The Case of the Non-native-like First-Language:
Neurophysiological Investigations of First-Language Attrition and Second-Language Processing

Kristina Kasparian
School of Communication Sciences and Disorders
Faculty of Medicine
McGill University, Montreal
June, 2015

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Communication Sciences and Disorders

For my grandmothers, Noemie and Rosine –

my twin pillars of strength and faith.

TABLE OF CONTENTS

Abstract 7
Résumé..9
Contribution of Authors 11
Acknowledgements..13
1. General Introduction...18
  1.1. A "critical period" for second-language acquisition?..19
  1.2. Event-related potentials (ERPs) in language research 23
  1.3. ERP evidence in favor of a critical period in L2 acquisition 29
  1.4. The role of proficiency: Casting doubt on the critical period claim.....................30
  1.5. Measuring proficiency: Inconsistencies in methodology 39
  1.6. The special case of first-language attrition...42
  1.7. Minority-language speakers and first-language attrition......................................43
    1.7.1. The unique linguistic situation of "attriters" 43
    1.7.2. Theoretical accounts of attrition..44
    1.7.3. Disentangling attrition effects from bilingualism effects 46
    1.7.4. Experiential factors affecting attrition 49
  1.8. Towards a neurocognitive account of attrition 52
  1.9. The present studies 57
    1.9.1. Aims and research questions..57
    1.9.2. Description of ERP studies..59
    1.9.3. Additional studies not included in this dissertation 63
2. General Methods 66
  2.1. Participants 66
    2.1.1. Attrition group 66
    2.1.2. Italian second-language (L2) learners 68
    2.1.3. Italian native-speakers...68
    2.1.4. English native-speakers...69
  2.2. Behavioral measures..70
    2.2.1. Language-background measures..70
    2.2.2. Proficiency measures...70

2.2.3. Working memory and reading performance **72**
2.3. Proficiency **74**
2.4. Procedure **76**
2.5. EEG recording and analysis **78**
3. Manuscript 1: *First-language attrition induces changes in online morphosyntactic processing and re-analysis: An ERP study of number agreement in complex Italian sentences* **79**
Abstract **80**
Introduction **81**
The present study **90**
Methods **92**
Results **97**
Discussion **107**
Conclusions **123**
Tables and figures **126**
Table 1 .... **126**
Table 2 .... **127**
Table 3 .... **127**
Table 4 .... **128**
Figure 1a **129**
Figure 1b **129**
Figure 2a **130**
Figure 2b **131**
Figure 3 .... **132**
Figure 4 .... **133**
Figure 5 .... **134**
Figure 6 .... **135**
Figure 7 .... **136**
Figure 8 .... **137**
Figure 9 .... **137**
Appendix **138**
References **140**
4. Bridge 1: *From morphosyntax to lexical-semantics* **149**

5. Manuscript 2: *Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late second-language acquisition* **151**

Abstract **152**

Introduction **153**

The present study **156**

Methods **163**

Results **167**

Discussion **177**

Conclusions **187**

Tables and figures **189**

Table 1 .... **189**

Table 2 .... **190**

Table 3 .... **191**

Figure 1a **192**

Figure 1b **192**

Figure 2 .... **193**

Figure 3a **194**

Figure 3b **195**

Figure 4a **196**

Figure 4b **197**

Figure 5a **198**

Figure 5b **199**

References **200**

6. Bridge 2: *Lexical access and integration in L1 attriters and L2 learners: Within-language and between-language effects?* **208**

7. Manuscript 3: *In My Other Language: An ERP Study of Italian-English Cognates and Homographs in L1-Attrition and L2-Acquisition* **210**

Abstract **211**

Introduction **212**

The present study **219**

Methods **223**

Results **230**

Discussion **246**

Conclusions **256**
Tables and figures **259**
Table 1 .... **259**
Table 2 .... **259**
Table 3 .... **260**
Table 4 .... **260**
Figure 1 .... **261**
Figure 2 .... **263**
Figure 3 .... **264**
Figure 4 .... **266**
Figure 5 .... **267**
Figure 6 .... **269**
Figure 7 .... **270**
Figure 8 .... **271**
Figure 9 .... **272**
References **273**
8. General Discussion **282**
8.1. The role of proficiency level in modulating ERP profiles of language processing **283**
8.2. Parallels between L1 attrition and L2 acquisition **286**
8.3. What is "attrition" and is there evidence of attrition at the neurocognitive level? **288**
8.4. L1 attrition effects observed in the brain versus behavior **291**
8.5. The P600 in L1 attrition and L2 acquisition: Implications for ERP research **294**
8.6. Evaluating the CPH and related claims of reduced neuroplasticity in adulthood **297**
8.7. Future directions **299**
8.7.1. Limitations **299**
8.7.2. New or unanswered questions **301**
8.8. Novelty and implications **302**
9. General References **304**
10. Appendices **321**

ABSTRACT

It is a widespread notion that mastering a second-language (L2) to the level of a native-speaker is more difficult in adulthood than in childhood. For more than half a century, researchers have attributed linguistic differences between "late" L2-learners (who acquire a new language after early childhood) and native-speakers to a neurobiological cause – with maturation, the brain has been argued to lose the plasticity required to change with experience, such that any subsequent language will rely on different brain areas or compensatory processing strategies than those underlying the native-language. Conversely, it is assumed that one's native-language has a privileged and stable status in the brain, as its neural connections are established and hard-wired in the early years of life. However, for the vast majority of L2 learners, the late-learned language is typically confounded with lower exposure, use and proficiency relative to the native-L1, and it is unresolved whether these experiential factors, rather than limitations of neuroplasticity, determine neurocognitive profiles of L2-processing.

The work presented in this dissertation was motivated by this controversial debate and turns to a unique population of bilinguals for a new perspective. First-generation immigrants who move to a new country in adulthood become predominantly exposed to and highly-proficient in the late-acquired L2, while experiencing gradual changes or "attrition" in their L1. Three studies using highly-sensitive event-related brain potenials (ERPs) were conducted to explore two main questions: (1) Do the neurocognitive mechanisms underlying L1-processing remain "native-like" despite attriters' shift in exposure, use and/or proficiency from the L1 to the L2?; (2) Does proficiency shape neurocognitive responses during language-processing, regardless of whether the language being processed is the L1 or L2?

We tested morphosyntactic (Study 1) and lexical-semantic (Study 2 and 3) processing of Italian and English in four groups of speakers: (1) Italian-English immigrants who were highly-proficient and predominantly-exposed to English and who reported negative changes to their Italian; (2) English-Italian late L2-learners who were highly-proficient in Italian; (3) monolingual Italian native-speakers residing in Italy, and (4) monolingual English native-speakers. The specific linguistic phenomena we chose to examine were elements that were either subtle or difficult aspects of grammar and vocabulary, or that were candidates of cross-linguistic competition / transfer between Italian and English.

This dissertation provides some of the first neurocognitive evidence of L1-attrition, although differences from non-attriting native-speakers may be subtle and observable at the brain level before they become overtly apparent in behavior. In showing processing differences in a group of native-speakers who lived in an exclusively monolingual context until adulthood, our findings are compatible with the view that there is ongoing neuroplasticity for language beyond an early developmental period. The possible downside of this plasticity is that the L1 may not be as stable as it is often assumed. The three studies also emphasize the crucial role of proficiency in modulating the brain's responses to language, regardless of age-of-acquisition, and highlight the possibility of parallels between L1 attrition and L2 acquisition. This research also offers insight into characteristics of attrition, which we argue cannot merely be ascribed to proficiency variation. Instead, our results indicate that attriters also engage in more conscious or controlled processes that depend both on experimental properties (i.e., conditions with enhanced linguistic conflict) as well as on experiential circumstances that are inherently part of attrition (e.g., increased attention, more cautious approach, second-thoughts). This work therefore has significant implications for research on the neurocognition of language, and highlights the usefulness of studying language-attrition as a bridge between first- and second-language processing.

