stimuli. In Experiment 2 we investigated whether the same L2 learners use automatic L2 phonetic processes when presented with contrasts embedded in a more complex lexical context that required them to engage in semantic analysis of the stimuli.

Experiment 2

Experiment 2 sought to examine whether the same L2 learners who had participated in Experiment 1 could discriminate non-native contrasts presented in a lexical context and when engaged in a task that directed attention away from phonetic-level analysis and, more specifically, required semantic analysis. During spoken word comprehension, phonetic processing must be robust and automatic so that listeners can engage in semantic analysis of the stimuli (Strange, 2011; Strange & Shafer, 2008). The ASP model predicts that, in these situations, without fully automatic L2 phonetic processing routines, late L2 learners will fail to discriminate non-native phonetic contrasts. Consequently, they may perceive minimal pairs and subtle mispronunciations (pseudo-words) that involve non-native phonetic contrasts (e.g., hair-air; happy-‘appy) as identical, even though they can discriminate the same contrasts in simpler stimuli and when task demands permit them to focus attention on phonetic information. In Experiment 2, we investigate whether the HP L2 learners from Experiment 1, who showed evidence of automatic L2 phonetic processing, and the IP L2 learners, who did not, would show evidence of automatic processing when they were engaged in semantic analysis of stimuli.

With this goal in mind, another ERP component, the N400, was examined. The N400 is a negative-going wave that peaks approximately 400 ms after the visual or auditory presentation of a word or word-like string (e.g., Kutas & Hillyard, 1984) and is a widely replicated and reliable index of lexical representation, activation, and access (Lau, Phillips & Poeppel, 2008). For example, the N400 elicited by low frequency words is larger than the N400 elicited by high frequency words, reflecting the degree of difficulty accessing the corresponding representations from lexical/semantic memory (Young & Rugg, 1992). The N400 elicited by pseudo-words (orthographically legal and pronounceable word-like strings) is even larger in amplitude and longer in duration than the N400 elicited by words and is thought to reflect a prolonged search through memory for potentially low-frequency and low-familiarity words that may match the word-like stimulus (Bentin, 1987; Holcomb & Neville, 1990; Holcomb, Grainger & O'Rouke, 2002). If, however, pseudo-words differ only with respect to a difficult to discriminate non-native contrast, then both might activate the same lexical representation and elicit a similar N400 response (Friedrich, Lahiri & Eulitz, 2008). Thus, the "N400 pseudo-word effect" characteristic of native speakers – a larger N400 elicited by pseudo-words compared to words – can be used to examine whether L2 learners can discriminate native and non-native phonetic contrasts in a lexical context.

To our knowledge, only one previous study has used the N400 pseudo-word effect to examine L2 phonetic processing; however, its findings are unclear. Sebastian-Gallés, Rodríguez-Fornells, Diego-Balaguer and Díaz (2006) compared Catalan- and Spanish-dominant bilinguals during processing of a vowel contrast that is phonemic in Catalan and allophonic in Spanish, /e/-/E/. Participants listened to Catalan words and pseudo-words that were derived by exchanging the vowels /e/ and /E/, while performing a lexical decision task. The authors predicted that only the Catalan-dominant group would successfully discriminate the pseudo-words from the words and that this would be reflected in their lexical decisions and N400 responses. However, the results only partially supported their predictions. First, no N400 pseudo-word effects were observed; for both groups, words and pseudo-words elicited similar N400 responses. The Catalan-dominant group, however, was more successful on the lexical decision task, and when they made an error (e.g., misclassified a pseudo-word as a word) they elicited an ERN effect (an Error Related Negativity), suggesting they had realized their mistake. No ERN response was observed for the Spanish-dominant group, suggesting they had not detected their incorrect lexical decisions.

Even though neither group displayed a N400 pseudo-word effect, the authors argued that the ERN and lexical decision results together indicate that only the Catalan-dominant bilinguals could discriminate the /e/-/E/ contrast. For this group, a lack of a N400 pseudo-word effect was interpreted to mean that they had established separate phonological variants for each word, one variant was activated by the words spoken with correct pronunciations and the other by the words spoken with a Spanish accent (i.e., the pseudo-words), resulting in no N400 pseudo-word effect. For the Spanish-dominant bilinguals, however, because they did not elicit a ERN and displayed poor lexical decision performance, the lack of N400 pseudo-word effect was interpreted to mean that they could not discriminate the contrast – words and pseudo-words were treated functionally as homophones, both activating the same underlying lexical representation and, in turn, no N400 pseudo-word effect was observed. Thus, the absence of an N400-effect, they argued, reflected different underlying representations in the two groups.

On the surface, this conclusion appears to be compatible with the predictions of the ASP model – discriminating L2 phonetic contrasts in a lexical context is impaired. However, some issues remain. First, in the Sebastian-Gallés et al. study, conclusions were based on effects occurring at relatively late stages of lexical processing (i.e., lexical decision performance and ERN responses), rather than N400 effects. These reflect post-lexical processing that may be influenced by a variety of non-linguistic factors that are largely independent of lexical representation and access (see, for example, the Bilingual Interactive Activation plus model, BIA+; Dijkstra & Van Heuven, 2002). Second, the null N400 effects could reflect the fact that approximately half of the words in this study were Spanish-Catalan cognates[13]. Thus, it is possible that upon hearing the Catalan words and pseudo-words, both Catalan and Spanish lexical items were activated and competed for processing. For the Spanish-dominant group, cross-language

---

[13] This is in contrast to a control condition that was also included in this study that involved a phonetic contrast that is phonemic in both Spanish and Catalan (/i-u/). For this condition only 25% of the words were Spanish-Catalan cognates. A significant N400 effect was observed here for both groups.

interference from their L1 could explain the poor lexical decision performance and null N400 and ERN effects (Marian & Spivey, 2003; Thierry & Wu, 2004; Weber & Cutler, 2004). As the Catalan-dominant bilinguals were also very proficient in Spanish, L2 activation might also explain the null N400-effects in this group as well (Libben & Titone, 2009; Titone Libben, Mercier, Whitford & Pivneva, 2012). Finally, because the lexical decision task required participants to orient their attention to the stimuli's phonetic form, this study does not directly test the ASP model's prediction that discrimination will be most difficulty when participants are performing a task that directs attention away from phonetic analysis.

The same participants from Experiment 1 participated in Experiment 2 in which they heard English words and pseudo-words that were created by manipulating the /h/-/Ø/ and /f/-/Ø/ contrasts (e.g., by removing the /h/ or /f/ from h- or f-initial words and by adding an /h/ or /f/ to vowel-initial words; **Table 1**). To address some of the issues raised above, we limited the possibility that N400 effects could reflect cross-language lexical activation by excluding French-English cognates or inter-lingual homophones from the stimuli and included a monolingual native English speaker control group. To ensure that attention was directed away from the phonetic features of the stimuli, participants performed a semantic categorization task ("press a button when you hear a fruit or vegetable word"), and they were not told about the pseudo-words or phonetic contrasts. We predicted that native English speakers would differentiate the /h/-/Ø/ and /f/-/Ø/ contrasts and, thus, would display a relatively large N400 for the pseudo-words compared to the words. As for the L2 speakers, we expected both the IP and HP groups to differentiate between words and pseudo-words involving the native /f/-/Ø/ contrast and to exhibit an N400 effect. However, based on the predictions of the ASP model and the results of Experiment 1, we expected the IP group would have difficulty differentiating the /h/-/Ø/ contrast and, thus, would not display a N400 pseudo-word effect in this condition, whereas the HP group would accurately discriminate the /h/-/Ø/ contrast and exhibit a N400 pseudo-word effect. Our predictions were based on the assumption that the HP, in contrast to IP, L2 learners would have acquired automatic L2 phonetic discrimination skills that would allow them to process a word's phonetic features while attending to its meaning.

Method

Participants. The same participants took part in Experiment 1 and 2.

Stimuli. The test stimuli consisted of 60 /h/- initial, 60 /f/- initial, and 90 vowel- initial words[14]; 35% of the vowel-initial words began with the letter “e”; 26% began with the vowel “a” (26%), “i” (13.3%), “o” (11.7%) or “u” (13.3%). There were also 240 pseudo-words that were derived from the words by removing the initial consonant from the /h/- and /f/- initial words and by adding an /h/ or /f/ to vowel initial words (see **Table 1** and **Appendix 1**). The words varied in length from one to four syllables and were 688 ms (SD = 159 ms) in duration, on average. Only high frequency content words that would be familiar to beginning L2 learners of English were included; French-English cognates and inter-lingual homophones were excluded. Pseudo-words that sounded like a word or name in

---

[14] Thirty of the vowel-initial words served in both the /f/ and /h/ conditions.

French or English were also excluded (e.g., head- Ed) to avoid spurious lexical activations.

From this initial corpus, two lists were created that each contained 90 words (30 /h/-initial, 30 /f/-initial, and 30 vowel-initial) and 120 pseudo-words (30 missing /h/, 30 missing /f/, 30 with an additional /h/, and 30 with an additional /f/). The lists were counterbalanced so that each item was presented as a word in one list and a pseudo-word in the other and no participant heard a given item (as a word or derived pseudo-word) more than once. The /h/, /f/ and vowel-initial words and pseudo-words were distributed in the two lists so that the lists were equivalent in terms of word frequency, phonological neighbourhood density, word and pseudo-word duration, as well as frequency of occurrence of the first vowel of the vowel-initial words and /h/-less and /f/-less pseudo-words. Phonological neighbourhood density and word frequency measures were obtained from the English Lexicon Project Database (Balota et al., 2007). Word frequency was based on the Kučera and Francis (1967) and Hyperspace Analogue to Language (Lund & Burgess, 1996) frequency norms. A multivariate ANOVA using word type (h-, f- or vowel-initial) and list (1, 2) as fixed factors and the two measures of word frequency as dependent variables, revealed no significant effects or interactions, verifying that word frequency was similar for all word types within each list and between lists ($ps > .10$). A univariate ANOVA investigating phonological neighbourhood density (again using word type and list as factors) revealed that, overall, vowel-initial words had fewer phonological neighbours (M = 2.0) than h-initial (M=7.1) and f-initial (M = 6.0) words [$F(2, 166) = 9.43$, $p <$

.001]. However, importantly, there were no differences between items in the two lists (*p* > .10). Because the pseudo-words for one list were derived from the words of the other list, this ensured that any N400 differences that emerged between words and pseudo-words would be a result of word/pseudo-word status rather than possible word frequency or phonological neighbourhood density differences. Finally, to compare item duration, a univariate ANOVA was conducted using lexicality (words, pseudo-words), initial letter (h, f, vowel) and list as factors. The 3-way interaction was not significant (*p* > .10), indicating that for each condition, item duration did not differ between lists. However, as expected, this also revealed that vowel-initial items were shorter than h- and f- initial items [*F*(2, 166) = 9.43, *p* < .001]; and, overall, words were also approximately 50 ms shorter than pseudo-words [655 ms vs. 711 ms; *F*(2, 166) = 9.43, *p* < .001] see "ERP analysis" section for how we controlled for these differences.

Stimuli were presented in a semantic priming paradigm (see procedure) such that, in both lists, each word and pseudo-word was followed by an English word that was semantically related to the word (or word from which the pseudo-word was derived; 210 in total). These words were presented twice in each list, once following the words and pseudo-words described above, and once following semantically unrelated words. These items are not reported here and will be described in more detail in a future paper. For the purpose of the present Experiment, this means that the critical word and pseudo-words were presented with a large number of other words, decreasing the likelihood that participants were aware of the contrasts. An additional 43 word pairs were also included in each list that were the “targets” to which participants responded in the semantic categorization task. One word in each pair was a fruit or vegetable word (e.g., raspberry) and the other word was a semantically-unrelated high frequency word (e.g., chair). ERP responses to these words are not reported here.

All stimuli were recorded by a female native Canadian English speaker in a sound attenuating chamber using a digital recorder. She read the stimuli from a list and was instructed to pronounce both words and pseudo-words naturally. The stimuli were transferred to a computer and edited into individual tokens using Cool Edit 1996 software package (Syntrillium Software; Phoenix, AZ). A 15 ms silence was left at the beginning and end of each word. All stimuli were then normalized to 70 dB using Praat software package version 5.2.35 (Boersma & Weenink, 2009). The order of stimuli within each list was randomized and equally distributed into 6 presentation blocks of approximately 9 minutes each. Block and item order was reversed for half of the participants (i.e., they were presented with mirror images of the original lists).

Procedure. Participants completed Experiment 1 and 2 in the same testing room, seated in front of the computer monitor that displayed the stimuli. Experiment 2 was actually conducted before Experiment 1 so that participants would not be aware of the /h/-/Ø/ or /f/-/Ø/ contrasts before testing for Experiment 2 began. They were told, with oral and written English instructions, that they would hear a series of English words presented in pairs. They were asked to listen and respond by pressing the left mouse button when they heard a fruit or a vegetable word. They were not explicitly told about the pseudo-words before testing and were not required to perform any task with these items.

The experiment began with the presentation of 9 practice word/pseudo-word pairs, followed by a short break in which they could ask questions. As in Experiment 1, stimuli were presented binaurally at 70 dB via inserted headphones (Etymotic Reseach) that were set to a comfortable volume by each participant. Each trial began with the presentation of a fixation cross that remained in the centre of the computer screen for the duration of the trial. One second later, a word or pseudo-word was presented. This was followed by an inter-stimulus interval of 600 ms and the second word in the pair. The symbol "???" appeared on the screen one second after the offset of the second word, prompting participants to respond. After 2 seconds this was replaced by the symbol "--", prompting participants to blink their eyes if necessary. This prompt disappeared after 2 seconds and was replaced by the fixation cross, signalling the start of the next trial. Each presentation block began with the presentation of 3 word pairs that were not analyzed to ensure that participants were settled into the experiment before presenting stimuli that entered into data analysis. The experiment lasted approximately 1- 1 1/2 hours, including short breaks between blocks.

EEG Recordings and ERP Analysis. EEG recording and ERP pre-processing (e.g., mastoid re-referencing, filtration, rejection) procedures were the same as in Experiment 1. ERPs were computed for each participant in response to 4 conditions: ***H word*** (average response to /h/-initial and vowel-initial words), ***H pseudo-word*** (average response to /h/-less- and /h/-plus-vowel pseudo-words), ***F***

word (average response to /f/-initial and vowel-initial words), ***F pseudo-word*** (average response to /f/-less- and /f/-plus-vowel pseudo-words; see **Table 1** for examples. By collapsing across sub-conditions, we were able to control for (1) length differences between words and pseudo-words and between vowel- and consonant-initial items and (2) for the initial letter of the items. These factors could have influenced the latency and amplitude of the N400 as well as earlier ERP components, (e.g., the N1-P2 complex) that may have carried into the N400 time window. Thus, we can be more confident that any effects reported below are likely due to lexical status. It is also theoretically motivated insofar as production data show that native French speakers tend to both delete /h/ from /h/-initial words and hypercorrect by adding an /h/ to vowel-initial words (Janda & Auger, 1992).

Participants were included in ERP analyses if they contributed a minimum of 30 artifact-free trials in each of these 4 conditions. For each participant and condition, artifact-free ERP responses were time-locked to the onset of the first word of each pair. This was done for a 1400 ms interval, including a 100 ms pre-stimulus baseline interval. Statistical analyses were conducted on the mean amplitude of the ERP waveforms within two time windows – 500-700 ms and 700-900 ms – which were selected based on previous literature of N400 pseudo-word effects (e.g., Friedrich, Lahiri & Eulitz, 2008; Holcomb & Neville, 1990) and visual inspection of the grand averages. As in Experiment 1, 12 lateral (F3, F4, F7, F8, C3, C4, T7, T8, P3, P4, P7, P8) and 3 midline (Fz, Cz, Pz) electrodes were selected for analysis, as effects were most prominent at these sites. For each time window, repeated-measures ANOVAs were performed on lateral and midline sites separately. For the lateral electrodes the between-subjects factors were Group (Gp: NS, HP and IP) and the within-subjects factors were: Contrast (/h/-/Ø/ or /f-Ø/), Lexical Status (word or pseudo-word), Anterior-Posterior (anterior, central, parietal), and Hemisphere (Left or Right). For the midline sites, the factors were: Gp, Con, Lex and Anterior-Posterior (anterior, central, parietal). Results are reported for main effects and interactions that involve the factor Lex; results of midline analyses are reported when they yielded results not revealed in the analyses of lateral electrodes. Greenhouse-Geisser correction was applied to all analyses involving the Ant-Post factor and corrected *p* values are reported.

Results

Behavioural Results. Accuracy in the semantic categorization task was defined as correct identification of a fruit or vegetable word within the word pair in which it was presented. Accuracy scores (calculated as percentage of correct hits) were analyzed using a univariate ANOVA with Gp (NS, HP, IP) as a between subjects factor. There was no significant Gp difference [*F*(2, 39) = 2.28, *p* > .10], indicating similar performance by all three groups (NS: *M* = 80.31 % *SD* = 10.73; HP: *M* = 84.29 % *SD* = 12.02; LP: *M* = 71.01 % *SD* = 23.63). The rather low and variable performance by the IP group in particular may reflect the fact that the word frequency (as measured by Hyperspace Analogue to Language) of the fruit and vegetable items was significantly lower than that of the h- f- and vowel-initial words [*F*(3, 210) = 2.93, p < .05]. However, performance was clearly above chance for all groups, indicating that participants were attending to the words in the experiment.

ERP Results. The ERP grand mean waveforms for the NS, HP and IP groups for the /h/-/Ø/ and /f/-/Ø/ contrasts are presented in **Figures 4 to 6**. For all ERPs, a negativity peaking at approximately 150 ms after stimulus onset (N1) and a positivity at approximately 240 ms (P2) can be seen followed by a negativity starting around 350 ms (N4). As can be seen in **Figures 4 to 6,** the latter negativity persisted longer for pseudo-words compared to words in all cases except the IP group in response to the /h/-/Ø/ contrast. This enhanced negativity will henceforth be referred to as the N400 pseudo-word effect. Statistical analyses of this effect in the 500-700 ms and 700-900 ms time windows are described below.

Insert Figures 4 - 6 about here

A repeated measures ANOVA on the ERP data elicited at lateral electrodes between 500-700 ms revealed a main effect of Lex [$F(1,39) = 7.10$, $p < .05$], indicating a larger negativity in response to pseudo-words compared to words across all electrode sites. However, this was qualified by a highly significant Lex x Gp interaction [$F(2,39) = 9.05$, $p \leq .001$], pointing to significant effects for the NS [$F(1,14) = 12.99$, $p < .005$] and HP [$F(1,11) = 5.68$, $p < .05$] groups only. For the IP group, the effect of Lex did not reach significance [$F(1,14) = 3.77$, $p = .075$].

Between 700-900 ms, a main effect of Lex was again observed [$F(1,39) = 24.42$, $p < .001$], indicating a broadly distributed N400 pseudo-word effect.

However, this was qualified by a number of interactions involving the factors Con and Gp, pointing to differences between groups in the effects elicited by the /h/-/Ø/ and /f/-/Ø/ contrasts (see **Table 5** for values from the global ANOVA). Most importantly, a Con x Gp x AP interaction was observed, suggesting group differences at posterior sites, as expected for N400 effects. Therefore, follow-up analyses were conducted for each group and contrast separately. For the NSs, as expected, a main effect of Lex was observed for both /f/-/Ø/ [$F(1,14) = 4.88$, $p < .05$] and /h/-/Ø/ [$F(1,14) = 7.43$, $p < .05$]. As can be seen in **Figure 5,** similar effects were also observed for the HP group; specifically, broadly distributed and highly significant N400 pseudo-word effects for both /f/-/Ø/ [$F(1,11) = 42.74$, $p < .001$] and /h/-/Ø/ [$F(1,11) = 15.61$, $p < .005$]. As well, for the /f/-/Ø/ contrast, a significant Lex x Lat interaction [$F(1,11) = 7.16$, $p < .05$] indicated a somewhat larger effect over medial [$F(1,11) = 40.69$, $p < .001$] than lateral [$F(1,11) = 32,62$, $p < .001$] electrodes, consistent with N400 effects reported by others. The IP group, in contrast, displayed a small N400 pseudo-word effect for the /f/-/Ø/ but not for the /h/-/Ø/ contrast, as seen in **Figure 6**. For the IP group, statistical analysis conducted on the ERP effects for /h/-/Ø/ revealed no effects or interactions with the factor Lex, indicating similar N400 responses for both words and pseudo-words (i.e., no N400 pseudo-word effect). For /f-Ø/, Lex x Lat [$F(1,14) = 7.59$, $p < .05$] and Lex x AP [$F(2,13) = 8.50$, $p < .005$] interactions indicate a N400 pseudo-word effect that was significant at posterior medial electrodes [$F(1,14) = 5.55$, $p < .05$][15]. Similarly, at midline electrodes, a

---

[15] At posterior electrodes a marginally significant negativity was observed [lex: $F(1,14) = 4.2$, $p = .06$] in addition to a lex x lat interaction [$F(1, 14) = 9.17$, $p < .01$] and at medial sites a marginal

significant negativity was observed at electrode Pz [*F*(1,14) = 5.61, *p* < .05; Lex x AP: *F*(2,13) = 1.77, *p* = .081] for /f-Ø/, while no significant effects were observed for /h/-/Ø/ [*ps* > .10].

Insert Table 5 about here

The Lex x Con x AP x Gp interaction was also followed up by separate analyses comparing the NS and HP groups and the NS and IP groups directly. As before, these analyses revealed similar N400 pseudo-word effects for the NS and HP groups, indicating native-like effects for the HP group. In contrast, comparing the NS and IP groups revealed a significant Lex x Con x AP x Gp interaction [*F*(2,27) = 3.75, *p* < .05], pointing to a significant group difference for the /h/-/Ø/ contrast at posterior electrodes [Lex x Group: *F*(1,28) = 5.21, *p* < .05] but not for the /f/-/Ø/ contrast [Lex x Group: *F* < 1]. This suggests the IP group engaged in native-like processing for the /f/-/Ø/ but not the /h/-/Ø/ contrast, whereas for the HP group, native-like processing was observed for both contrasts.

Correlation Analyses. To investigate the relationship between L2 oral proficiency and the N400 pseudo-word effect further, we correlated the amplitude of the N400 difference wave for all L2 learners in response to the /h/-/Ø/ and /f/-/Ø/ contrasts (at electrode Pz within the 700-900 ms time window) with the 6 proficiency ratings (i.e., general pronunciation, pronunciation of H words, vocabulary, grammar, fluency and overall impression; averaged across both

---

main effect was observed [*F*(1,14) = 3.31, *p* = .09], in addition to a lex x AP interaction [*F*(2,13) = 7.14, p ≤ .001]. Both interactions pointed to a significant posterior medial negativity.

raters). The amplitude of the N400 in response to the /h/-/Ø/ contrast was significantly correlated with most aspects of oral L2 proficiency assessed (using a p value corrected for multiple comparisons). Participants with a large N400 pseudo-word effect for the /h/-/Ø/ contrast were those who received high ratings on overall pronunciation [$r = -.680, p < .001$], nativelikeness [$r = -.605, p \leq .001$], vocabulary [$r = -.551, p = .003$], pronunciation of "h" words [$r = -.509, p = .007$] and fluency [$r = -.507, p = .007$]. The amplitude of the N400 pseudo-word effect for the control /f/-/Ø/ contrast was not significantly correlated with pronunciation [$r = -.366, p = .060$]; however, it did correlate with grammar [$r = -.529, p \leq .005$]. That there was only one significant correlation for the /f/ contrast is not surprising since this contrast in phonemic in both the L1 and L2 and, thus, may be insensitive to L2 proficiency levels. **Figure 7** displays the relationship between pronunciation ratings and the N400 pseudo-word effect for the /h/-/Ø/ contrast.

Insert Figure 7 about here

Discussion

The main finding of Experiment 2 was that the HP late L2 learners, like the NSs, exhibited a significant N400 pseudo-word effect for both the native /f/-/Ø/ contrast and the non-native /h/-/Ø/ contrast. These results indicate that HP L2 learners can discriminate non-native phonetic contrasts even when engaged in a semantic categorization task that directed their attention away from relevant phonetic information. This, in turn, implies that they have acquired the ability to process non-native phonetic contrasts automatically. In contrast, the IP learners displayed a N400-effect for the native /f/-/Ø/ contrast only, indicating successful discrimination of the native but not the non-native contrast. In contrast to the results of Experiment 1, these results suggest that a high level of L2 proficiency is critical for L2 learners to engage in automatic L2 phonetic processing while engaged in other levels of linguistic analysis.

The results for the HP L2 learners are particularly noteworthy because participants were not told about the pseudo-words or phonetic contrasts before the experiment and were required to listen for word meaning rather than phonetic form, making it unlikely that attention could have facilitated discrimination. Despite this, the HP group's N400 pseudo-word effect was statistically identical for the /h/-/Ø/ and /f/-/Ø/ contrasts and no significant differences were observed in the effects exhibited by the HP or NS groups, together indicating native-like processing for high proficiency L2 learners. That the difference between the HP and IP groups is related to their level of proficiency is underscored by the significant correlation between pronunciation ratings and the amplitude of the N400 pseudo-word effect for the /h/-/Ø/ contrast, suggesting that the mechanisms underlying this response become increasingly available as L2 oral proficiency improves. These findings reinforce the notion that L2 proficiency, rather than an early age of L2 acquisition, may be the necessary prerequisite for native-like automatic phonetic processing. We will return to this point in the general discussion.

In contrast to the HP group, the IP group exhibited a small N400 pseudo-word effect for the /f/-/Ø/ but not the /h/-/Ø/ contrast. One could argue that the lack of effects for the /h/-/Ø/ contrast may indicate that words in the /h/ condition were less frequent or familiar than words in the /f/ condition. However, as discussed in the Methods section, words in both conditions were matched in word frequency. Moreover, if words in the /h/ condition were less familiar, one would expect overall increased N400s for both words and pseudo-words, simply because unfamiliar words would have been equally difficult to retrieve and, thus, processed in a similar way as pseudo-words. Instead, the ERP plots show similar N400 amplitudes for words in both the /f/ and /h/ conditions, and a larger effect for pseudo-words for the /f/ condition only. Thus, it appears that the lack of N400 pseudo-word effects for the /h/-/Ø/ contrast reflects the IP groups' difficulty differentiating between words and pseudo-words that involve a non-native phonetic contrast.

Comparing the results of Experiment 1 and 2, the discrimination of the /h/-/Ø/ contrast during spoken word comprehension may have been difficult for the IP group for two reasons. First, the results of Experiment 1 suggest phonetic processing of non-native phonetic contrasts may have been less automatic than the processing of native contrasts and so attention-guided mechanisms may have been necessary for accurate integration and categorization of L2-specific phonetic information. Consequently, the ability to discriminate L2-specific phonetic categories suffered when attention was directed away from the phonetic features of the stimuli, as found in Experiment 2. Second, the IP group exhibited a rather small N400 pseudo-word effect for even the native /f/-/Ø/ contrast suggesting that L2 words in general (regardless of whether they contain phonetic features that are present in their L1) may be less clearly represented and require more lexical activation for the IP group compared to the HP group. Arguably, the IP group may have needed to focus more attention on lexical-semantic processing in Experiment 2 in order to discern word meaning, leaving less attention available for phonetic-level processing.

General Discussion

In two Experiments we investigated whether late L2 learners who have attained high L2 proficiency process difficult non-native phonetic contrasts using similar neuro-cognitive mechanisms as native speakers. In particular, we examined whether high proficiency late L2 learners, like native speakers, recruit both automatic and attention driven mechanisms during L2 phonetic processing. By comparing processing in different stimulus and task conditions, we sought to examine the conditions under which native-like processing is possible for L2 learners at different stages of L2 proficiency. The most important finding was that high proficiency (HP) L2 learners exhibited a similar neuro-cognitive processing profile as native speakers (NSs) during L2 phonetic processing in both stimulus/task conditions. In particular, the results from Experiment 2 suggest that the HP L2 learners recruited automatic processes even when attention was directed to semantic, rather than phonetic, level analysis. In contrast, the intermediate proficiency (IP) L2 learners were more successful in Experiment 1 when presented with syllables during a task that allowed them to focus attention on phonetic information, compared to Experiment 2 when the stimulus and task demands were more complex and attention was directed away from critical phonetic information. These results show that native-like automatic processes are available to late L2 learners when processing non-native phonetic contrasts and become increasingly available as L2 proficiency improves.