RÉSUMÉ

Il est généralement pris pour acquis que la maîtrise d'une langue seconde (L2) au niveau d'un locuteur natif est plus difficile à atteindre à un âge adulte que pendant l'enfance. Pendant plus d'un demi-siècle, les chercheurs ont attribué ces différences entre locuteurs natifs et bilingues tardifs (c'est à dire ayant acquis la langue seconde après la petite enfance) à une cause neurobiologique: Au cours du développement, le cerveau perdrait progressivement de sa plasticité requise pour se modifier selon l'expérience, de sorte que toute langue acquise subséquemment serait basée sur des aires cérébrales ou des stratégies de traitement différentes de celles sous-tendant la langue première. En contraste avec cette hypothèse, il est généralement proposé que la langue première (L1) jouit d'un statut privilégié et stable dans le cerveau, et que ses connections neurales sont fixes et bien établies. Toutefois, pour la majorité des personnes bilingues, la langue seconde s'accompagne d'une exposition, d'un usage et d'une expertise moindres par rapport à leur langue première, et la question de savoir dans quelle mesure ces facteurs liés à l'expérience sont à la source des profiles de traitement de L2 au delà des limites de plasticité cérébrale demeure irrésolue.

Ce débat controversé est au cœur de la présente thèse, laquelle se concentre sur un bassin unique de personnes bilingues pour l'étudier selon une approche nouvelle. En particulier, les immigrants adultes de première génération sont exposés à L2 de façon prédominante et en atteignent un niveau de maîtrise élevé tout en subissant des changements graduels – c'est à dire une « attrition » – dans L1. Nous présentons trois études exécutées selon la méthode très sensible des potentiels évoqués (PÉs) afin d'explorer deux questions de recherche principales, à savoir (1) Les mécanismes neurocognitifs sous-tendant le traitement de L1 demeurent-ils à un « niveau natif » malgré le changement radical d'exposition et/ou de compétence de L1 à L2?, et (2) le niveau de compétence est-il à la sources des réponses neurocognitives de traitement du langage, que celui se passe en L1 ou en L2 ?

Nous avons testé le traitement morphosyntaxique (Étude 1) et lexico-sémantique (Études 2 et 3) de l'italien et de l'anglais dans quatre groupes de locuteurs: (1) des immigrants italo-anglophones auparavant très compétents et exposés principalement à l'anglais mais rapportant des changements négatifs sur leur maîtrise de l'italien; (2) des apprenants italo-anglophones tardifs de l'anglais mais parfaitement compétents en italien; (3) des locuteurs monolingues natifs de l'italien résidant en Italie et (4) des locuteurs monolingues de l'anglais. Les phénomènes

linguistiques particuliers que nous avons choisi d'étudier portaient sur des aspects grammaticaux ou lexicaux subtils ou difficiles, ou à risque d'interférence ou à de transfert entre l'anglais et l'italien.

La présente thèse fournit la première série de preuves neurocognitives d'attrition dans L1, bien que le contraste avec les locuteurs natifs sans attrition soit subtil et observable au niveau cérébral avant de se manifester de façon explicite au niveau comportemental. Au travers de différences de traitement dans un groupe de locuteurs natifs ayant vécu dans un environnement exclusivement monolingue jusqu'à l'âge adulte, nos données sont compatibles avec la notion d'une neuroplasticité linguistique persistant au delà d'une période développementale précoce. L'effet potentiellement négatif d'une telle plasticité est que L1 ne serait pas aussi stable qu'on l'imagine. Nos trois études soulignent le rôle crucial joué par le niveau de maîtrise linguistique dans la modulation des réponses cérébrales liées au langage quel que soit l'âge d'acquisition de la langue seconde, et mettent en lumière les parallèles possibles entre l'attrition de L1 et l'acquisition de L2. Notre recherche fournit aussi des informations importantes quant au phénomène d'attrition proprement dit, lequel ne peut se réduire à des variations au niveau de la maîtrise linguistique. En effet, nos résultats indiquent que les personnes sujettes à l'attrition font appels à des processus plus conscients et contrôlés dépendant à la fois de facteurs expérimentaux (c'est à dire des conditions présentant un conflit linguistique accru) et de circonstances liées à l'expérience faisant partie intégrante du phénomène d'attrition (par exemple une attention accrue, une approche plus prudente ou des remises en question). Le présent travail a donc des conséquences significatives dans la recherche sur la neurocognition du langage et illustre l'utilité d'étudier le phénomène d'attrition linguistique en tant que phénomène transitoire entre le traitement de la langue première et celui de la langue seconde.

CONTRIBUTIONS OF AUTHORS

The three manuscripts in this dissertation were co-authored by Kristina Kasparian (student / first author) and Karsten Steinhauer (supervisor / senior author). The first manuscript was additionally co-authored by Francesco Vespignani, and the third manuscript was co-authored by Francesca Postiglione, both of whom were second authors.

All studies conducted in the framework of this dissertation, including those not reported here, were conceptualized and designed by Kristina Kasparian and Karsten Steinhauer. The main premise and research questions, along with ideas for the specific experiments, were contributed by Kristina Kasparian and further developed under extensive supervisory input from Karsten Steinhauer. All funding that made participant recruitment, testing and hiring of research assistants possible was provided by funds administered to Dr. Steinhauer. The two authors were the primary (and equal) contributors to all data analyses and interpretation, and will be listed as corresponding authors on publications emerging from this dissertation. All data are stored in Dr. Steinhauer's laboratory.

Dr. Vespignani contributed as host supervisor for three months while Kristina Kasparian collected data in Italy. He provided the lab space, equipment, training and input that allowed for data collection and analysis during that phase. Dr. Vespignani was included as a co-author on Study 1 given his substantial contributions to the experimental design and stimuli (largely based on his prior work), as well as on data interpretation. Dr. Vespignani provided valuable feedback on the intellectual content of several drafts of the manuscript.

Dr. Postiglione made a substantial contribution to the conception and design of Study 3 and, together with Kristina Kasparian, created the list of experimental stimuli. Dr. Postiglione also provided the in-progress computerized corpora that were used for all Italian studies that were part of this research project (including those not reported here). Her feedback on the design and stimuli used in other experiments (e.g., Study 2, verb priming study) was also valuable. With respect to Study 3, Dr. Postiglione contributed feedback during the drafting of the manuscript, and her comments were particularly helpful for the Introduction and Methods sections.

All participant recruitment and testing was performed by Kristina Kasparian, with the involvement of research assistants that have been duly acknowledged. All manuscripts were written by Kristina Kasparian and edited by respective co-authors, the largest editorial role being that of Dr. Steinhauer. All co-authors gave their final approval on the manuscript versions that were included in this thesis and submitted to peer-review journals. All studies in this dissertation are considered original scholarship and distinct contributions to knowledge.

ACKNOWLEDGEMENTS

When a research project develops over the course of six years, it comes as no surprise that a multitude of individuals contribute important building blocks along the way. It seems only fair, then, that the acknowledgments section be proportional to the length of the dissertation!

First and foremost, my deepest gratitude goes to my supervisor, Dr. Karsten Steinhauer. Out of all the parallel universes where things would have played out quite differently, I feel extremely fortunate to have decided to share these pivotal years with you. I remember when I initially came to you with an idea – it wasn't perfect, but you saw its potential and let me run with it. Looking back on that now and comparing that initial proposal with the "finished" product is the most striking measure of exactly how much I have learned from you, how very much I have benefitted from your brain and your experience. Thank you for your remarkable commitment not only to science, but especially to the training and development of your students. Thanks for being a source of wisdom and perspective, yet letting me make my own mistakes; for being demanding, yet caring; for pushing me to think independently, yet not being far behind me if I needed your guidance; for continuously saying you trusted me implicitly, for backing up my decisions, and always encouraging me to follow my gut. Thank you for our stimulating meetings, after which I often could not decide if my brain was exhausted or if it was on fire. You developed my strengths and helped work on my weaknesses, and I am convinced I would not be where I am now without your presence. I feel blessed that, for the large part of this long journey, we have been on the same page. I am so grateful for your unwavering support and generosity, both of which were never confined to merely your supervisory role. I will certainly carry forward all that you gave me, as a scientist and as a person. I think, though, that after all this, we have only just *begun* to tame the Monster!

I also owe a special thank you to Dr. Francesco Vespignani. Collaborating with you has been an immense privilege. Your support and insightful feedback was a crucial aspect of this work, whether it came from down the hall or from six hours into the future all the way across the Atlantic ocean. Thank you for hosting me during my research stay in Italy, for welcoming me into your lab (and town!), for giving me a whole new set of lab-mates and colleagues with whom I could discuss my work, and for teaching me how to make savory *fiori di zucca*. Grazie di cuore per tutta la tua generosità e tutto il tempo che hai dedicato al nostro progetto. I look forward to continued collaborations with you!