The results from the IP group support a central prediction of the ASP model – discrimination of non-native phonetic contrasts depends on the stimulus and task context in which they are presented. Consistent with the ASP model, the IP group successfully discriminated the non-native /h/-/Ø/ contrast in Experiment 1 when the contrast was presented as syllables in an attended paradigm. However, the same learners had difficulty discriminating the same contrast in Experiment 2 when it was presented as words and pseudo-words during a task that directed attention away from phonetic analysis. Also consistent with the ASP model, the IP group showed clear evidence that they could engage in attention-driven stages of L2 phonetic processing (as evidenced by N2b/P3b responses), but not automatic L2 processing, especially when discrimination was required in parallel with word comprehension (as evidenced in Experiment 2). This supports the notion that difficulties discriminating L2 phonetic contrasts reflect difficulties with automatic stages of phonetic processing in particular, rather than an inability to process L2 contrasts at all. Even in the absence of fully automatic L2 processing routines, L2 phonetic discrimination can be successful in situations that allow L2 learners to focus their attention on the relevant phonetic information.

Although the ASP model accurately predicted the results of the IP group, the results of the HP group were not predicted by the ASP model. Like the NSs, the HP L2 learners showed evidence of engaging both attention-driven and automatic neuro-cognitive processes. This is consistent with the findings from ERP Experiments of L2 grammar processing that report native-like automatic processing in high but not low proficiency late L2 learners (Rossi et al., 2006; Steinhauer et al., 2009; 2012). Automatic processing was particularly noteworthy in Experiment 2, insofar as it occurred even though participants were unaware of the contrasts and they were engaged in a task that required semantic, rather than phonetic, analysis. In contrast to these findings, the ASP model argues that "phonetic perception of non-native contrasts may never become as automatic and robust as perception of native contrasts" (Strange & Shafer, 2008, pg. 185). These findings suggest that the ASP model should be extended to account for the possibility of native-like automatic L2 phonetic by high proficiency L2 learners.

In contrast to the ASP model, our findings suggest that the neuro-cognitive mechanisms that are available to late L2 learners during L2 phonetic discrimination change as L2 proficiency improves. For L2 learners at low levels of L2 proficiency, our results are consistent with the ASP model in its current form. At this stage, L2 learners make use of primarily attention-driven processing mechanisms. This allows them to successfully discriminate non-native contrasts in simple stimulus contexts and during tasks that orient attention to the relevant acoustic-phonetic differences between contrasting phonetic segments. However, they have progressively more difficulty with phonetic discrimination as task and stimulus demands necessitate the use of automatic processes. Our extension of the ASP model differs from the current ASP model in that we argue that it is possible for late L2 learners to demonstrate native-like patterns of discrimination and automatic neuro-cognitive processing, even in cognitively-demanding situations. Moreover, the development of automatic processing, particularly in complex task conditions, coincides with increases in L2 proficiency skills, as evidenced by the HP group's high ratings on the speech sample task.

This conceptualization also views automatic processing itself as falling along a continuum such that L2 learners may show evidence of automatic processing in some situations but not in others. That automaticity falls along a continuum could account for the IP group’s trend towards a MMN in Experiment 1 (as evidenced by the non-significant interaction with the factor group in the global ANOVA), and their clear lack of an N400 pseudo-word effect in Experiment 2. It would be interesting in future research to examine if L2 learners who exhibit an N400 pseudo-word effect in quiet listening environments (as in Experiment 2) are able to do so when presented with a dichotic listening paradigm in which they must simultaneously discriminate contrasts and ignore irrelevant background noise. We predict that very advanced L2 learners would exhibit automatic processing even in such highly demanding situations. Longitudinal Experiments of L2 learners’ ability to discriminate non-native phonetic contrasts in a variety of stimulus and task contexts as they advance in L2 proficiency are necessary to test these predictions and to investigate whether there is an upperlimit to late L2 learners' ability to engage in native-like automatic phonetic discrimination.

Another direction for future research is to investigate the factors that allow L2 learners to engage in automatic L2 phonetic processing in a wider range of situations. One important factor may be current L2 use. In the present Experiment, the HP group reported nearly double the amount of daily L2 use at the time of testing as the IP group. Interestingly, the groups did not differ in the amount of English they reported using during childhood or adolescence. This is consistent with Piske, MacKay and Flege (2001) who report that current L1-L2 usage patterns exert a significant effect on L2 pronunciation skills, independent of the age at which the L2 was acquired. Conversely, this is at odds with the notion of a critical period that restricts high levels of L2 phonological attainment to early learners (Long, 1990). Future Experiments should explore whether high attainment is possible for all individuals given enough L2 exposure, or if it reflects underlying cognitive differences, such as the ability to maintain new phonological sequences in short-term memory (O'Brien, Segalowitz, Collentine & Freed, 2006).

In conclusion, the experiments here show that late L2 learners who have attained high L2 proficiency process non-native phonetic contrasts using similar neuro-cognitive mechanisms as native speakers even when attention is directed away from phonetic form. These experiments also suggest that at intermediate levels of proficiency, L2 learners can successfully discriminate non-native phonetic contrasts when the stimuli and task allows them to use attention-driven processing mechanisms, although they will have difficulty in tasks that necessitate the use of automatic processes. We argue that native-like automatic phonological processing is possible for late L2 learners and becomes increasingly available as L2 proficiency skills improve.

References

Alho, K., Paavilainen, P., Reinikainen, K., Sams, M., Naatanen, R., (1986). Separability of different negative components of the event-related potential associated with auditory stimulus processing. *Psychophysiology, 23,* 613– 623.

Balota, D. A, Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B. *et al.* (2007). English Lexicon Project. *Behavioral Research Methods, 39*(3), 445-459.

Beddor, P. S. & Gottfied, T. L. (1995). Methodological issues in cross-langauge speech perception research with adults. In W. Strange (Ed.), *Speech Perception and linguistic experience: Issues in cross-language research* (pp. 207-232). Timonium, MD: York Press.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language, 31* (2), 308-327.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and linguistic experience: Issues in cross-language research* (pp.191-204). Timonium, MD: York Press.

Birdsong, D. (2006). Age and second language acquisition and processing: a selective overview. *Language Learning, 68* (4), 706-755.

Boersma, P & Weenink, D. 2009. Praat: doing phonetics by computer (Version 5.2. 35) [Computer program]. Retrieved January 3, 2009

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second Language*

Acquisition and Linguistic Theory (pp. 4-63). Oxford and Malden, MA: Blackwell.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience, 12,* 635-647.

Dijkstra, T. & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175-197.

Donchin E. Surprise! Surprise? (1981). *Psychophysiology, 18,* 493-513.

Flege, J. E. 1995. Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech Perception and linguistic experience: Issues in cross-language research* (pp.233-277). Timonium, MD: York Press.

Flege, J. E., & Schmidt, A. M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica, 52,* 90-111.

Friederici, A.D., Steinhauer, K. & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, 99,* 529–34.

Friedrich**,** C. K., Lahiri, A., & Eulitz. C. (2008). Neurophysiological Evidence for Underspecified Lexical Representations: Asymmetries with Word Initial Variations. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1545–1559.

Gomes, H., Molholm, S., Ritter, W., Kurtzberg, D., Cowan, N. & Vaughan, H. G. Jr. (2000). Mismatch negativity in children and adults, and effects of an attended task. *Psychophysiology, 37* (6), 807-816.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention of phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 57-77). Amsterdam: John Benjamins Publishing Company.

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brian Research, 1360,* 89-105.

Hisagi, M. & Strange, W. (2011). Perception of Japanese temporally-cued contrasts by American English listeners. *Language and Speech, 54*(2), 241-264.

Holcomb, P. J. & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5* (4), 281-312.

Holcomb, P. J. Grainger, J. & O'Rouke, T. (2002). An Electrophysiological Study of the Effects of Orthographic Neighborhood Size on Printed Word Perception. *Journal of Cognitive Neuroscience, 14* (6), 938-950.

Janda, R. D., & Auger, J. (1992). Quantitative evidence, qualitative hypercorrection, sociolinguistic variables--And French speakers' 'eadhaches with English h/Ø. *Language & Communication, 12*(3-4), 195-236.

Johnson, J. S. & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

Kazanina, N., Phillips C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *PNAS, 103*(30), 11381-11386.

Kučera, H., & Francis, W. N. (1967). Computational analysis of present-day English. Providence, RI: Brown University Press.

Kuhl, P. K. & Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. In W. Strange (Ed.), *Speech Perception and linguistic experience: Issues in cross-language research* (pp.121-154). Timonium, MD: York Press.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161-163.

LaCharité, D., & Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. In A. Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of BUCLD 23* (pp. 373-385). Somerville: Cascadilla.

Lenneberg, E. H. (1967). *The biological foundations of language.* New York: Wiley.

Libben, M. & Titone, D. (2009). Bilingual language processing in context: Evidence from eye movement recordings during reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35,* 381-390.

Lau, E. F. Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9,* 920-933.

Logan, J. S. & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech Perception and linguistic experience: Issues in cross-language research* (pp. 351-378). Timonium, MD: York Press.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12,* 251-285.

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic space from lexical co-occurence. *Behavioral Research Methods, Instruments & Computers, 28*(2), 203-208.

Mah, J. (2011). Segmental representations in interlanguage grammars: the case of francophones and English /h/. Unpublished doctoral dissertation. Department of Linguisitics, McGill University.

Mah, J., Goad, H., Steinhauer, K. (2006). The trouble with [h]: Evidence from ERPs. In: O'Brien, M.G. (ed.), *GASLA-2006 Proceedings,* 80-87. Somerville, MA, USA: Cascadilla Proceedings Project.

Marian, V. & Spivey, M. (2003). Competing activation in bilingual language processing: within- and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97-115.

Molnar, M., Baum, S. Polka, L. & Steinhauer, K. (2009). Automatic auditory discrimination of vowels in simultaneous bilingual and monolingual speakers as measured by the mismatch negativity (MMN). *Journal of the Acoustical Society of America, 125*(4), 2754-2754.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24* (4), 933-947.

Näätänen, R & Gaillard, A. W. K. (1983). The orienting reflex and the N2 deflection of the event-related potential (ERP). In A. W. K. Gaillard and W. Ritter (Eds.), *Tutorials in event related potential research: Endogenous components* (pp. 119-141). Amsterdam: North Holland.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., Alho, K., (1993). Attention and mismatch negativity. *Psychophysiology, 30,* 436-450.

Näätänen, R. Paavilainen, P. Rinne, T. & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118,* (12), 2544-2590.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432-434.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I. (2001). “Primitive intelligence’’ in the auditory cortex. *Trends Neuroscience, 24,* 283-288.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience, 14,* 1105-1107.

Novak, G., Ritter, W., & Vaughan, H. G. Jr. (1992). The chronometry of attention-modulated processing and automatic mismatch detection. *Psycholphysiology, 29,* 412-430.

Novak, G., Ritter, W., Vaughan, H. G. Jr., Wiznitzer, M. L. (1990). Differentiation of negative event-related potentials in an auditory discrimination task. *Electroencephalography Clinical Neurophysiology, 7,* 255-75.

O'Brien, I., Segalowitz, N., Collentine J., & Freed, B. (2006). Phonological memory and lexical, narrative and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*(3), 377-402.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Endingburgh Inventory. *Neurophysiologica, 9,* 97-113.

Penfield, W. & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, N.J.: Princeton University Press.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12*(6), 1038-1055.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29* (2), 191-215.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics, 52*(1), 37-52.

Ritter, W., Simson, R., & Vaughan, H. G., Jr. (1983). Event-Related Potential Correlates of Two Stages of Information Processing in Physical and Semantic Discrimination Tasks. *Psychophysiology, 20* (2), 168- 179.

Rossi, S., Gugler, M.F., Friederici, A.D. & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*, 2030-48.

Sebastian-Gallés, N. Rodríguez-Fornells, A. Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon, *Journal of Cognitive Neurosceince, 18,* (8), 1277-1291.

Sharma, A., & Dorman, M. F. (2000). Neurophysiological correlates of cross-language phonetic perception. *Journal of Acoustic Society of America, 107,* 2697-2703.

Sharma, A., Kraus, N., McGee, T., Carrell, T., & Nicol, T. (1993). Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalograph and clinical neurophysiology, 88* (1), 64-71.

Shytrov, Y., Hauk, O., & Pulvermüller, F. (2004). Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. *European Journal of Neuroscience, 19,* 1083-1092.

Steinhauer, K., White, E. J. & Genesee, F. (2012). Proficiency and first language transfer effects in late second language acquisition: ERP data from a balanced word category violation design. Unpublished manuscript. McGill University, Montreal.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13-41.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics,* doi:10.1016/j.wocn.2010.09.001

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards, & M.L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153-191). Philadelphia: John Benjamins.

Sussman, E. S. (2007). A new view on the MMN and attention debate: the role of context in processing auditory events. *Journal of Psychophysiology, 21* (3-4), 164-175.

Sussman, E. Kujala, T. Halmetoja, J. Lyytinen, H. Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research, 190,* 128-140.

Sussman, E., Winkler, I., (2001). Dynamic process of sensory updating in the auditory system. Cognitive Brain Research, 12, 431-439.

Tees, R. C., & Werker, J. F. 1984. Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(4), 579-590.

Thierry, G & Wu, Y. J. (2004). Electrophysiological evidence for language interference in late bilinguals. *Neuroreport, 15*(10), 1555-1558.

Titone, D., Libben, M., Mercier, J., Whitford, V., Pivneva, I. (2012). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1412-1431.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4,* 105-122.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50,* 1-25.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37,* 35-44.

White, L., Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12,* 233–265.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999a). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36,* 638-642.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V. et al, (1999b). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7,* 357-369.

Yoshida, K. A. (2004). *Shallow vs. Deep: Bilingual contrast processing.* Unpublished master's thesis. University of British Columbia, Vancouver, BC.

Young, M. P. & Rugg, Michael D. (1992). Word frequency and multiple repetition as determinants of the modulation of event-related potentials in a semantic classification task, *Psychophysiology, 29* (6), 664-676.

Table 1: Example Stimuli used in Experiments 1 and 2

	Experiment 1			**Experiment 2**	
	Standard	Deviant	Example	Word	Pseudo-word
H condition (/h-0/)	ha	a	ha$_1$ ha$_1$ ha$_1$ **ha$_2$ a$_1$** ha$_2$	Happy	‘Appy
	a	ha	a$_1$ a$_2$ a$_1$ a$_1$ **a$_2$ ha$_1$** a$_2$	English	H-english
F condition (/f-0/)	fa	a	fa$_1$ fa$_1$ fa$_1$ **fa$_2$ a$_1$** fa$_2$	Father	‘Ather
	a	fa	a$_2$ a$_2$ a$_1$ a$_1$ **a$_2$ fa$_1$** a$_2$	Answer	F-answer

In Experiment 1 the non-native (/h/-/Ø/) and native (/f/-/Ø/) contrasts were presented as syllables in an attended categorical odd-ball paradigm. In this design each stimulus serves as the standard in one stimulus block and the deviant in another and multiple tokens of each syllable were presented randomly and equally in each block (identified here by 1 and 2). This design decreases the reliance on acoustic-level level processing and requires participants to group stimuli into distinct L2-specific phonetic categories.

In Experiment 2 the same contrasts were presented as high frequency content words and derived pseudo-words. The pseudo-words were created by adding a /h/ or /f/ to the start of a vowel initial word or by removing initial the /h/ or /f/ from h- or f- initial words. Participants conducted a semantic categorization task (i.e., press a button when you hear a fruit or vegetable word) and were not told about the contrasts or pseudo-words before testing began.

Table 2: Participant Information for the High and Intermediate Proficiency L2 Speakers

	High Proficiency (HP) n=12		**Intermediate Proficiency (IP)** n=15	
	M	SD	M	SD
Age at Time of Testing[a]	27.3	4.8	25.7	3.0
Age of Acquisition in years (AoA)	19.0	3.7	18.9	3.2
Years of English use	8.3	4.3	6.8	5.0
Childhood English exposure (% of reported daily language use)				
Age 0-4 home	2.5	5.8	0.1	0.5
Age 0-4 school	0.8	2.9	0.0	0.0
Age 5-11 home	2.9	6.2	2.2	5.6
Age 5-11 school	5.4	5.0	4.2	4.6
Age 12-14 home	4.6	6.2	2.3	5.6
Age 12-14 school	10.4	5.4	12.0	5.9
Age 15-18 home	15.4	17.8	9.7	25.9
Age 15-18 school	29.2	22.7	15.8	9.3
Current daily English use (%) with:				
Friends	34.6	10.8	15.1	21.7
Family	7.1	16.0	6.4	21.8
School/Work	51.7	30.0	28.1	23.6
Spare Time	44.6	21.1	33.3	30.3
TV/Radio	62.9	25.4	43.3	35.5
Reading	62.5	22.0	25.0	26.8
Overall/Total**	**44.6**	7.9	**24.3**	18.7
Self Rate English abilities (7-point scale)				
Pronunciation	5.3	0.8	3.5	1.5
Fluency	5.5	0.9	3.9	1.8
Listening	6.2	0.6	4.7	1.4
Reading	6.5	0.5	5.1	1.6
Vocabulary	5.6	0.8	3.7	1.7
Grammar	5.4	0.7	3.9	1.7
Overall/Total**	**5.8**	0.5	**4.1**	1.5
Cloze Test of English Proficiency (%)*	**89.1**	8.3	**71.4**	23.8
Average Speech Sample Ratings (13 point scale)				
General Pronunciation	10.6	0.7	6.6	1.7
Pronunciation H words	11.9	1.3	7.8	3.2
Fluency	11.7	1.3	8.2	2.5
Vocabulary	11.7	0.5	8.1	2.1
Grammar	11.8	0.7	7.6	2.5
Overall Impression	11.0	0.8	6.7	1.9
Overall/Total***	**11.5**	0.6	**7.5**	1.8

[a] The Native speakers were younger than both groups of L2 learners, who did not differ from each other [*M* = 22.6 years, *SD* = 4.3; *F*(1,2) 4.90, p < .05].
AoA is defined as age of first regular and intensive L2 exposure (c.f. White & Genesee, 1996).
Years of English exposure is defined as current age minus AoA.
$^{*}p < .05$; $^{**}p < .005$; $^{***}p < .001$

Table 3: Accuracy in detecting deviant stimuli in Experiment 1

	Standard	Deviant	**Native Speakers**		**High Proficiency**		**Intermediate Proficiency**	
H condition			*M*	*SD*	*M*	*SD*	*M*	*SD*
	ha	a	96.1	3.0	96.8	3.1	97.7	2.3
(/h-0/)	a	ha	65.1	36.3	70.5	35.6	61.5	37.1
F condition	fa	a	94.0	4.7	96.2	3.1	95.6	7.1
(/f-0/)	a	fa	89.5	11.5	92.5	11.1	80.7	26.6

`

Table 4: Global ANOVA for ERP effects in Experiment 1

		100-175 ms	175-250 ms	300-500 ms
	df	*F value*	*F value*	*F value*
Lateral Electrodes:				
Dev	1, 39	19.81****	18.35****	47.91****
Dev x Hemi	1, 39	1.54	> 1	17.30****
Dev x AP	2, 38	18.82****	40.14****	47.11****
Dev x AP x Hemi	2, 38	2.60	1.60	32.70****
Dev x Hemi x Lat	1, 39	> 1	> 1	32.65****
Dev x AP x Lat	2, 38	2.01	> 1	36.44****
Dev x AP x Hemi x Lat	2, 38	> 1	> 1	30.81****
Dev x Cont	1, 39	8.16**	3.72^{+}	> 1
Dev x Cont x Lat	1, 39	2.93^{+}	> 1	> 1
Dev x Cont x Hemi	1, 39	> 1	4.51*	> 1
Dev x Cont x AP x Hemi	2, 38	1.22	1.03	11.26****
Dev x Cont x AP x Lat	2, 38	> 1	2.11^{+}	4.07*
Dev x Cont x Hemi x Lat	1, 39	> 1	4.70*	2.50
Dev x Cont x AP x Hemi x Lat	2, 38	> 1	> 1	3.19^{+}
Dev x Cont x Hemi x Group	2, 39	> 1	> 1	2.55^{+}
Midline Electrodes:				
Dev	1, 39	12.67****	8.22**	56.46****
Dev x AP	2, 38	12.66****	23.52****	27.88****
Dev x Cont	1, 39	6.74*	2.16	> 1
Dev x Cont x AP	1, 39	1.58	1.05	2.17^{+}

Dev = Deviant (Deviant vs. Standard), Cont = Contrast (/h-0/ vs. /f-0/), Lat = laterality, Hemi = hemisphere, AP= anterior-posterior; Group (NS, HP, IP)
* $p < .05$, ** $p < .01$, *** $p <.005$, **** $p < .001$

Table 5: Global ANOVA for N400 pseudo-word effects within the 700-900 ms time window

	df	*F value*
Lateral Electrodes:		
Lex	1, 39	24.42****
Lex x Lat	1, 39	11.46***
Lex x AP	2, 38	4.81^{+}
Lex x Lat x Cont	1, 39	5.73*
Lex x Cont x AP x Group	4, 78	1.92*
Lex x Cont x Hemi x Group	2, 39	2.65^{+}
Lex x Cont x Lat x Hemi x Group	2, 39	2.58^{+}
Midline Electrodes:		
Lex	1, 39	25.45****
Lex x AP	2, 38	1.91^{+}
Lex x Cont x AP x Group	4, 78	3.47^{+}

Lex = Lexical Status, Lat = Laterality, AP = Anterior-Posterior, Hemi = Hemisphere, Cont = Contrast (/h-0/ vs. /f-0/), Group (NS, HP, IP)
* p < .05, ** p < .01, *** p < .005, **** p < .001

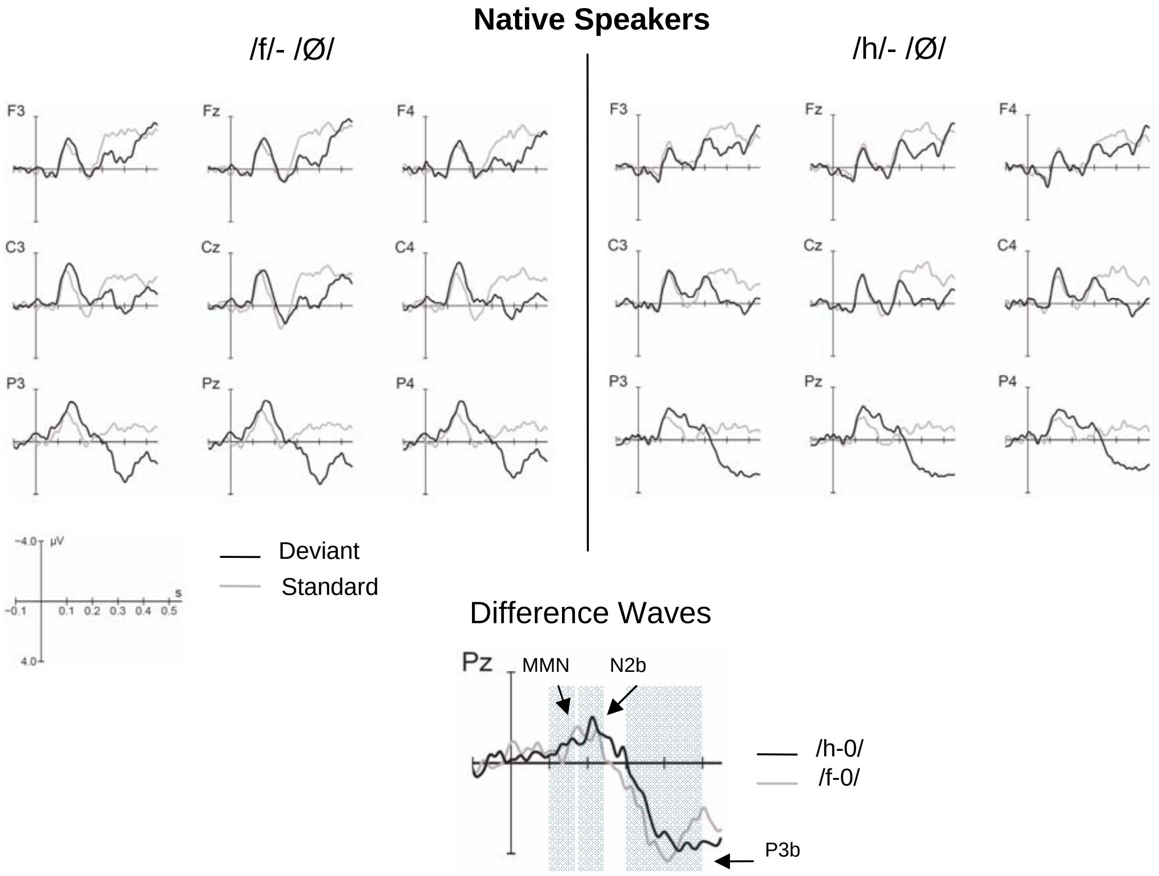


Figure 1. Grand average event-related brain potentials (ERPs) in response to the standard and deviant stimuli for the Native Speakers in response to for the /f-Ø/ and /h-Ø/ contrasts. In the bottom panel, difference waves (deviant-standard) show the MMN, N2b and P3b components.

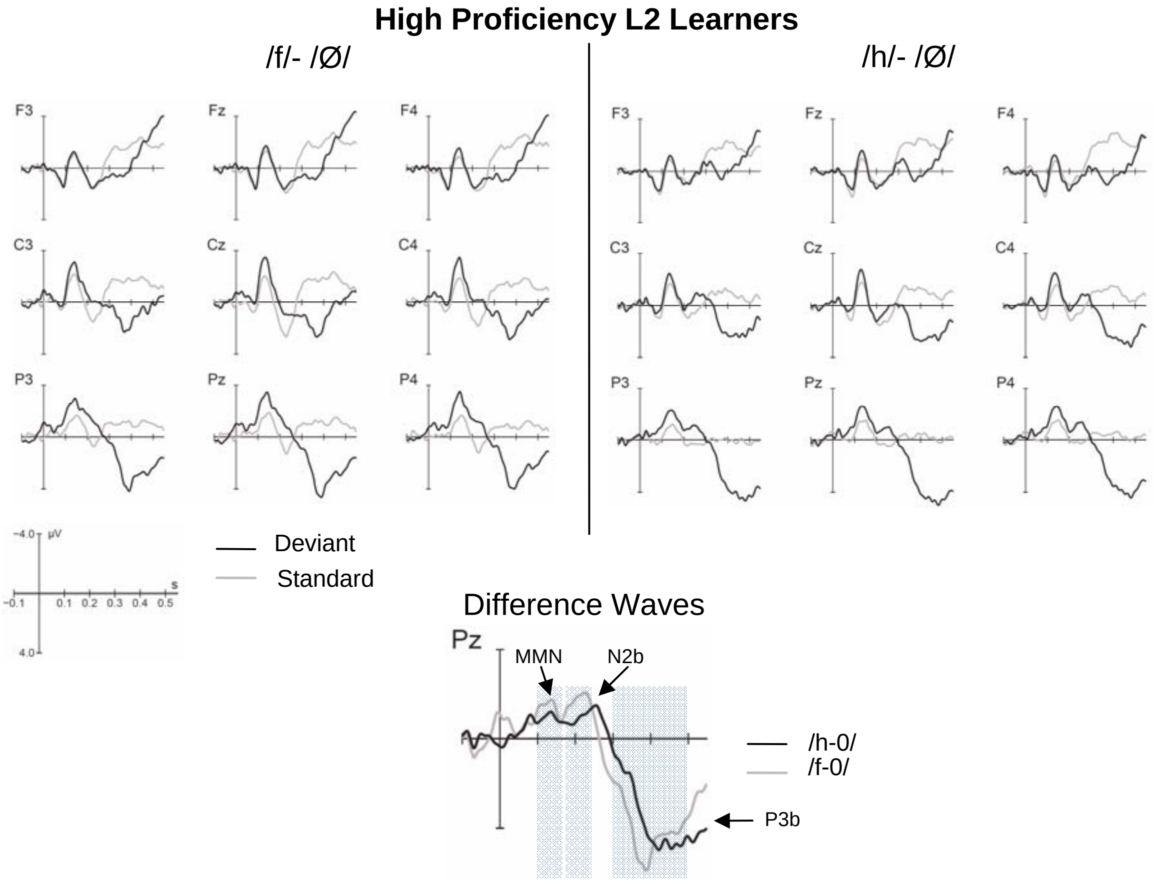


Figure 2. Grand average event-related brain potentials (ERPs) in response to the standard and deviant stimuli for the high proficiency (HP) L2 learners in response to for the /f-Ø/ and /h-Ø/ contrasts. In the bottom panel, difference waves (deviant-standard) show the MMN, N2b and P3b components for both contrasts.