I am also indebted to all those individuals who gave useful feedback at early stages of our experimental design: Dr. John Drury, Dr. Roy Lyster, Dr. Laura Gonnerman, Dr. Eleonora Rossi, Dr. Francesca Postiglione, Dr. Guillaume Thierry, Dr. Phaedra Royle, Dr. Isabell Wartenburger, Dr. Nicola Molinaro, Dr. Remo Job, Dr. Davide Crepaldi, Dr. Simone Sulpizio, Dr. Manuela Berlingeri, Dr. Isabella Cattinelli and Dr. Michael Ullman.

A warm thank you also to those who, through their writings, teachings and conference-presentations, introduced me to the fundamentals of first-language attrition and welcomed me into a field whose development was already long underway: Dr. Monika Schmid, Dr. Laurie Stowe, Dr. Kees de Bot, Dr. Silvina Montrul, Dr. Ayse Gürel, Dr. Loraine Obler, Dr. Mira Goral, Dr. Simone Sprenger and Dr. Conny Opitz.

To two supporters who played a *huge* role in my academic development prior to my PhD and to whom I will feel forever indebted: Dr. Fred Genesee and Dr. Frank Burchert. Thank you both for the unique academic opportunities you have given me, which served as my foundation and culminated in this moment. Most especially, thank you for your enthusiastic support which allowed me to be awarded a Richard H. Tomlinson Fellowship (McGill) and Vanier Canada Graduate Scholarship (CIHR) for my PhD studies. On a more personal note, thank you both for all the conversations we shared, and all the advice and life-lessons that accompanied your academic support.

To my partner in crime during the *hundreds* of testing sessions in Montreal: Kristina Maiorino. Thank you for being a true source of strength, positivity and perspective, and for always being ready to go above and beyond your obligations. I would probably still be in the data collection phase if it weren't for your help! Thank you for your professionalism, your thoroughness, your insightful ideas which often made me think of my work from another angle, and for making participants feel at ease with your bubbly character and stories about Rome. Our motto was true, in the end: Two Kristinas are better than one.

Another big thank you goes to Paolo Zandomeneghi and Filippo Vicari – my testing partners at my *other* lab in Rovereto, Italy. Thank you for all your help testing and scoring, for all the laughter, the magic tricks, the much-needed aperitivo evenings at *Andreatta* after long days in the lab, and the inevitable "paolofication" process that is still a part of me, years later.

I am also greatly appreciative of research assistants Linna Jin and Lucia D'Arienzo for their meticulous assistance with stimuli creation and in setting up the experimental tasks.

A great big thank you goes to the individuals and organizations that were instrumental in spreading the word about our project and in helping recruit participants: Dr. Francesca Mangiarini (Italian Scientific Community in Canada), Andrea Gianvenuti (Italian Cultural Institute), Pat Buttino (Leonardo Da Vinci Centre) and Bruno Villata (Concordia University). You helped get this project off the ground!

To all our participants, but especially to our bilingual groups of Italian native-speakers and Italian-learners who agreed to take part in not only *one* three-hour testing session but *two*: I am deeply appreciative that all of you – without exception - returned for the second testing session, and with minimal arm-twisting required! Thank you for your enthusiasm and your patience, and for sharing your personal stories and socio-linguistic intuitions with us.

For the last six years, the School of Communication Sciences and Disorders (SCSD) has been my second home and, although we are no longer housed in the historical and beloved "Beatty Hall", the fond memories of my life at SCSD will remain vivid. Thank you to all the Professors and members of the administrative staff both at SCSD and at the Centre for Research on Language, Brain and Music (CRBLM) for all your forms of support over the years. To our Director, Dr. Marc Pell – thank you so much for making me feel like a valued member of the SCSD community, and for your genuine support (and sound advice) over the years. It was a privilege to be enrolled in the School. My SCSD peers – what an incredibly intelligent and wonderful bunch of individuals you all are! I feel blessed to have spent time with you in classrooms and hallways, during our Writing Groups and PhD-days, and to have shared a part of my PhD journey with all of you. I have no doubt that all the world's success awaits you!

To my former and present lab-mates: Dr. Erin White, Dr. Nicolas Bourguignon, Dr. Efrat Pauker, Dr. Inbal Itzhak, Dr. Monika Molnar, Stefanie Nickels, Anastasia Glushko, Fayden Bokhari and Lauren Fromt. I thoroughly enjoyed working with you, sharing an office with you, discussing our work, our triumphs, our doubts and our life goals. Thank you for all the support, feedback, brainstorming, practice talk hours, equipment troubleshooting, as well as all your life-perspective.

My special crew – my "girls": Myrto Brandeker, Katja Häuser, Pan Liu and Aruna Sudarshan. Where would we be without our coffee breaks?! I could not have asked for a better group of "PhD friends" to laugh with, cry with, celebrate every birthday with, and ponder the meaning of it all with ("*it*" not being "*life*", but "*the PhD*"). We are all from different corners of the world and, no matter whether we will return to our respective corners or find new corners to call home, I am convinced that our bond will not dissolve at graduation.

Finally, because the road to a PhD degree is as much a personal journey as it is an academic one, a heartfelt thank you goes to all the key players in my life who selflessly hold their arms out to keep me from falling. You who taught me over these last few years to trust in myself and go with my heart, to leave my comfort zone with courage, to find strength and reassurance in myself rather than to expect it from others, to bloom where I am planted, to just *smile* and be myself, to focus on what I can control and to do my best with the cards I am dealt, to prioritize health and happiness above all other goals, to see life as a mix of sun and rain and to trust in its delicately orchestrated balance, to measure my accomplishments in terms of how far I have come rather than how far I still have to go, to be kind to myself, and to continue – every day – to pursue my dreams. Thank you for believing in me when I sometimes failed to believe in myself. You know who you all are, and you know that your friendship has enriched my life beyond words.

To my endlessly nurturing family who has graciously put up with my scholarly pursuits for *eleven* coffee-filled, deadline-delimited, anxiety-provoking university years – I honestly do not know how you continue to support me in all my endeavors (but please don't stop). My sweet and crazy parents: I know you would give me the moon and that you would have given me a PhD degree long ago if it had been in your control – thank you simply does not cut it. Last but certainly not least – limitless gratitude is reserved for Mister. I don't know whether to thank you or apologize! Thank you for your unconditional support and patience, and for rarely ever complaining about our much-too-late suppers, my much-too-frequent travels, and for my bringing work (and all associated feelings) home with me every day. Thank you for sitting in the front row at conference presentations and smiling ear to ear, for inventing your own ERP components because you can't remember the names of the real ones (RX2000?), for listening to me hash out arguments or practice my talks, and for happily mingling with my academic peers without ever feeling out of place. As I watch you explain my research to people we know, I have come to realize that you deserve an honorary PhD degree for having stood by my side through all this, and for having effortlessly integrated yourself into this important part of my life! Thank you for knowing me, for believing in me and for making sure I never lose that precious feeling that the world is our oyster and that nothing is impossible.

Funding acknowledgements

I wish to express my gratitude for the financial support I received from the Vanier Canada Graduate Scholarship (Canadian Institutes of Health Research), the Richard H. Tomlinson Fellowship (Faculty of Medicine, McGill University), the Graduate Scholar Stipend (Centre for Research on Brain, Language and Music) and the Michael-Smith Foreign Study Supplement (CIHR). In addition, numerous conference presentations were supported by travel grants from the CRBLM and the SCSD Abelson Fund.

Our research was made possible thanks to grants awarded to Dr. Karsten Steinhauer by the Canada Research Chair program and the Canada Foundation for Innovation (CRC/CFI; project # 201876), the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPGP 312835-05 and RGPIN 402678-11), the Social Sciences and Humanities Research Council of Canada (SSHRC; # 435-2013-2052), and the Fonds de Recherche Société et Culture, Québec (FQRSC; # 2010-SE-103727). Thank you to all reviewers and organizations who saw the value of this research and chose to support it.