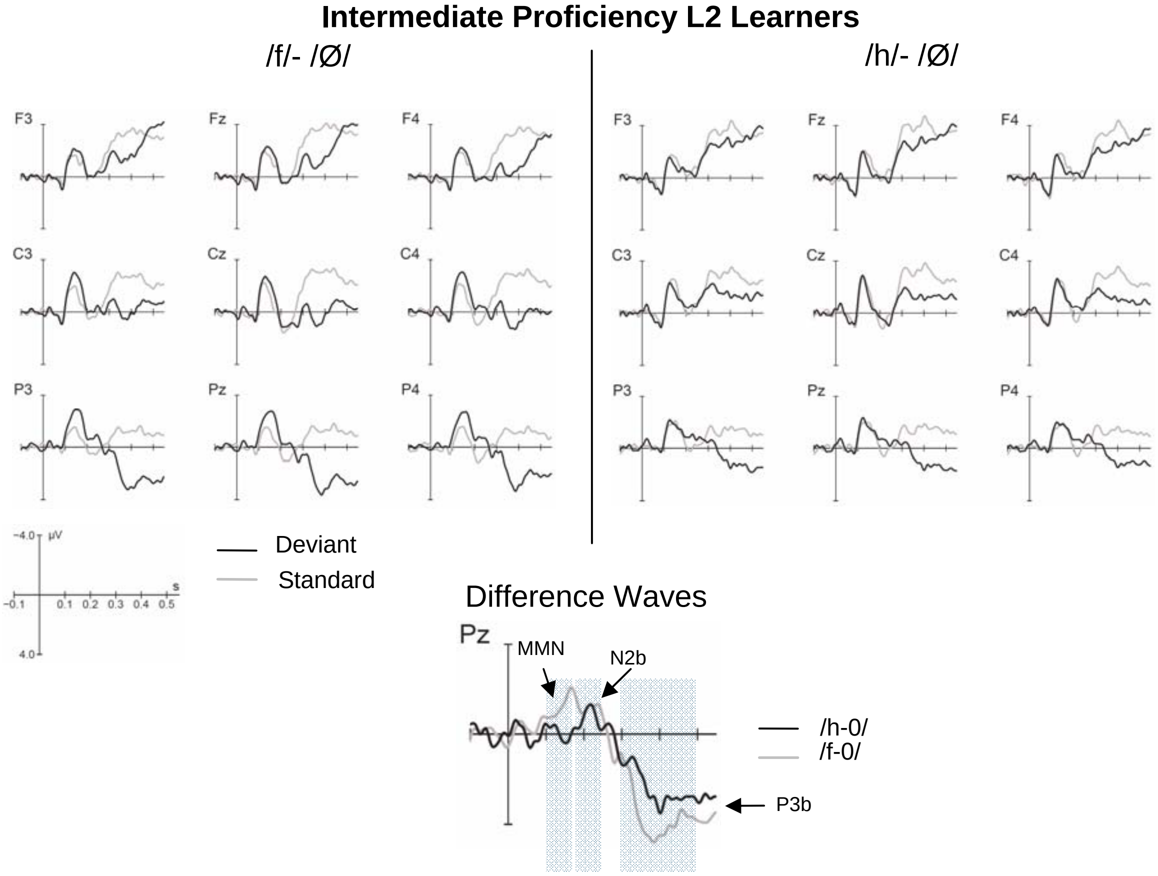
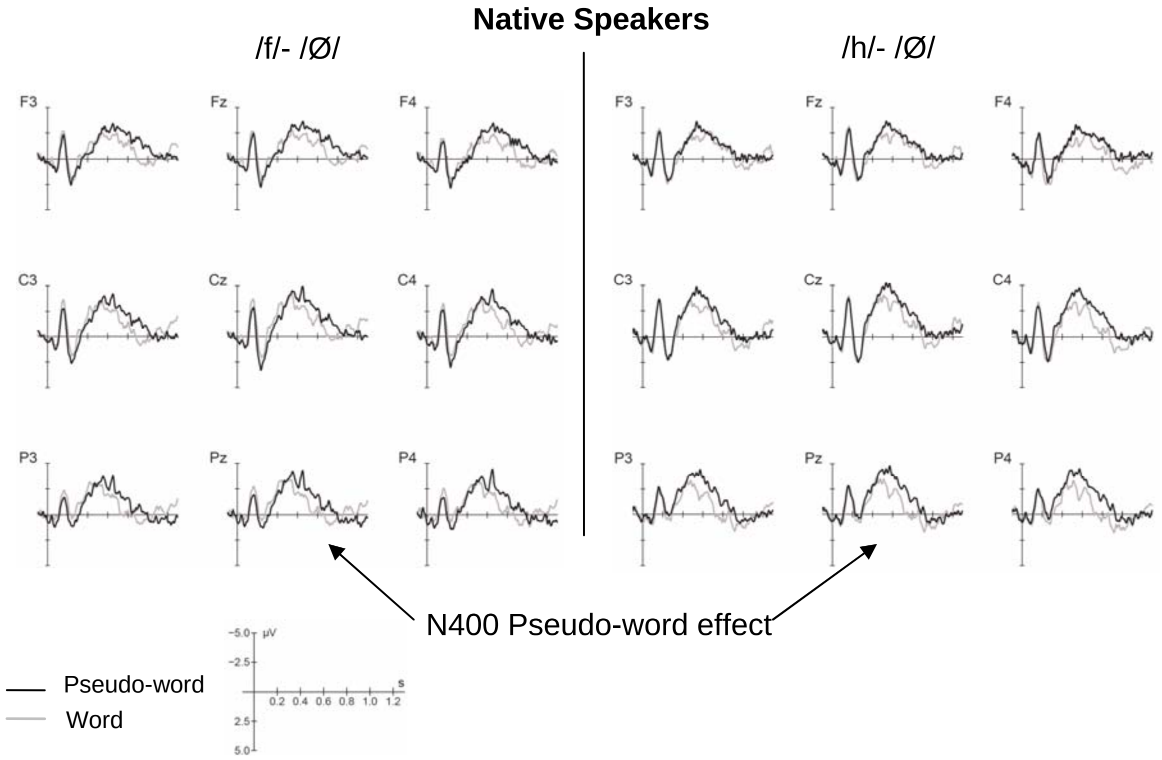


Figure 3. Grand average event-related brain potentials (ERPs) in response to the standard and deviant stimuli for the intermediate proficiency (IP) L2 learners in response to for the /f-Ø/ and /h-Ø/ contrasts. In the bottom panel, difference waves (deviant-standard) show the MMN, N2b and P3b components for both contrasts, although the negativity within the early time window is somewhat attenuated for the /h-Ø/ contrast.



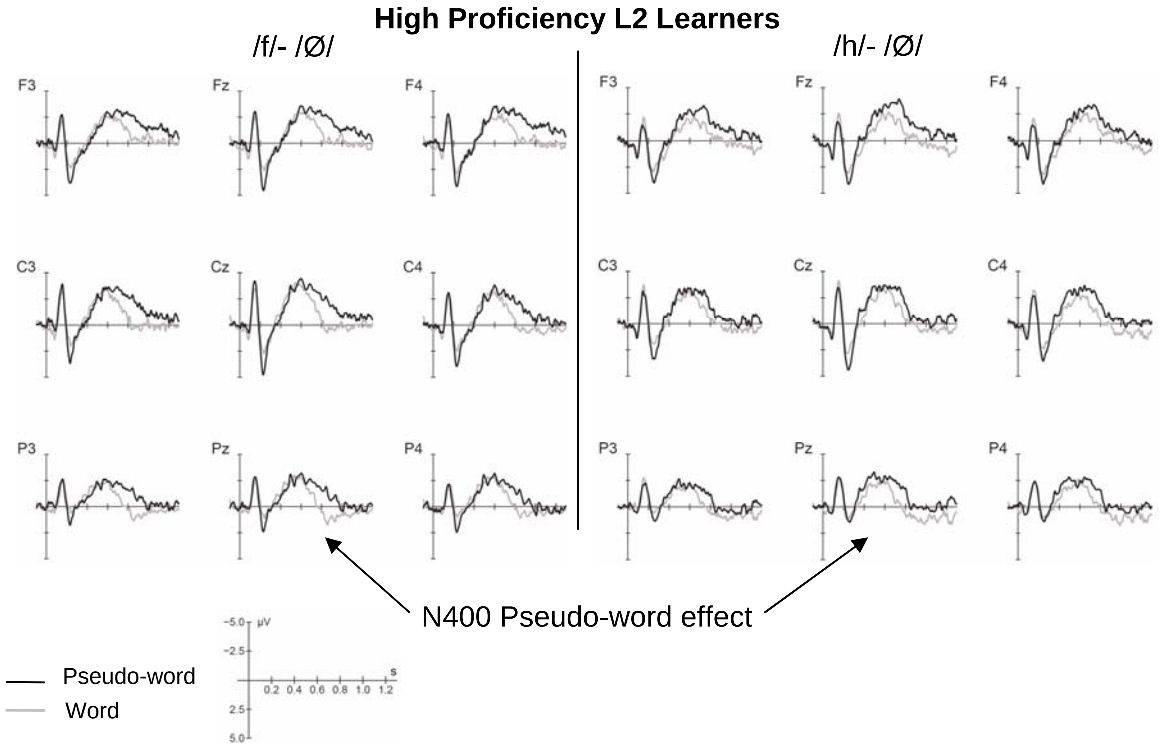
Native Speakers

/f/- /Ø/

/h/- /Ø/

N400 Pseudo-word effect

Figure 4. Grand average event-related brain potentials (ERPs) for the Native Speakers in response to words and pseudo-words for the /f-Ø/ and /h-Ø/ contrasts in Experiment 2. An N400 pseudo-word effect can be seen for both contrasts.



High Proficiency L2 Learners

N400 Pseudo-word effect

Figure 5. Grand average event-related brain potentials (ERPs) for the High Proficiency L2 learners in response to words and pseudo-words for the /f-Ø/ and /h-Ø/ contrasts in Experiment 2. An N400 pseudo-word effect can be seen for both contrasts.

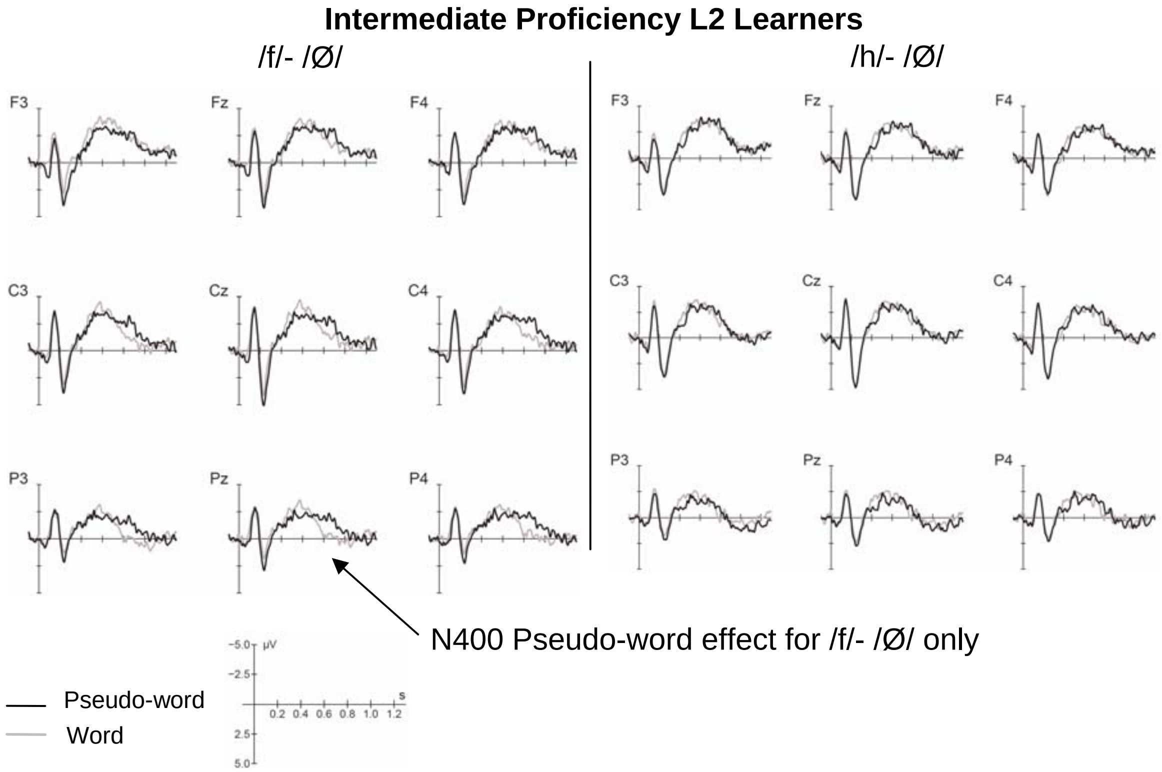


Figure 6. Grand average event-related brain potentials (ERPs) for the Native Speakers in response to words and pseudo-words for the /f-Ø/ and /h-Ø/ contrasts in Experiment 2. An N400 pseudo-word effect can be seen for the /f-Ø/ contrast only.

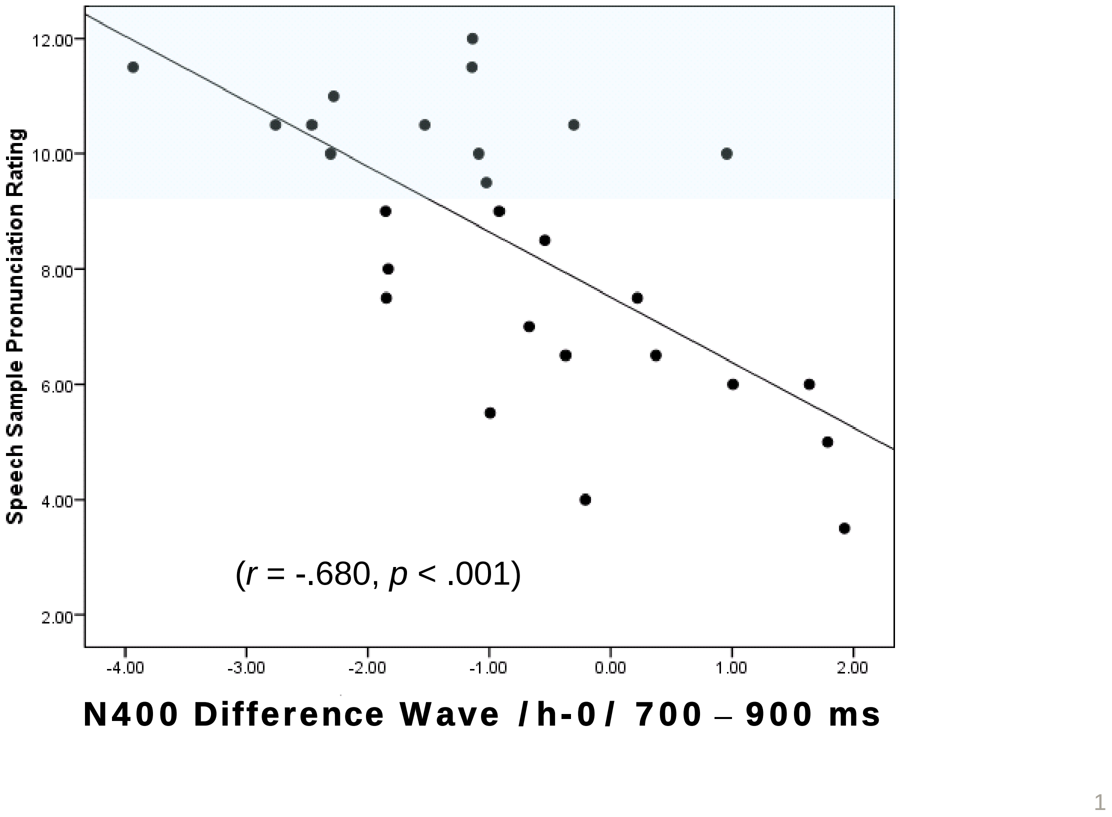


Figure 7. Correlation between the L2 learners' pronunciation ratings and the amplitude of their N400 pseudo-word effect for the /h-Ø/ contrast (at electrode Pz within the 700-900 ms time window).

Appendix 1: Sample Stimuli used in Experiment 2

Contrast	Word	Pseudo-word
/h/- /Ø/	Half	‘alf
	Hammer	‘ammer
	Happy	‘appy
	Haunted	‘aunted
	Height	‘eight
	Helmet	‘elmet
	Holiday	‘oliday
	Horse	‘orse
	Huge	‘uge
	Human	‘uman
	Afraid	H-afraid
	Anything	H-anything
	Egg	H-egg
	English	H-english
	Evening	H-evening
	Improve	H-improve
	Only	H-only
	Unsafe	H-unsafe
	Uncle	H-uncle
	Ugly	H-ugly
/f/- /Ø/	Factory	‘actory
	False	‘alse
	Father	‘ather
	Feather	‘eather
	Fence	‘ence
	Fever	‘ever
	Fifteen	‘ifteen
	Foot	‘oot
	Forgot	‘orgot
	Full	‘ull
	Alien	F-alien
	Answer	F-answer
	Earth	F-earth
	Easy	F-easy
	Exciting	F-exciting
	Idea	F-idea
	Improve	F-improve
	Others	F-others
	Understand	F-understand
	Upset	F-upset

General Discussion

The studies in this dissertation were undertaken to address two general questions about late L2 learners: (1) how does the neuro-cognitive basis of L2 morpho-syntactic processing change as late learners at low/intermediate proficiency levels become more proficient in their L2 (Study 1); and (2) to what extent can high proficiency late L2 learners recruit the neuro-cognitive mechanisms underlying native speakers' phonological processing, if an L2 speech sound does not exist in their L1 (Study 2)? These questions are linked to the broader issue of whether maturational constraints in the brain limit L2 acquisition in late (i.e., post-puberty) L2 learners. According to maturational and "critical period" accounts of AoA effects in L2 acquisition, the mature state of an adult's brain at the time of L2 acquisition restricts the extent to which L2 morpho-syntactic and phonological processing can engage the neuro-cognitive substrates used by native speakers (Johnson & Newport, 1989; Bley-Vroman, 1989; Neville & Bavelier, 2000; Sanders, Weber-Fox & Neville, 2008). By examining the neuro-cognitive bases of L2 processing by late learners at progressive stages of L2 proficiency, the present studies examined an alternative hypothesis; namely, L2 acquisition is associated with changes in neural processing that can ultimately lead to native-like processing profiles regardless of AoA. There follows a summary of the main findings from these studies and a discussion of how they further our understanding of these issues, ending with a discussion of how some limitations of the present studies motivate questions for future work.

General Summary

A number of noteworthy findings emerged from the present studies. In Study 1, it was demonstrated that, after intensive L2 instruction, Korean- and Chinese- late L2 learners of English began to exhibit significant P600 effects when presented with violations of English regular past tense. This is noteworthy because past tense verbal morphology operates differently in Korean and English and does not exist in Chinese; thus, this is a grammatical structure that would be predicted to be difficult, if not impossible, for these groups of learners to acquire and process like native speakers (Chen, Shu, Liu, Zhao & Li, 2007; Ojima, Nakata, Kakigi, 2005; Tokowicz & MacWhinney, 2005; Sabourin & Stowe, 2008; for reviews of studies investigating the role of L1-L2 similarities on neuro-cognitive processing, see Hernandez & Li, 2007, and Kotz, 2009). The implication of these findings is that late L2 learners can in fact call upon at least some of the neuro-cognitive mechanisms used by native speakers during L1 processing and apply them to the processing of L2 grammatical structures that either operate differently in the L1 and L2 or that are not present in the L1 at all and, thus, cannot be directly transferred from their L1.

Three lines of evidence suggested that the ability to exhibit P600 effects is linked to higher levels of L2 proficiency. First, P600s were observed in these learners after, but not before, intensive L2 instruction. Second, the results of the correlation analyses showed that the amplitude of this effect was larger for individuals with relatively high levels of behavioural performance at the end of the intensive L2 course in comparison to learners with relatively low levels of performance. Third, the amplitude of the P600 was larger for trials that corresponded to correct behavioural responses in comparison to “all trials” (i.e., trials with correct and incorrect behavioural responses), suggesting that larger P600s were elicited in response to violation sentences that the L2 learners correctly identified as ungrammatical. Although it is difficult to determine causality, it can be inferred from the results of Study 1 that the ability to recruit the neuro-cognitive mechanisms underlying the P600 is associated with higher levels of L2 proficiency. These findings reveal the capacity for qualitative neural changes that can co-occur with L2 learning in adults, even when processing L2 grammatical structures that are unique to their L2 and, thus, cannot be directly transferred from their L1. They further reveal that the capacity to engage native-like neuro-cognitive processing is preserved in adulthood, such that neural changes co-vary with the success of L2 learning and not primarily with the age at which it takes place.

In Study 2, it was found that proficiency also plays an important role in influencing the extent to which late L2 learners engage native-like neuro-cognitive mechanisms when presented with a difficult L2 phonetic contrast. Native-like processing was observed for high proficiency late L2 learners, but not intermediate proficiency L2 learners, even under difficult task/stimulus conditions. Specifically, native French speakers who had acquired English after puberty were tested in their processing of an English-specific phonetic contrast that is notoriously difficult for native French speakers to acquire: /h/ versus /Ø/ (Janda & Auger, 1992; LaCharité & Prévost, 1999; Mah, 2011). The L2 learners were classified as intermediate or high proficiency based on their pronunciation skills and then, along with native English speakers, were tested on their processing of the non-native phonetic contrast in two stimulus/task conditions. It was found that when the contrast was presented as simple syllables in an attended discrimination paradigm (Experiment 1), both the intermediate and high proficiency L2 learners showed evidence of engaging the same series of neuro-cognitive processes as native speakers (i.e., MMN, N2b and P3b responses). The ERP components elicited during this study are thought to reflect both attention-driven (i.e., N2b/P3b) and automatic (i.e., MMN) processing. However, in Experiment 2, when the contrast was presented in words and pseudo-words in a semantic monitoring task, only the high proficiency L2 learners, like native speakers, showed evidence of discriminating the phonetic contrast, as revealed by the N400 pseudo-word effect. This was taken as evidence for automatic L2 processing in the high proficiency learners because it was elicited even when they were engaged in a task that directed their attention away from phonetic analysis. Moreover, the amplitude of the N400 across all learners was significantly correlated with pronunciation ratings, further reinforcing the importance of proficiency in mediating this effect. That no significant differences were observed between the native speakers and high proficiency L2 learners in any of the ERP components examined for was taken as evidence for native-like processing.

The results of Study 2 suggest both that native-like L2 phonological processing is possible for late L2 learners. Moreover, very high levels of proficiency may be necessary in order for L2 learners to recruit native-like automatic neuro-cognitive processes when they are engaged in complex tasks that direct attention to other levels of analysis, such as semantics. The conclusions of Study 1 and 2 will now be discussed with respect to the research questions that motivated this research.

Neuro-cognitive Changes and L2 Learning

One goal of the studies in this dissertation was to elucidate whether and how the neuro-cognitive basis of L2 processing changes as late L2 learners become more proficient in their L2. Steinhauer, White and Drury (2009) proposed a model of the neuro-cognitive stages that a hypothetical L2 learner may pass through as he/she advances in L2 proficiency, as revealed by distinct ERP patterns exhibited in response to morpho-syntactic violations. It was proposed that as late L2 learners become more proficient in an L2, the ERP components they elicit will begin to approximate those of native speakers in a systematic way. Specifically, it was predicted that L2 learners who initially show no ERP effects for L2 morpho-syntactic violations would begin to exhibit P600s that increase in amplitude as proficiency improves. This model predicts further that, at very high levels of proficiency, learners might also display LAN effects (associated with automatic grammar processing) that become increasingly left-lateralized as proficiency advances to native-like levels.

These hypotheses were formulated largely on the basis of cross-sectional data. The results of Study 1 provide important corroborative longitudinal evidence from L2 learners of a natural language as they advance from relatively low to intermediate L2 proficiency. At the start of their intensive L2 course, the behavioural performance of the L2 learners in Study 1 on the grammaticality judgement task was low (around chance level) and was associated with no discernible difference in ERPs for violation and well-formed sentences. However, by the end of the course, improvement in behavioural performance was associated with significant increases in the amplitude of P600 effects. Moreover, the amplitude of P600 effect increased as a function of individual behavioural improvement. This was taken as evidence for grammaticalization of the morpho-syntactic structures under investigation and the use of the same neuro-cognitive processes underlying the P600 in native speakers when presented with sentences that violate those structures. The finding of larger P600 effects for trials that corresponded to correct behavioural responses (i.e., "response contingent analyses") compared to "all trials" (i.e., trials with correct and incorrect behavioural responses) provided further evidence for a tight link between the neuro-cognitive processes underlying the P600 and behavioural performance. These findings corroborates the results of Tanner, McLaughlin, Herschensohn and Osterhout (2012) who found a similar correlation between P600 amplitude and behavioural performance for morpho-syntactic structures that are similar in the L1 and L2. The results of Study 1 suggest that this relationship reflects L2 proficiency rather than L1-L2 typological similarities.

It is interesting to note from Study 1 and Study 2 that, in the domains of L2 morpho-syntax and phonology, a similar sequence of neuro-cognitive processes emerged as learners became more proficient in their L2. In both studies, it was found that late L2 learners who attained intermediate levels of L2 proficiency appeared to engage primarily controlled, attention-driven neuro-cognitive processes. This was reflected in Study 1 by P600 effects observed after L2 instruction and in Experiment 1 of Study 2 by N2b/P3b effects observed in both the intermediate and high proficiency L2 learners when they were engaged in a task that permitted attention to be directed to relevant phonetic information. However, what appears to differentiate high proficiency L2 learners from those who have attained intermediate levels of L2 proficiency is the ability to also engage automatic aspects of processing. For phonology, automatic aspects of processing were observed in Experiment 2 of Study 2 when phonetic discrimination was assessed while participants were engaged in a task that required semantic, rather than phonetic, analysis. In this case, only the high proficiency late L2 learners showed evidence of automatic L2 phonological processing. Similarly for morpho-syntactic processing, the results of Steinhauer, White and Genesee (2012) showed that high, but not intermediate, proficiency L2 learners displayed the LAN response that is associated with automatic aspects of grammar processing in native speakers. In other words, it appears that during the course of L2 acquisition, late L2 learners experience a series of changes in neuro-cognitive activity, which leads to patterns of neuro-cognitive activity that progressively resembles that displayed by native speakers (i.e., reliance on attention-driven processing, followed by emergence of automatic processing). In both Study 2 and the Steinhauer et al. (2012) study, the high- and intermediate-proficiency learners also differed in their reported L2 use at the time of testing. Consistent with some behavioural studies, discussed in the General Introduction (e.g., Birdsong & Molis, 2001; Piske, MacKay & Flege, 2001), these findings highlight a non-trivial role for practice in attaining high levels of L2 proficiency and in engaging 'native-like' automatic aspects of L2 processing.

It is important to note that in Experiment of Study 2, no significant differences were observed between native speakers and both groups of L2 learners in terms of their MMN response, suggesting that even the intermediate proficiency group was able to engage automatic L2 phonetic processing when the task conditions were simple and allowed participants to focus on phonetic analysis. These findings contrast with Mah's (2011) previous study that also tested the processing of the /h/-/Ø/ contrast in native French speaking late L2 learners of English, but did not observe MMN effects when participants listened to the stimuli while watching a silent movie. The results of Experiment 1 also contrast with the results of Experiment 2 (Study 2) in which the intermediate proficiency L2 learners did not show a significant N400 pseudo-word effect when the task required them to engage in semantic, rather than phonetic, analysis.

These results from Study 2 have two important implications. First, they suggest that stimulus and task conditions may influence the extent to which L2 learners can recruit the neuro-cognitive mechanisms used for L2 phonetic processing. More specifically, they suggest that L2 learners who show evidence of engaging native-like processes when task conditions permit them to direct attention to relevant phonetic information may not recruit the same processes when their attention is directed elsewhere – on semantic analysis, for example. This is important because it suggests that it may not be the processing of L2 structures *per se* that is difficult for L2 learners, but in marshalling the necessary neuro-cognitive processes in complex listening conditions which require that processing resources be shared with other levels of processing. Second, they suggest that the ability to deploy automatic L2 neuro-cognitive processing may fall on a continuum such that phonetic processing becomes increasingly automatized as learners become progressively more proficient in the L2 and thus, increasingly successful in a wider range of stimulus and task conditions. These points will be discussed further with respect to future research directions.