1. GENERAL INTRODUCTION

Multilingual speakers worldwide would intuitively agree that learning a language to the level of a native-speaker is a more difficult feat in adulthood than in childhood. While there may be no doubt that a new language becomes more challenging to master with advancing age, whether or not this is due to a maturationally-constrained window in our neurobiological development has been a highly controversial issue for over half a century. A key notion that has received much attention in language research is that fundamental differences exist between how first- (L1) versus second- (L2) languages are organized and processed in the brain if L2-learning occurs beyond a "*critical-period*" for language-learning (Lenneberg, 1967; Penfield & Roberts, 1959). It has been argued that this period is limited to the early years of childhood, after which the brain loses the *plasticity* necessary to adapt with experience. The neurocognitive substrates set for the L1 are therefore believed to be unavailable for any L2 learned at a "late" *age-of-acquisition* (AoA), resulting in learners' reliance on different brain areas or compensatory processing strategies, and thus culminating in non-native profiles of language-processing. The corollary of this view is that the L1 has a privileged status, as it was acquired from birth and hard-wired in the brain within this maturationally-constrained window, and is therefore expected to remain stable once brain networks for language are no longer plastic (Marchman, 1993; Penfield, 1965).

Supporting this theory, a number of studies have shown that "late L2-learners" deviate from native-speakers in the way their brain represents and processes the L2, compared to native-speakers or early-learners of the language. However, a major shortcoming is that the late-acquired L2 is typically also the language that speakers are less exposed to, less experienced with and less proficient in relative to the L1, therefore making it difficult to determine whether differences observed in neurocognitive processing mechanisms in late L2 learners are due to the late age at which the language was acquired, or rather to the low level of mastery and low exposure to that language. In recent years, an increasing number of studies have advocated for ongoing neuroplasticity in adulthood and have emphasized the crucial role of proficiency level in modulating the brain's responses to language. However, it remains a controversial empirical question whether (and to what extent) it is AoA that crucially shapes the way the brain processes the L2, or whether other experiential factors that are typically confounded with AoA (such as proficiency level, amount of exposure, typological similarities or differences, etc) are stronger predictors of "nativelikeness".

The present dissertation turns to a unique population of bilinguals for a novel perspective on the neuroplasticity debate – first-generation immigrants who move to a new country and become immersed and dominant in the L2 that is the majority language of their new environment, while experiencing significantly reduced exposure to their native (minority) L1. After some time spent in the new environment, these speakers report changes or difficulties in their L1 (a phenomenon defined as "*attrition*"), whereas their proficiency level in the late-learned L2 continues to strengthen. First-language "*attriters*" shed new light on the question of a critical period vs. ongoing neuroplasticity for language in adulthood in two ways. First, by examining L2-processing in these individuals, we are able to assess the differential contribution of AoA, proficiency level, and exposure in a unique L2-learning situation where the late-acquired L2 is the predominantly-used language (whereas exposure to the L1 is limited or severed), and where proficiency in the late-acquired L2 is not by default lower than proficiency in the native-L1. Secondly, examining their *L1* allows us to determine whether their processing profiles remain native-like despite their shift in exposure, dominance and proficiency towards the L2, or whether differences from native-speakers can be detected at the brain level, in a situation where AoA plays no role. Attriters also allow us to elucidate the role of proficiency level in modulating not only L2- but also *L1*-processing. The study of L1-attrition therefore constitutes a logical and necessary bridge between L1 and L2 acquisition research, and may reveal that attrition and acquisition lie on the same continuum, with proficiency level influencing the nativelikeness of neurocognitive correlates of language, regardless of whether the language being processed is the L1 or L2.

Using highly-sensitive *electrophysiological (ERP) measures* of brain activity, this work is among the first neurophysiological investigations of L1-attrition, examining attriters' linguistic abilities and processing patterns in a range of lexical and grammatical features of both their L1 *and* their L2, compared to native-speakers of each language, as well as to a bilingual control group of highly-proficient late L2 speakers of the same pair of languages, but with opposite AoA profiles.

1.1. **A "Critical Period" for second-language acquisition?**

The "*Critical Period Hypothesis*" (CPH; Lenneberg, 1967; Penfield & Roberts, 1959) suggests that outside of a limited biologically-programmed time-window in the early years of development, the ability to acquire language at the proficiency-level of a native-speaker declines.

The strong version of this hypothesis posits a firm endpoint of the critical period – traditionally assumed to be puberty – which marks a turning-point in language acquisition outcomes.

It is important to distinguish between *critical-period* effects and AoA-effects that are *not* argued to be due to maturational limits but to those environmental circumstances which typically go hand in hand with learning a language at an older age compared to early childhood, such as differences in the amount and type of exposure (e.g., explicit teaching compared to implicit learning through immersion), having less time to practice, etc. The views that this dissertation aims to address are those which posit that age-effects are due to a progressive loss of plasticity of the neural substrates underlying language processing, as a result of brain maturation (Lenneberg, 1967, see Harley & Wang, 1997 for a review). A variety of related claims have explained AoA-effects in this vein. According to the *"Fundamental Difference Hypothesis*" (Bley-Vroman, 1989), the brain areas used to represent and process the L1 become unavailable beyond the critical period, and different, compensatory neural substrates must be called upon for representing and processing the L2. A similar claim that L1 and L2 acquisition differentially rely on distinct cognitive processes has been put forward by Paradis (2003) and by Ullman (2001). According to these researchers, cognitive changes occur with age and result in a shift in the relative use of two different memory systems in the brain, leading to observable differences in late-learners compared to early-learners. Whereas native-speakers rely on the *declarative or explicit memory system* for semantic knowledge and on the *procedural or implicit memory system* for grammatical knowledge, adult L2 learners instead largely rely on declarative memory, even for grammatical processing. However, unlike the strong version of the CPH, this view does not necessarily advocate for a termination of the ability to rely on procedural/implicit memory after a limited period early in life. Rather, late-learners' reliance on the procedural/implicit memory system could increase with greater practice and proficiency in the language (Ullman, 2001; White et al., 2007).

Another account of the critical-period is that it is the process of learning the L1 itself that stabilizes the neural connections and reduces brain plasticity. According to this "*neural commitment*" or "*entrenchment*" hypothesis, the more fully-developed the L1 system in the brain, the less these language areas can be modified by exposure to a later-acquired L2, and, consequently, the more strongly the L1 is expected to affect the L2. This cross-linguistic influence may either be positive, in terms of the L1 facilitating L2 acquisition, or the effects may be negative, if there is competition and interference between the two languages. Thus, still related to the notion of brain plasticity, it has been suggested that it becomes more difficult to learn a language with increasing age, but that this is directly a result of L1 acquisition having fixed the neural connections that are also involved in subsequent L2 acquisition (Hernandez, Li & MacWhinney, 2005; Kuhl, 2004; Marchman, 1993; Penfield, 1965).

Behavioral evidence in support of a critical period for L1 acquisition comes from rare cases of extreme social isolation (Curtiss, 1977; Davis, 1947). These studies have shown that when exposure to language is substantially delayed, ultimate attainment of that language is irregular and incomplete, and no amount of subsequent experience can compensate for the initial delay (Long, 1990). However, such cases of late L1 acquisition in linguistically-isolated children cannot be taken as evidence in favor of a critical period, as these children also suffered sensorial, social and affective deprivation, the cognitive consequences of which may have been considerable (Harley & Wang, 1997, as cited in Köpke, 2004). A more compelling source of evidence comes from cases of congenitally deaf children who are born to hearing parents and, therefore, are exposed to sign language at varying ages in childhood, but not in infancy. These children experience delayed L1 acquisition but grow up in otherwise normal circumstances. Studies have shown that children who were exposed to sign language in their first years of life performed significantly better in both comprehension and production than children who were only exposed to sign language in mid-childhood or later (Boudreault & Mayberry, 2006; Mayberry, 1993; Mayberry & Eichen, 1991; Mayberry & Lock, 2003; Mayberry, Lock & Kazmi, 2001).

Although most researchers agree on the drastic impact that delayed language exposure has on acquiring a *first* language (i.e., in establishing the cognitive foundation for language and communication), the existence of a critical period and the (im)possibility of “native-like” attainment are controversial notions in the acquisition of a *second* language. Numerous studies in L2 research have favoured the CPH in their reports that AoA is a major determinant of ultimate proficiency, with late L2 acquisition resulting in important differences from native speakers in phonology, morphology and syntax (De Keyser, 2003; Flege, 1991; Johnson & Newport, 1989; Meisel, 1997; Oyama, 1976). It has also been suggested that AoA differentially affects various aspects of L2 skills; achieving native-sounding pronunciation is typically notoriously difficult in late-learners, as is the acquisition of complex, subtle aspects of grammar, whereas lexical and semantic knowledge may be less affected by AoA (see Newport, Bavelier & Neville, 2001). Based on this, researchers have advocated for "*multiple critical periods*" differentially affecting sub-domains of language (Eubank & Gregg, 1999; Long, 1990; Walsh & Diller, 1981). In sum, behavioral evidence has suggested that unlike early L2 learners, late-learners' proficiency level in a language acquired after puberty typically falls short of being "native-like", at least in some aspects of their linguistic abilities.