Age of L2 Acquisition and L2 Proficiency Effects on L2 Neuro-Cognitive Processing

The most important conclusion that can be drawn from the present studies is that L2 proficiency, rather than AoA alone, plays a critical role in influencing the neuro-cognitive mechanisms that late L2 learners use during L2 processing. Specifically, as L2 proficiency increases, patterns of neuro-cognitive activity progressively resemble those of native speakers. As discussed earlier, Study 1 demonstrated that the emergence of one aspect of native-like morpho-syntactic processing (i.e., P600 effects) was systematically tied to improvements from relatively low to intermediate levels of L2 proficiency. Moreover, in Study 2, high L2 proficiency was associated with native-like neuro-cognitive processing profiles, as evidenced by the lack of significant differences between native speakers and the high proficiency group for all of the ERP components examined.

These findings are particularly compelling in light of two important points that should be reiterated. First, all of the L2 learners tested in Studies 1 and 2 had acquired their L2 after puberty, a time that is often thought to mark a turning point in the capacity for native-like processing (Johnson & Newport, 1989; Lenneberg, 1967), and they were tested on two domains of language, morpho-syntax and phonology, that are thought to be particularly susceptible to AoA effects (Clahsen & Felser, 2006; Johnson & Newport, 1989; Lamendella, 1977; Long, 1990; Neville & Bavelier, 2000; Sanders et al., 2008; Scovel, 1989). Second, the particular structures that were examined in these studies are thought to be difficult for the L2 learners who participated in these studies because they are not expressed in a similar way in the learners' L1 and, therefore, cannot be directly transferred (i.e., regular past tense verbal morphology for native Korean and Chinese speakers, and the English specific /h/-/Ø/ phonetic contrast for native French speakers). Thus, if maturation of the brain restricts the ability to acquire and process L2 morpho-syntactic and phonological information in a native-like manner (Bley-Vroman, 1989; Johnson & Newport, 1989; Neville & Bavelier, 2000; Sanders et al., 2008), clear differences between native speakers and *all* L2 learners should have been observed in the present studies. Such a conclusion would have been in line with some previous ERP studies that did not control for L2 proficiency and concluded that an early AoA is a prerequisite for native-like L2 processing (e.g., Weber-Fox & Neville, 1996; Hahne & Friederici, 2001).

However, the results of the present studies revealed striking similarities between the neuro-cognitive bases of morpho-syntactic and phonological processing in native speakers and these late L2 learners. These findings corroborate the results of a growing number of studies that report native-like L2 morpho-syntactic and phonological processing profiles in high proficiency L2 late learners (Bowden, Sanz, Steinhauer & Ullman, 2012; Dowens, Vergara, Barber & Carreiras, 2010; Friederici, Steinhauer & Pfeifer, 2002; Morgan-Short, Sanz, Steinhauer & Ullman, 2010; Morgan-Short, Steinhauer, Sanz & Ullman 2012; Steinhauer et al., 2012; Winkler et al., 1999). Native-like L2 processing, thus, appears to be contingent on high levels of L2 proficiency, rather than an early AoA. In effect, age effects in L2 attainment do not appear to be due to a categorical inability for the brain to engage native-like neuro-cognitive processes after a certain AoA, and L2 processing by late L2 learners does not appear to be "fundamentally different" from that of early learners and native speakers (as proposed by Bley-Vroman, 1989).

Future Directions

The present studies, like many other ERP studies of L2 processing, used stimuli and task conditions that were relatively simple and imposed few additional processing demands. For example, in Study 1, the test sentences were relatively short and in the active voice. As well, participants were tested in a quiet listening booth and were engaged in a grammaticality judgement task that directed their attention to the grammatical structure of the sentences. This methodology was intentional because the purpose of this study was to test whether it is possible, in principle, for late L2 learners to exhibit P600 effects in response to tense violations. If no P600 had been observed at session 2, even in such simple task/stimulus conditions, it would have been compelling support for maturational accounts of L2 acquisition. However, that P600s *were* observed in Study 1 is merely a first step towards arguing against maturational accounts because it may be that sentence processing in these conditions differs considerably from real life language processing where processing requirements are more complex and demanding.

Study 2 sought to examine these factors and, specifically, to investigate how stimuli and task conditions impact processing. In fact, as discussed earlier, it was found that for intermediate proficiency L2 learners, stimulus/task factors influenced the extent to which they engaged the same neuro-cognitive mechanisms used by native speakers. This was revealed by MMN, N2b and P3b effects in Experiment 1 and the absence of a N400 pseudo-word effect in Experiment 2 when processing the same /h/-/Ø/ phonetic contrast.

These findings are important because they suggest that although L2 morpho-syntactic and phonological processing may involve similar neuro-cognitive mechanisms in native speakers and intermediate to high proficiency late L2 learners, the range of situations in which these processes can be applied by native and L2 speakers may still differ. A similar conclusion was drawn by Dowens et al. (2009) in a study of gender agreement processing by highly proficient late L2 learners. They found that these learners exhibited a native-like LAN/P600 response for violations between articles and nouns that were contained in a single phrase. However, the same learners exhibited a P600 but no LAN in response to violations between nouns and adjectives that were further apart in the test sentence (i.e., in different phrase). Native speakers, in contrast, exhibited a LAN/P600 pattern in response to violations in both sets of test conditions. The authors suggested that the cost of maintaining the agreement features in working memory in the long distance noun-adjective condition may have been more taxing for the L2 learners than the native speakers. As a result, they did not exhibit the LAN response that is associated with automatic processing of the violation in the more difficult condition. Differences between conditions in the Dowens et al. study and between the results of the intermediate proficiency group in Experiments1 and 2 of Study 2 suggest that it is not the processing of L2 grammatical and phonological structures alone or *per se* that is difficult for L2 learners, but engaging these processes under taxing conditions (see also McDonald, 2006).

It may be that some behavioural studies report differences between native speakers and late L2 learners because the stimulus/task conditions are more complex than those used in other ERP studies that report native-like processing for late L2 learners (see Abrahammson & Hyltenstam, 2009, for a discussion). Future ERP studies could test this by comparing L2 morpho-syntactic and phonological processing in simple and complex conditions that tax memory capacity (e.g., long distance dependencies), decoding ability (e.g., listening to stimuli presented with white noise) or during tasks that direct attention to other levels of linguistic analysis (e.g., semantic monitoring task). Investigating the relationship between AoA and L2 proficiency, on the one hand, and L2 neuro-cognitive processing, on the other hand, in situations that impose high or low cognitive load would significantly advance our understanding of how multiple factors work together to influence L2 processing.

While the present studies examined neuro-cognitive processing by late L2 learners in different linguistic domains (namely, morpho-syntax and phonology), different groups of L2 learners were examined in each domain. This allowed for an investigation of how the presence or absence of particular linguistic structures in the learners' L1 influenced the processing of the target grammatical and phonological structures. It was also due to practical constraints inherent to ERP research, in which a larger number of trials are required per condition than in many behavioural studies. However, Abrahamsson and Hyltenstam (2009) have argued that although late L2 learners may perform like native speakers when tested in limited areas of the target language, the ability to achieve native-like attainment in a range of linguistic domains is difficult, if not impossible, for late L2 learners. In fact, this level of "global native-like-ness", they argued, may be uncommon among even early L2 learners. Thus, it is not necessarily the case that late learners who are highly proficient in one domain (i.e., morpho-syntax) are also highly proficient in another (e.g., phonology). Moreover, Steinhauer et al. (2009) have argued that different levels of "structure specific" proficiency are to be expected even within each domain, partly as a function of the type of L1-L2 paring (which determines the specific kinds of transfer effects possible).

A goal for future research would be to investigate the extent to which L2 learners display native-like profiles of neuro-cognitive activity in a range of linguistic domains and a range of structures within each domain of the L2. This would allow for a better understanding of the full scope of neuro-cognitive processes that are available to individual late L2 learners, how they may differ from those used by native speakers, and how they may be affected by AoA and L2 proficiency. A few ERP studies are already underway examining the processing of different L2 morpho-syntactic structures as a function of the presence/absence of the structure in the L1 and the learners' level of behavioural proficiency with respect to those structures in the L2 (Kasparian, Bourguignon, Drury and Steinauer; 2010; 2011).

Another point for future research relates to the choice comparison group. In the present study, as in the majority of ERP studies of L2 processing, neuro-cognitive processing by late L2 learners was compared to that of monolingual English speakers. While necessary and useful, this approach also has limitations. More specifically, it is unclear whether any differences that are observed between native speakers and late L2 learners using this approach are due to the effect of storing of two languages in the same brain or the effect of acquiring one of them later in life (Grojean, 1989). In this respect, simultaneous bilinguals (i.e., individuals who have been native speakers of two languages since birth) would be a useful and interesting additional comparison group. Studies of phonological processing illustrate this point. For example, voice onset timing (VOT) is a phonological feature that can be used to discriminate phonetic categories in English and French, although the exact position of the boundary differs between languages. Caramazza, Yeni-Komshian, Zurif and Carbone (1973) showed that French- early L2 learners of English (AoA ≤ 7) did not perform like monolinguals of French or English when tested on their perception of VOT contrasts in both languages, suggesting that even L1 phonetic perception can be influenced by the act of acquiring another language. Similarly, ERP evidence (Molnar, 2009) and word production data (Bosch & Ramon-Casas, 2011) from simultaneous bilinguals has shown that the experience of learning two languages, even from birth, can influence the way vowel categories are represented in both languages compared to monolinguals.

When possible, future studies should attempt to control for this potential confound. This may be particularly important when testing the processing of speech sounds that are categorized differently in the two languages and, thus, potentially compete for representation and processing (Hernandez, Li & MacWhinney). For example, using simultaneous bilinguals as the comparison group may be more appropriate than monolinguals when testing vowel contrasts that are phonemic in one language and allophonic in the other, or contrasts that use features such as VOT, which may be phonemic in both languages but nevertheless categorizes sounds in each language differently. These effects of storing two languages in the same brain may account for the low incidence of native-like performance in L2 learners tested in some previous studies of VOT perception and production (e.g., Abrahamsson & Hyltenstam, 2009).

Finally, another important avenue for future research would be to compare the neuro-cognitive processes used by early (i.e., pre-puberty) and late (i.e., post-puberty) L2 learners at progressive stages of L2 proficiency. In the present studies, this was done for late learners only. Similarly, examining developmental changes in neuro-cognitive processing in children as they acquire their L1 would shed light on the question of whether language acquisition follows a similar path in early (L1) and late L2 learners. Indeed, it has been argued that L2 morpho-syntactic processing by late L2 learners differs from L1 processing by both adult and child native speakers in that late L2 learners rely less on grammatical parsing routines and more on lexical storage and semantic information during grammatical processing (Clahsen & Felser, 2009; Ullman, 2001). However, relatively few ERP studies have compared these populations directly and there is little ERP evidence of L2 morpho-syntactic processing for child L2 learners. Longitudinal studies comparing the neuro-cognitive changes associated with increased L1/L2 proficiency would allow for a critical examination of the link between proficiency and processing at different ages in different learner groups.

In a related vein, surprisingly few developmental ERP studies of morpho-syntactic processing have been conducted with children, even in their L1. However, those that exist suggest age-related and, thus, probably proficiency-driven, differences in neuro-cognitive processing that are similar to those demonstrated in this dissertation for low and high proficiency late L2 learners. For example, using an auditory phrase violation paradigm, Pakulak Sanders, Paulsen and Neville (2005) compared morpho-syntactic processing in 3 to 5 year old and 6 to 8 year old L1 children who were grouped into high and low proficiency groups based on their grammatical proficiency scores on a standardized test. In response to the violations, high proficiency 6-8 year olds displayed the LAN-P600 pattern observed previously with adults, whereas 3-5 year olds displayed a P600 and no LAN. Neither LAN nor P600 effects were observed for low proficiency children in either age group. Similar age-related changes in L1 morpho-syntactic processing were reported by Hahne, Eckstein and Friederici (2004) with LAN/P600 effects observed for 7-13 year olds and P600s but no LANs for 6 year olds (however, for a discussion of methodological issues pertaining to the stimuli used in both these studies, see Steinhauer & Drury, 2011; Steinhauer et al., 2012). These findings suggest that controlled, attention-driven aspects of L1 processing become available at an earlier age than automatic neuro-cognitive processing – exactly the pattern discussed earlier with respect to late L2 learners. Moreover, the Pakulak et al. study demonstrates a clear role for proficiency, independent of age, in eliciting these components. More work is clearly needed to elucidate the sequence of neuro-cognitive stages that learners pass through during both L1 and L2 acquisition and the extent to which they differ depending on the age or proficiency of the learner.

Summary: The State of Maturational Accounts of L2 Acquisition

In summary, the results of the present studies do not provide support for maturational accounts of L2 acquisition. That is, the maturational state of the brain does not appear to categorically restrict the capacity for late L2 learners to process L2 morpho-syntactic and phonological structures using similar neuro-cognitive processing mechanisms as native speakers, at least within the limits tested in this dissertation. To the contrary, the present results suggest that the adult brain is capable of acquiring new L2 information and of changing in systematic ways with learning, such that L2 morpho-syntactic and phonological processing increasingly resembles that of native speakers as L2 proficiency advances.

An important outstanding question is the extent to which late L2 learners are able to access native-like neuro-cognitive processes in more complex task/stimulus conditions that tax L2 processing capacity (e.g., when processing complex sentences, subtle linguistic anomalies, stimuli that are presented with noise, or when engaged in tasks that direct attention to other, less relevant, levels of linguistic analysis). Comparing late L2 learners at progressive stages of L2 proficiency to early L2 learners or simultaneous bilinguals in cognitively simple and complex conditions will help to elucidate the extent to which age effects in L2 acquisition are due to neuro-cognitive differences in L2 morpho-syntactic and phonological processing *per se* or differences in how this knowledge interacts with more general processing capacities. Such an inquiry would also provide a closer approximation of L2 processing in real-world situations.

The results of the studies in this dissertation are difficult to align with strict interpretations of the critical period hypothesis, which would predict large scale differences in the neuro-cognitive systems that late (i.e., post-puberty) L2 learners use during L2 processing compared to native speakers. On the contrary, the present results reveal the capacity for late learners to experience systematic changes in neuro-cognitive processing that are associated with improvements in L2 proficiency and which may ultimately lead to native-like processing profiles. Although future research is necessary to investigate the extent to which these results generalize to other L1-L2 pairings and to processing in more complex listening conditions, the results here suggest that at the very least, the maturational state of the brain at the time of L2 acquisition may not restrict all aspects L2 morpho-syntactic and phonological processing by late L2 learners. This conclusion, although incompatible with maturational accounts of L2 acquisition, is consistent with current research in neuroscience that demonstrates the capacity for various neural systems to be modified by new experiences and learning opportunities throughout the lifespan (e.g., Mahncke, Bronstone & Merzenich, 2006; van Praag, Kempermann & Gage, 2000).

It is likely that not one but a number of factors contribute to the oft observed finding that adults tend to achieve lower and more variable levels of L2 attainment than children. The results of the present studies suggest that widespread decrease in neural plasticity at the time of puberty within the systems supporting L2 morpho-syntactic and phonological processing may not be the primary cause of AoA effects in L2, as argued by some previous studies (see Neville & Bavelier, 2000; Sanders et al., 2008). Other important factors may include differences in the amount and kind of L2 input that children and adults tend to receive (Marinova-Todd, Marshall & Snow, 2000)[16], differences in learning environments (Morgan-Short et al., 2012), and individual differences that relate to genetic factors that predispose certain individuals to varying levels of neural plasticity (e.g., the brain-derived neurotropic factor that is thought to be involved with synaptic plasticity and learning; Tyler, Alonso, Bramham & Pozzo-Miller, 2002), cognitive skill (e.g., phonological short-term memory and working

---

[16] Ellis (1995; Ellis, Loewen & Erlam, 2006) have suggested that L2 learning can be greatly facilitated by L2 instruction that helps learners to notice grammatical features in the input, comprehend their meaning and engage in the cognitive process of comparing the forms in the input with the forms they produce in their output. This may be particularly important for the acquisition of L2 morpho-syntactic or phonological structures that do not exist in the L1 or which operate differently in the L1 and L2, and thus pose particular challenges for late L2 learners.

memory; O'Brien, Segalowitz, Collentine & Freed, 2006; Mcdonald, 2006), language aptitude (Skehan, 1989), and attitudes and motivation (Gardner, 1985). The goal for future research should be to understand how these factors work together to shape language processing in learners at different ages. Such an understanding may better enable us to facilitate L2 attainment, regardless of AoA.

General References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny. *Language Learning, 59* (2), 249-306.

Abutalebi, J., Cappa, S. F. & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism 4,* 179–90.

Bialystok, E., & Hakuta, K. (1999). Confounded age: linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 161-181). Mahwah, New Jersey: Erlbaum.

Bialystok, E., & Miller, B. (1999). The problem of age in second language acquisition: Influences from language, task, and structure. *Bilingualism: Language and Cognition, 2,* 127-145.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68*(4), 706-755.

Birdsong, D. (1999). Introduction: The whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Critical Period Hypothesis and second language acquisition* (pp. 1-22). Mahwah, New Jersey: Erlbaum.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56,* 9-49.

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

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). Cambridge, UK: Cambridge University Press.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31,* 175-198.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: the case of very advance late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp.133- 159). Mahwah, New Jersey: Erlbaum.

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories? *Journal of Phonetics, 39*(4), 514-526.

Bowden, H. W., Sanz, C., Steinhauer, K., & Ullman, M. T., (2012) Can you attain native-like brain processing of syntax when you learn a foreign language through college? Unpublished manuscript. University of Tennessee, Knoxville.

Caramazza, A., Yeni-Komshian, G. H., Zurif E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in

French□English bilinguals. Journal of Acoustical Society of America, 54(2), 421-428.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*(2), 161-174.

Clahsen, H. & Felser, C. (2006). How native-like is non-native processing? *Trends in Cognitive Sciences, 10*(12), 564-570.

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters.

Doupe, A. J. & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Reviews Neuroscience, 22,* 567-631.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2009). Morphosyntactic processing in late second language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870-1887.

Ellis, R (1995). Interpretation tasks for grammar teaching. *TESOL Quarterly, 29*(1), 87-105.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28,* 339-368.

Eubank, L. & Gregg, K. R. (1999). Critical Periods and (Second) language Acquisition: Devide et Impera. In D. Birdsong (Ed.), *Critical Period Hypothesis and second language acquisition* (pp. 65- 99). Mahwah, New Jersey: Erlbaum.

Friederici, A. D. (2002). Towards a neural basis of auditory sentences processing. *Trends in Cognitive Science 6*(2), 78-84.

Friederici, A. D., Steinhauer, K. & Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *PNAS, 99*(1), 529-534.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold Publishers.

Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual language development and disorders: A handbook on bilingualism and second language learning*. Baltimore: Brookes Publishing.

Grojean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*, 3–15.

Hagoort, P., Brown, C. and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and Cognitive Processes 8, 439–83.

Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience, 16*(7), 1302-1318.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194-205.

Hahne, A. & Friederici, A. D. (2001). Processing a second language: late learners' comprehension mechansims as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*(2), 123-141.

Hakuta, K., Bialystok, E. & Wiley, E. (2003). Critical evidence: A test of the critical period hypothesis for second-language acquisition. *Psychological science, 14*(1), 31-38.

Harley, B. & Wang, W. (1997). The critical period hypothesis: where are we now? In de Groot, A.M.B., & Kroll, J. F. (Eds.) *Tutorials in bilingualism Psycholinguistic Perspectives* (pp. 19- 51). Mahwah, New Jersey: Erlbaum.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin, 133*(4), 638-650.

Hernandez, A., Li, P. & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9*(5), 220-225.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539–588). Malden, MA: Blackwell.

Janda, R. D., & Auger, J. (1992). Quantitative evidence, qualitative hypercorrection, sociolinguistic variables--And French speakers' 'eadhaches with English h/Ø. *Language & Communication, 12*(3-4), 195-236.

Johnson, J.S. & Newport, E.L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60-99.

Kasparian, K., Bourguignon, N., Drury, J. E. & Steinhauer, K. (2010). On the influence of proficiency and L1-background in L2 processing: An ERP study of nominal morphology in French and Mandarin learners of English. *2nd Neurobiology of Language Conference (NLC), San Diego, (CA, USA), November 2010.*

Kasparian, K., Bourguignon, N., Drury, J. E., & Steinhauer, K. (2011). Effects of L2-proficiency and L1-background on L2 sentence processing: An ERP study of adjective-noun word order in French and Mandarin late learners of English *International Symposium on Bilingualism (ISB8), Oslo, Norway, June 2011*

Kim, K. H. S., Relkin, N. R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature, 388,* 171-174.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16*, 1412–1425.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language, 109*, 86-74.

Krashen, S.D., Long, M.H., & Searcella, R.C. (1982). Age, rate, and eventual attainment in second language acquisition. In S.D. Krahsen, R.C. Scarcella, M.H. Long (Eds.) *Child-adult differences in second language* *acquisition series on issue in second language research*. Rowley, Massachusetts: Newbury House.

Kuhl, P., (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5,* 831-843.

LaCharité, D. & Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. In A. Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of BUCLD 23* (pp. 373-385). Somerville: Cascadilla.

Lamendella, J. (1977). General principles of neuro-functional organization and their manifestations in primary and non-primary language acquisition. *Language Learning, 27,* 155-196.

Lau, E. F. Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9,* 920-933.

Lennenberg, E. (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.

MacWhinney, B. (2005). *A unified model of language acquisition*. In J. F. Kroll & A. M. B. De Groot (Eds.), Handbook of bilingualism: Psycholinguistic approaches (pp.49-67). Oxford: Oxford University Press.

Mah, J. 2011. Segmental representations in interlanguage grammars: the case of francophones and English /h/. Unpublished doctoral dissertation. Department of Linguisitics, McGill University.

Mahncke, H., W., Bronstone, A. & Merzenich, M. M. (2006) Brain plasticity and functional losses in the aged: scientific bases for a novel intervention. *Progress in Brain Research, 157*, 81-109.

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*(2), 215-234.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly, 34*, 9-34.

Mayberry, R. I. & Eichen, E.B. (1991). The long-lasting advantage of learning sign language in childhood: another look at the critical period for language acquisition. *Journal of Memory and Language, 30,* 486-512.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgement performance by late second language learners. *Journal of Memory and Language, 55,* 381-401.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7,* 703–704.

Molnar, M., Baum, S. Polka, L. & Steinhauer, K. (2009). Automatic auditory discrimination of vowels in simultaneous bilingual and monolingual speakers as measured by the mismatch negativity (MMN). *Journal of the Acoustical Society of America, 125*(4), 2754-2754.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning, 60,* 154-193.

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933-947.

Moyer, A. (2007). Do language attitudes determine accent? A study of bilinguals in the USA. *Journal of Multilingual and Multicultural Development, 28*(6), 502-518.

Näätänen, R. Paavilainen, P. Rinne, T. & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118,* (12), 2544-2590.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I. (2001). "Primitive intelligence'' in the auditory cortex. *Trends Neuroscience, 24,* 283-288.

Neville, H. J., & Bavelier, D. (2000). Specificity and plasticity in neurocognitive development in humans. In M.S. Gazzaniga (Ed.) *New Cognitive Neurosciences,* (pp. 83-96). Cambridge, MA: MIT Press.

O'Brien, I., Segalowitz, N., Collentine J. & Freed, B. (2006). Phonological memory and lexical, narrative and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*(3), 377-402.

Ojima, S., Nakata, H. & Kakigi, R. (2005). An ERP study on second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience 17,* 1212–28.

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language 31,* 785–806.

Otten, L.J. & Rugg, M.D. (2005). Interpreting Event-Related Brain Potentials. In T.C. Handy (Ed.) *Event-Related Potentials: A Methods Handbook* (pp.3-16). Cambridge,MI: MIT Press.

Oyama, S. (1979). The concept of the sensitive period in developmental studies. *Merrill-Palmer Quarterly, 25*(2), 83-103.

Pakulak, E & Neville, H. (2010). Proficiency differences in syntactic processing monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience, 22*(12), 2728-2744.

Pakulak, E & Neville, H. (2011). Maturational Constraints on the Recruitment of Early Processes for Syntactic Processing. *Journal of Cognitive Neuroscience, 23*(10), 2752-2765.

Pakulak, E., Sanders, L., Paulsen, D. J., & Neville, H. (2005). Semantic and syntactic processing in children with high and low grammatical proficiency as indexed by ERPs. *Journal of Cognitive Neuroscience Supplement, 2005.*

Patkowski, M. S. (1982). The sensitive period for the acquisition of syntax in a second language. In S. D. Krahsen & R. C. Scarcella (Eds.), *Issues in*

Second Language Research (pp. 52-63). Rowley, Massachusetts: Newbury House Pub. Inc.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms.* New York: Atheneum.

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain, 121*(10), 1841-1852.

Piske, T., MacKay, I. R. A. & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29* (2), 191-215.

Pulvermueller, F. & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning, 44,* 681-734

Sabourin, L., & Stowe, L. A. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching (IRAL), 43* (4), 329-353.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24* (3), 397-430.

Sanders, L. D., Weber-Fox, C. M. & Neville, H. J. (2008). Varying degrees of plasticity in different subsystems within language. In J. R. Pomerantz (Eds.), *Topics in Integrative Neuroscience: From cells to cognition* (pp.125-153). Cambridge, UK: Cambridge University Press.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech.* Rowley, MA: Newbury House

Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor* (2nd Ed.). Clevedon, UK: Multilingual Matters.

Skehan, P. (1989). *Individual differences in second-language learning*. London: Edward Arnold.

Statistics Canada. (2001). 2001 Census: analysis series. Profile of languages in Canada: English, French and many others. Retrieved May 16, 2012 from http://www12.statcan.ca/english/census01/products/analytic/companion/lang/pdf/96F0030XIE2001005.pdf

Statistics Canada. (2006). Detailed Mother Tongue, Knowledge of Official Languages: Montreal [Census Table]**.** Retrieved May 16, 2012 from http://www12.statcan.gc.ca/census-recensement/2006/dp-pd/tbt/Rp-eng.cfm?TABID=1&LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=759969&GK=0&GRP=1&PID=89200&PRID=0&PTYPE=88971,97154&S=0&SHOWALL=0&SUB=0&Temporal=2006&THEME=70&VID=0&VNAMEE=&VNAMEF=&D1=0&D2=0&D3=0&D4=0&D5=0&D6=0

Steinhauer, K. & Drury, J. E. (2011). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language, 120*(2), 135-162.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13-41.

Steinhauer K., White E., King E., Cornell S., Genesee F. & White L. (2006). The neural dynamics of second language acquisition: Evidence from event-related potentials. *Journal of Cognitive Neuroscience Supplement, 2006.*

Steinhauer, K., White, E. J. & Genesee, F. (2012). Proficiency and first language transfer effects in late second language acquisition: ERP data from a balanced word category violation design. Unpublished manuscript. McGill University, Montreal.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics,* doi:10.1016/j.wocn.2010.09.001

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards, & M.L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153-191). Philadelphia: John Benjamins.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2012). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. Unpublished manuscript. University of Washington, Seattle.

Tokowicz, N. & MacWhinney, B. (2005). Implicit and Explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition 27,* 173–204.

Tyler, W. J., Alonso, M., Bramham , C. R. & Pozzo-Miller, L. D. (2002). From acquisition to consolidation: on the role of brain-derived neurotropic factor signalling in hippocampal- dependent learning. *Learning & Memory, 9,* 224-237.

Rossi, S., Gugler, M. F., Friederici, A. D. & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience 18,* 2030–48.

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. Neuroreport, 9(16), 3556-3560.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition, 4*(1), 105-122.

van Praag, H., Kempermann, G. & Gage, F. H. (2000). Neural consequences of environmental enrichment. *Nature Reviews Neuroscience, 1*(3), 191-198.

Weber-Fox, C.M. & Neville, H.J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231-256.

Werker, J. F. & Curtin, J. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197-234.

White, L. (2003). *Second language acquisition and universal grammar.* Cambridge: Cambridge University Press.

White, L. & Genesee, F. (1996). How native is near native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12* (3), 233-265.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36,* 638-642.