In contrast, while the studies cited above have focused on documenting the variable success of L2 acquisition and on demonstrating a negative correlation between AoA and linguistic ability, other researchers have aimed to *falsify* the CPH by showing that native-like attainment is indeed possible in late L2 learners (Birdsong, 1999; White and Genesee, 1996). According to Long (1990), one single post-critical-period L2 learner with language abilities indistinguishable from those of native speakers would be sufficient to reject the strong view of the CPH or any related hypothesis postulating neurobiologically-programmed maturational constraints. This view has also been debated, as it has been argued that such few individuals cannot be considered as counterevidence to the CPH, but rather as exceptions to the rule. In this context, Selinker (1972) suggested that these successful learners are so unique (only approximately 5% of late L2 learners) and make use of such different psychological processes in their language learning that "[they] may be safely ignored" and excluded from L2 theory construction.

Still others argue that even those L2 learners which *appear* to be indistinguishable from native speakers in their language performance are not truly "native-like" if one scrutinizes their linguistic abilities with the use of sophisticated measures and more demanding tasks than are conventionally employed in most of the L2 literature. Pioneers of this view are Hyltenstam and Abrahamsson (2000, 2003; see also Abrahamsson & Hyltenstam, 2009). These researchers argue that findings of successful "native-like" L2 acquisition have either been based on language tests that have been much too easy and which have involved simple structures that are not notoriously difficult for L2 learners, or because the language performance data have not been analyzed in sufficient detail (Abrahamsson & Hyltenstam, 2009). These researchers make a distinction between "perceived nativelikeness" and "scrutinized nativelikeness". In a large-scale study, Abrahamsson and Hyltenstam (2009) assessed the language abilities of Spanish-Swedish bilinguals with different AoAs (< 1-47 years) and revealed that only a handful of bilinguals who had acquired Swedish after the age of 12 were rated as "native-like" in their language production by native-speakers of Swedish. However, when the linguistic abilities of these L2 learners were scrutinized in detail with a battery of 10 complex and cognitively demanding tasks (such as VOT perception/production, processing language in babble noise / white noise, written/auditory grammaticality judgment tasks, and familiarity with idioms/proverbs), *none* of the late-learners performed within the native-speaker range. Furthermore, the results also showed that only a few of the early-learners (AoA < 12) exhibited actual native-like abilities on all measures of L2 proficiency. Based on these findings, Abrahamsson and Hyltenstam argue that native-like ultimate attainment in an L2 is, in principle, never attained by post-critical-period learners and is, moreover, much less common among child-learners than was previously believed.

In sum, the controversial debate about whether or not there is a critical period for L2 acquisition is still unresolved. However, the extreme theoretical viewpoint that there is a decisive age cut-off marking the abrupt termination of the limited window for optimal language development is no longer favored, and current researchers tend to discuss critical- or sensitive-period effects in terms of a gradual decline in language outcomes with increasing age-of-acquisition (Birdsong, 2006; Harley & Wang, 1997). It is still in question whether differences in language outcomes in older learners are in fact due to limitations on brain plasticity, or other experiential factors whose importance has been considerably downplayed in the research to date.

1.2. Event-related-potentials (ERP) in language research

As the CPH and related views such as the "*Fundamental difference hypothesis*" are neurobiological theories claiming neuroanatomically- and neurofunctionally-based differences between L1 and (late) L2 learners, it seems logical to turn to cognitive neuroscience to shed light on these issues (Doughty and Long, 2003; Eubank and Gregg, 1999, as cited in Steinhauer, White & Drury, 2009). Whereas behavioural methods focus on "*ultimate attainment*" and on whether linguistic *performance* is native-like, neurocognitive methods focus on the *underlying mechanisms* involved in L2 learning and processing. Do L2 learners (of varying AoAs and proficiency levels) process the L2 in the same way as native-speakers, or do they engage in qualitatively different and compensatory processes?

In recent years, neuroimaging and neurophysiological methods have begun to contribute to the critical-period debate by examining whether the brain areas and cognitive mechanisms underlying L1 processing are still available and used for processing the L2, and whether the overlap between L1 and L2 processing in the brain depends on age of L2 acquisition (Abutalebi et al., 2001; Kim et al., 1997). As the studies comprising this dissertation are neurophysiological in nature, the overview of the literature will focus almost exclusively on event-related brain potential (ERP) research, except where other imaging methodologies (e.g. fMRI, PET) have made important contributions to the debate.

Event-related brain potentials (ERP) measure the real-time electrophysiological dynamics of cognitive processes in the brain. Due to their high temporal resolution (in the range of milliseconds), ERPs are particularly useful to examine the temporal dynamics of language-related processes as they unfold in time. In contrast to behavioral methods which are measured at discrete points in time, a continuous measure of brain activity such as ERPs is able to tap into implicit linguistic processes and may reveal more about language processing than behavioral tasks alone.

Studies examining the ERP correlates of language processing – either in an L1 or L2 – have traditionally adopted the "*violation paradigm*"; in test sentences, *violations* (ungrammaticalities or anomalies that do not fit the sentence context) become evident at a specific target word, and responses to these sentences are directly contrasted with responses to matched control sentences that do not contain a violation. The rationale for using this paradigm is that violations should disrupt or increase the workload of the processing mechanisms of interest (Kotz, 2009). The standard task in these studies is an acceptability judgment task where the participant decides at the end of the sentence whether it is an acceptable sentence or not, while the brain's electrophysiological activity continues to be recorded. Certain types of linguistic violations engage particular systems or processes in the brain and elicit specific ERP components, which vary in polarity, latency and scalp distribution (Coles & Rugg, 1995). By examining the ERP responses elicited in late L2 learners by the presentation of specific correct or incorrect structures, one can determine which neurocognitive processes they engage in, and how these processes may deviate from those occurring in native speakers. Such deviations could either be described in terms of 1) the time-course (onset and duration) of ERP responses, 2) the presence, absence or amplitude of certain effects, or 3) in terms of approximately which areas in the brain are associated with these responses.

In native-speakers, difficulties in lexical and conceptual-semantic processing have been consistently shown to elicit an N400 effect – a centro-parietal negativity that peaks at approximately 400 ms after the presentation of a target word (Kutas & Hillyard, 1980). The N400 was first observed in response to an end-of-sentence target word that was incongruous with the preceding context (*e.g., He spread his warm bread with* *<u>socks</u>; Kutas & Hillyard, 1980) and has since then been reliably and crosslinguistically reported to occur in response to semantic anomalies occurring within a sentence. The N400 has been associated with the declarative memory system, rather than with procedural memory (Ullman 2004; see Bowden et al., 2014). Its *amplitude* reflects costs of lexical activation and/or semantic integration, and has been shown to be affected by factors such as word class (e.g., Münte et al., 2001), word frequency (e.g., Van Petten & Luka, 2006), and the degree of contextual constraint provided by the sentence (e.g., Connolly et al., 1992), with a larger N400 amplitude reflecting increased difficulty in integrating the target word into the current context (Van Berkum, Hagoort & Brown, 1999; Chwilla, Hagoort & Brown, 1998; Holcomb, 1993).

In some cases, lexical-semantic anomalies also elicit a centro-parietal late-positivity observed between 600 and 900 ms (*P600*). The P600 is typically associated with controlled processes of integration, reanalysis and repair in *morphosyntactic* processing (Osterhout & Holcomb, 1992). However, in recent years, this clear dichotomy between "semantic N400" and "syntactic P600" has increasingly been challenged by several studies reporting late-positivities elicited in response to lexical-semantic anomalies (Kim and Osterhout, 2005; Kuperberg, Sitnikova, Caplan, Holcomb, 2003; Kuperberg, 2007). Given the replicability of these findings, researchers have extended the "morphosyntactic reanalysis" account of the P600 to a more *general* process of re-attending and reanalysis of problematic sentences, where expectations about upcoming words are disconfirmed (De Long et al., 2011; Federmeier et al., 2007; Kolk et al., 2003; van Herten et al., 2005; 2006). In this vein, the "*monitoring theory*" (Van de Meerendonk et al., 2009), proposes that speakers engage in a monitoring process during online language comprehension; when conflict arises between what is expected and what is encountered, the language system is brought into a state of indecision, and re-analysis is triggered in order to check the input for potential perceptual errors. The stronger the implausibility, the stronger the conflict and the re-analysis process that follow. When mild conflicts are resolved successfully, an N400 effect is elicited. However, when the stimulus is deeply implausible, a stronger conflict arises, integration fails and re-analysis becomes necessary, thus giving rise to the P600 effect (Vissers, Kolk, Van de Meerendonk and Chwilla, 2008; Kolk and Chwilla, 2007; Vissers, Chwilla & Kolk, 2006; 2007). The N400 and P600 components in response to lexical-semantic anomalies will be examined in detail in Study 2, which investigates lexical-semantic processing in Italian, as well as in Study 3, examining lexical-semantic processing in English.

Morphosyntactic violations have been found to elicit two distinct ERP components said to reflect two stages of morphosyntactic processing. The first component is known as the "*left-anterior negativity"* (LAN) and has been associated with automatic rule-based parsing taking place during the early stages of sentence processing (Hahne & Friederici, 1999). It is typically elicited by word category violations (e.g., *He criticized Max's *of proof the theorem; Weber-Fox & Neville, 1996*) and violations of subject-verb agreement (e.g., *The children *plays in the garden*; Friederici, 2002; Neville et al., 1991). Variability has been reported regarding its distribution, with some reports of a sustained bilateral-anterior topography (ANs; Hagoort, Wassenaar, & Brown, 2003) and other findings of a left temporal-parietal distribution (LTNs; Neville et al., 1991). As discussed by Steinhauer (2014), it is unresolved whether LANs, ANs and LTNs reflect similar underlying cognitive processes, and whether some of these distributional differences may be related to the modality of the stimuli (written or auditory).

Variability has also been reported in terms of the LAN's peak latency, with an early effect reported by some researchers around 100-250 ms (termed "*early left anterior negativity*" or *ELAN*) and later effects occurring around 300-500ms (LAN). Some have claimed that the ELAN is distinct from the LAN and is associated with very early and highly automatic processing of word category information (Hahne & Friederici, 2001). Recently, however, serious doubts have been cast on the existence of the ELAN, as its presence has been attributed solely to methodological limitations such as inadequately matched target words and baseline problems arising from differences in pre-target contexts between conditions. If the words preceding the target word differ between a violation condition and its correct control condition, it is likely that ERP differences on the target word actually stem from differences elicited by the *preceding* context rather than the processing of the target word itself (Steinhauer & Drury, 2012). Indeed, in studies reporting ELAN effects (Hahne & Friederici, 2001; Rossi et al., 2006), the negativities that have been termed ELANs by the authors actually occur *earlier* than the onset of the violation, clearly demonstrating that these differences must have been due to the context (for a detailed discussion, see Steinhauer & Drury, 2012; see also Steinhauer & Connolly, 2008). A recent study by White, Genesee & Steinhauer (2012) which innovatively circumvented context-driven artifacts by creating a balanced design consisting of *two* correct control conditions and *two* violation conditions, and completely counterbalancing target words across sentences, failed to replicate the ELAN effect (see also Bowden et al., 2014; Steinhauer & Drury, 2012; White at et., 2012). Meticulous stimuli construction and the creation of balanced experimental designs was an approach we took to heart in the experimental studies that make up this dissertation.

The second component elicited in morphosyntactic violations is the P600, which has been found to occur more reliably than the LAN in response to different types of grammatical violations and across many languages (Kuperberg et al., 2003; Osterhout & Mobley, 1995; Sabourin & Stowe, 2008). As mentioned above, it is typically associated with more elaborate processing, such as having to re-analyze or repair the sentence to arrive at its correct interpretation (Kaan et al., 2000; Osterhout & Holcomb, 1992). In contrast to the LAN, which has been argued to be early and automatic (Friederici, 2002), the P600 has been claimed to reflect controlled and attention-driven aspects of sentence processing (Hahne & Friederici, 1999). The P600 has been found to occur in response to syntactically ambiguous sentences and garden path sentences that require a re-parsing of the sentence (Osterhout & Holcomb, 1992; Osterhout et al., 1994) as well as in response to various types of morphosyntactic violations that require repair (*e.g.* verb agreement; verb, case, and pronoun inflection, verb argument), with larger P600s being associated with costlier repair (e.g., Carreiras, Salillas & Barber, 2004; Hagoort & Brown, 2000; Mancini, Vespignani, Molinaro, Laudanna & Rizzi, 2009; Molinaro, Vespignani et al., 2008; Silva-Pereyra & Carreiras, 2007).

Often, though not always, a biphasic ERP pattern consisting of the LAN followed by the P600 is elicited in response to morphosyntactic anomalies. However, it is important to bear in mind that ERP components may overlap over time. Osterhout (1997, 2004) has questioned the reliability of the LAN as an index of morphosyntactic violation detection that is distinct from the N400 and has instead argued that the LAN is what is "left over" after a temporally-overlapping right-lateralized positivity (P600) cancels out the broadly-distributed N400 (Tanner & Van Hell, 2014; Steinhauer, 2014). A second issue that has been highlighted by several researchers (Nieuwland & Van Berkum, 2008; Osterhout 1997; Osterhout et al., 2004; Tanner & Van Hell, 2014) is that, while a group average may reveal a LAN+P600 pattern, such a biphasic pattern may be a byproduct of individual differences in participants' inherent processing strategies, with individual datasets showing *either* an N400 or a P600, rather than a true biphasic pattern (Nieuwland & Van Berkum, 2008; Osterhout 1997; Osterhout et al., 2004; Tanner & Van Hell, 2014). This finding highlights the importance of investing time in the exploration of individual datasets in order to fully understand the group pattern – an approach we have taken in the current dissertation.

Another ERP component that may be relevant to the data presented in the current studies is the P300 – a component that has been established as a reliable response to unpredictable stimuli across modalities. Researchers have identified two subcomponents with different scalp distributions that make up the larger P300-component: (1) a *frontally-maximal positivity* (P3a) elicited when perceptually novel stimuli are interspersed among more frequent stimuli, and (2) a *parietally-maximal positivity* (P3b) whose amplitude has been shown to correlate with the relevance of the stimuli to the specific task the participant has been instructed to perform (Courchesne et al., 1975; Donchin, 1981; Dien et al., 2004; see Van Petten & Luka, 2012). Positivities in the P300-family have been argued to reflect cognitive processes underlying the disconfirmation of an expectation ("surprise" or "orientation response"), and to be associated with a shift in attention and the updating of working memory (WM) during processing (Donchin & Coles, 1988). In the linguistic domain, several studies examining syntactic ambiguity resolution (e.g., garden-path sentences, object relative-clauses compared to subject-relative clauses) have revealed that encountering a disambiguating part of a sentence that forces a less-preferred syntactic analysis elicits a P300-like positivity, even if these unpreferred sentences are grammatically well-formed (Mecklinger et al., 1995; Osterhout & Holcomb, 1992; Osterhout et al., 1994). The amplitude of these positivities has been shown to positively correlate with the processing load (Osterhout et al., 1994). Currently, it remains unclear whether the P600 is functionally and neurally distinct from the P300-family, or whether, at the very least, it may sometimes comprise P300 components, depending on the nature of the stimuli and the experimental task (see Steinhauer & Connolly, 2008). The P300 is relevant to our series of studies, as we may expect a P3a in response to sentence stimuli where strong syntactic expectations are disconfirmed and where working memory must be updated, such as in the processing of long-distance agreement relationships (Study 1 in Italian) or of non-canonical relative clauses (not included in this dissertation).

The distinction between the prototypical (posterior) P600 and an earlier (more fronto-central) positivity has not consistently been made, nor has the earlier fronto-central positivity consistently been termed a P3a (rather than an "early P600"). The majority of studies typically quantify the P600 only in one time-window. In contrast, supporters of the claim that the P600 is not a monolithic component have taken modulations in scalp distribution and timing as evidence that different positivities reflect distinct neurocognitive processes. This is the approach we have taken in Study 1. Further details on ERP profiles of lexical-semantic and morphosyntactic processing (N400, LAN, P600, P300) will be provided in the context of our individual studies.

1.3. ERP evidence in favor of a critical-period in L2 acquisition

Early neuroimaging (fMRI, PET) studies argued in favor of the CPH and demonstrated that a late age of acquisition (AoA) resulted in fundamental differences between L2 learners and native speakers in the neurocognitive mechanisms involved in language processing. In a pioneering fMRI study, Kim and colleagues (1997) investigated whether the neural substrates underlying L1 and L2 processing were overlapping or separate in early and late L2 learners. During a silent production task, late L2 learners were shown to activate two distinct, adjacent centers in frontal brain regions (Broca's area) for the L1 and the L2. Based on these findings, it was concluded that AoA determines the functional organization the bilingual brain, and that after the end of the critical period, Broca's area is no longer plastic enough to accommodate a new language, and adjacent brain areas must be used for processing the L2.

Neurophysiological studies have also highlighted AoA as a critical determinant of native-like language processing. A seminal ERP study by Weber-Fox and Neville (1996) explored semantic and syntactic processing in five groups of Chinese L2 learners of English, differing in AoA (1-3, 4-6, 7-10, 11-13, > 16 years). In response to semantically anomalous sentences, all L2 groups, regardless of AoA, elicited an N400 typical of native-speakers, although this effect was delayed in the older AoA groups (11-13 and > 16). The authors therefore suggested that semantic processing largely relies on the same neurocognitive substrates in late L2 learners as in native-speakers, although lexical retrieval or semantic integration processes may be slightly slower in L2 learners with higher AoAs. In contrast, L2 syntactic processing showed an even more profound divergence due to late AoA and was associated with substantial differences in latency, amplitude as well as distribution of the syntactic LAN + P600 components. In response to phrase structure violations, L2 learners with an early AoA (< 11 years) showed a LAN-like negativity similar to that which is typically observed in native speakers. Late L2 learners, however, revealed a negativity which was either bilaterally distributed (AoA: 11-13) or greater over the right hemisphere (AoA > 16). Therefore, although all L2 groups showed this early negativity, only the youngest AoA groups elicited a LAN which was similar to that of native-speakers. Furthermore, the P600 in learners with an AoA under 11 years was similar to the P600 elicited in native speakers; however, it was delayed in the 11-13 group and absent altogether in the learners who acquired English after 16 years of age. These findings were taken as evidence in support of the CPH and that the recruitment of neural mechanisms underlying language processing is only native-like when the L2 is acquired at a young age.

Similar results were reported by Hahne and Friederici (2001) in their study with Japanese late learners of German (AoA = 21). Whereas lexical-semantic processing was not found to be significantly affected by the learners' late AoA, as indexed by an N400 similar to native-speakers, important differences between the late L2 learners and native-speakers were revealed in the domain of syntactic processing. In response to word category violations, late L2 learners failed to show either the early negativity or the P600 effect, but instead showed a greater P600 for correct sentences. This finding was in line with the claim that late L2 learners engage in less automated language processing, have limited access to implicit, procedural memory systems (Ullman, 2001), and ultimately show difficulty with syntactic integration (Osterhout & Holcomb, 1992).

Taken together, these studies supported the existence of a critical period for L2 acquisition and also demonstrated that AoA does not seem to affect all aspects of L2 acquisition to the same extent. Rather, different linguistic domains appear to display different critical periods, with lexical-semantic processing being less vulnerable to AoA-effects than syntactic processing.

1.4. The role of proficiency: Casting doubt on the Critical Period claim

The picture may have appeared to be rather straightforward had it not been for one major shortcoming in these early neurolinguistic studies: in all the studies described above, late L2 acquisition is confounded with a low proficiency level in the L2, making it impossible to determine whether non-native-like ERP profiles are due to the learners' late AoA or, rather, to their insufficiently advanced command of the language.

In the Weber-Fox and Neville study, for example, AoA was negatively correlated with L2 proficiency level; the later the age at which learners acquired the L2, the lower their proficiency level, as measured with standardized tests of English grammar, self-reported proficiency ratings, and accuracy in the grammaticality judgment task during ERP recording. Thus, based on these studies, it is impossible to disentangle the effects of AoA and proficiency level on the neurocognitive correlates of L2 processing, and to determine whether individuals who *are* capable of reaching very high levels of L2 proficiency are indeed indistinguishable from native-speakers in their brain activation patterns, or whether differences in the brain persist even at high levels of proficiency, thus supporting the claim for neurobiologically-programmed maturational effects.

The influence of L2 proficiency on the recruitment of native-like neural mechanisms during language processing was first revealed in an fMRI study by Perani and colleagues (1998), who showed that the brain activations of high-proficiency L2 learners overlapped with those of native-speakers, even if they had acquired their L2 after puberty. Taken together with similar findings from subsequent neuroimaging studies (Chee et al., 1999; see Perani & Abutalebi, 2005 for a review), it has been argued that the same network of brain areas underlying L1 acquisition remain plastic and available for processing the L2. However, although neuroimaging methods such as fMRI or PET are characterized by high spatial resolution, their temporal resolution is poor (unlike ERPs) and, as a result, these methods cannot reveal much about the specific psycholinguistic sub-processes that unfold in real-time, and which, in turn, may be differentially dependent on proficiency level and/or AoA.

The first ERP study to address the AoA/proficiency confound was one by Friederici, Steinhauer and Pfeifer (2002) in which they employed an artificial miniature language paradigm. Though limited in scope, artificial languages are useful experimental paradigms as these miniature languages can be learned up to high proficiency levels, allowing researchers to investigate the effect of increasing proficiency within the same set of participants within a relatively short period of time, in contrast to real-life longitudinal studies. In addition, they allow for researchers to control for potential confounds inherent in L2 acquisition (e.g., amount and context of exposure). Native-speakers of German (mean age = 21) were trained on an artificial language called *Brocanto* with a simple but highly-controlled grammar and a fourteen word vocabulary. Learners were trained until they reached a high level of proficiency and were compared to a control group of German speakers who were only trained on vocabulary (symbol-word mappings) and not grammar. In the trained participants, syntactic violations elicited a LAN followed by a P600, similar to effects reported for native-speakers in natural languages. Conversely, the control group without grammatical training did *not* detect the ungrammatical sentences and failed to show a LAN/P600 response pattern. Though conducted using an artificial language, this study challenged the CPH by demonstrating that individuals who attain high levels of proficiency in a language *do* engage in brain mechanisms that are similar to native-speakers, even if this language was acquired in adulthood.

The findings from the Brocanto study were successfully replicated in studies with natural languages. Rossi and colleagues (2006) conducted an auditory ERP study with German-Italian and Italian-German bilinguals who had acquired the L2 after the age of 10 and had attained either high or low L2 proficiency levels. Despite their late AoA, high proficiency L2 learners of both languages displayed a native-like LAN effect followed by a P600 in response to all three kinds of morphosyntactic violations tested (word category violations, number agreement violations and combined violations), although some amplitude differences were found. The low proficiency L2 learners, in contrast, showed qualitative differences from native-speakers for morphosyntactic violations, as the LAN was absent and the P600 effect was significantly smaller and delayed. The authors argue that these findings do not fit with the CPH or any other claim advocating for fundamental differences in the neural underpinnings of language processing due to AoA. Rather, native-like ERP patterns seem to be driven by L2 proficiency level. Despite the methodological limitations of this study and the need for the very early negativities (ELANs reported in response to word-category violations even in low-proficiency L2 learners) to be reinterpreted as contextually-driven artefacts (as discussed above), Rossi and colleagues' findings remain important for having shown, for the first time in an ERP study with natural languages, that adult L2 learners *are* able to elicit the same ERP response patterns as native-speakers, provided that they are sufficiently proficient in the language, even if they acquired it in adulthood.

The impact of L2 proficiency level was also explored by Steinhauer, White and Drury (2009) who examined morphosyntactic processing in French and Chinese late learners of English

(AoA of 15 and 19 years, respectively) categorized either as high- or low-proficiency speakers on the basis of a cloze test. In native-speakers of English, syntactic word category violations (e.g. *The man hoped to *meal the enjoy with friends*) elicited a biphasic LAN/P600 pattern. Low-proficiency L2 learners failed to show an early LAN and only showed a P600 effect, in line with previous findings that low-proficiency late learners are unable to engage in automatic and implicit grammar processing (Hahne & Friederici, 2001). On the other hand, the biphasic LAN/P600 pattern in high-proficiency L2 learners was indistinguishable from the native speakers, indicating that high-proficiency learners *can* indeed rely on native-like neurocognitive mechanisms, even those processes which are rapid and automatic, contra to predictions of the CPH.

This debate, however, is not yet resolved, as this conclusion has not gone unchallenged by other researchers who continue to emphasize the role of age-of-acquisition, showing persistent differences in the neurocognitive responses of even high-proficiency late learners. An fMRI study by Wartenburger and colleagues (2003) tested semantic and morphosyntactic processing (agreement violations in number, gender and case) in 3 groups of Italian-German bilinguals: (1) early high proficiency (AoA = birth); (2) late high-proficiency (AoA = 18); and (3) late low-proficiency (AoA = 20). During syntactic processing, early high-proficiency speakers did not show any differences in their L2 brain activation patterns compared to L1 activation patterns, whereas late high-proficiency speakers required the recruitment of additional neural resources to complete the task, thus revealing an effect of AoA. In contrast, for lexical semantic processing, differences in the pattern of brain activity between L1 and L2 were dependent on proficiency level, rather than AoA. Based on these results, the authors argued that the role of AoA for syntactic processing cannot be ruled out, and effects cannot be accounted for by L2 proficiency level alone; rather, both factors differentially impact the neurocognitive processing mechanisms underlying lexical-semantic and morphosyntactic processing in the bilingual brain.

Using ERPs, Hahne (2001) examined the processing of German phrase structure violations (e.g. *The shop was being *<u>on</u> closed*) in late-learners of German (AoA > 10). As expected, low-proficiency L2 learners showed neither a negativity (ELAN or LAN) nor a P600 effect. However, in contrast to the studies discussed above, high-proficiency L2 learners *also* failed to show the early, automatic negativity, and showed a slightly delayed P600. It was suggested that late, controlled syntactic processes (as indexed by the P600) can be native-like in L2 learners who have reached a high level of L2 proficiency, but that earlier and highly automatic processes (indexed by the LAN) cannot be achieved by late learners, regardless of proficiency level, due to maturational constraints on brain plasticity.

Another study by Ojima and colleagues (2005) assessed lexical-semantic violations (e.g. *This house has ten *cities in total*) and subject-verb agreement violations (e.g., *Turtles *moves slowly*) in late Japanese L2 learners of English. Native-English speakers as well as both high- and low-proficiency L2 learners of English elicited a significant N400 effect in response to semantic violations, though there were latency differences (i.e., delayed N400) from native-speakers, particularly for the low-proficiency learners. In response to morphosyntactic violations, low-proficiency L2 learners did not elicit *any* significant ERP effects, and although learners with high L2 proficiency showed a LAN between 350 and 550 ms, similar to native-speakers, they surprisingly failed to elicit a P600. This was the only study to report the presence of a LAN in the absence of a P600 in late L2 learners (and not vice-versa), and upon further inspection of the data, it can be seen that what the authors refer to as a "LAN" in the violation condition is in fact a relative *positivity* in the correct control condition. Moreover, as will be discussed in the concluding paragraph of this section, the absence of the P600 was due to the choice of time-window under investigation. Although these findings must, therefore, be interpreted with caution, the authors concluded that morphosyntactic processing must be constrained by AoA-effects, and, although proficiency does play a role in modulating ERP responses among L2 learners, the divergence from native-speakers cannot be explained on the basis of proficiency alone (see also Chen et al, 2007; Pakulak & Neville, 2011).

To account for such L1-L2 differences in morphosyntactic processing, Clahsen and Felser (2006a; 2006b) have put forth their "*Shallow Structure Hypothesis (SSH)*" in which they argue that the syntactic representations computed by late L2 learners during comprehension are more "shallow" (i.e., less accurately detailed) than those of native-speakers, leading late-learners to rely instead on more superficial cues, such as lexical-semantic information, verb biases and plausibility during syntactic processing. Importantly, Clahsen and Felser posit that late L2 learners may only resemble native-speakers in select aspects of syntactic processing, namely those governed by *local dependencies* (e.g., gender/number agreement within a clause). In contrast, due to critical-period effects, fundamental differences remain (even at high levels of proficiency) in how late L2 learners process structures involving *non-local dependencies* (such as in relative clauses where a certain syntactic constituent does not appear in its canonical position, e.g., *The elephant that the boy rode* ___ *was no longer at the zoo*), as a result of their shallow representations of these complex syntactic properties.

In sum, despite much evidence that L2 proficiency is the only determining factor in predicting native-like neurocognitive mechanisms in L2 processing, this view remains challenged by researchers who continue to stress the impact of a late AoA, and who present evidence of persisting differences in the kinds of neurocognitive processes engaged in by late L2 learners compared to native speakers, regardless of proficiency level.

Additional insights into the specific effects of proficiency level on the nativelikeness of L2 processing can be gained from the few *longitudinal* ERP studies that have been conducted so far. Compared to the cross-sectional designs reviewed above, the advantage of such longitudinal designs is that they can reveal more about potential changes that may occur in ERP responses with increasing proficiency levels, and may allow us to map these changes onto predictable stages in L2 development. This is likely to be more informative than studies offering a snapshot at a single level of proficiency, where it may be unclear what stage of L2 development the speakers are in, thereby resolving the inconsistencies in ERP patterns that have been reported across studies (e.g. absence/presence of LAN, absence/presence of P600). Furthermore, evidence of systematic changes in ERP profiles would constitute the best evidence that brain mechanisms involved in language processing do remain plastic, even in adulthood, and that any differences between L1 and L2 speakers cannot be attributed to neurobiological maturational constraints.

In a study focusing only on the acquisition of new *lexical* knowledge, McLaughlin, Osterhout & Kim (2004) investigated whether increasing proficiency (i.e., increasing familiarity with new words as a result of classroom-instruction) would be reflected in changing ERP profiles. In a priming experiment, the authors showed that increasing vocabulary proficiency (i.e. learning the distinction between familiar words vs. legal pseudowords, and between semantically-related vs. unrelated words) was mirrored in N400 effects that became more pronounced and more native-like. Interestingly, despite this change in ERP patterns, L2 learners continued to perform at chance-level when performing an overt lexical decision task (i.e., judging whether a string is a real word or a non-word), indicating that ERPs reveal more than traditional behavioral measures do, and may tap into changes in neurocognitive processing patterns *before* they show up in linguistic behavior.

Another longitudinal ERP study by Osterhout and colleagues (2006) investigated the early stages of lexical-semantic and syntactic development in English late learners of French during a 9-month introductory-level university French course. Each L2 learner was tested after approximately 1, 4, and 8 months of instruction. Native French speakers showed an N400 in response to semantic anomalies and a large P600 effect for syntactic anomalies (person agreement violations). A subset of the L2 learners (categorized as "fast learners" based on their behavioral data) also showed a robust N400 in response to semantic violations, starting from the first testing session. In terms of syntactic processing, after 1 month of instruction, ungrammatical sentences elicited an N400 effect rather than a P600. After 4 months of instruction, the N400 was replaced by a small P600. The P600 amplitude was subsequently shown to increase with increasing L2 proficiency, indicating that real-time syntactic processing undergoes systematic changes, even in late learners, as proficiency levels increase. These stage-like changes (from N400 responses to native-like P600 responses) with increased exposure and proficiency were also reported in a cross-sectional study examining subject-verb agreement processing in beginning and intermediate English late-learners of German after 1, 2 or 3 years of classroom instruction (Tanner, Osterhout & Herschensohn, 2009).

Another longitudinal study examined morphosyntactic processing in Korean late-learners of English, both before and after a nine-week intensive English course (White, Genesee & Steinhauer, 2012). Participants read sentences that were either grammatical (e.g., *The teacher did not <u>start</u> the lesson*; and *The teacher had not <u>started</u> the lesson*) or contained violations of English regular past tense (e.g., *The teacher did not *<u>started</u> the lesson*; and *The teacher had not *<u>start</u> the lesson*). While no P600 effects emerged in response to violation sentences at the start of instruction, significant P600s were elicited at the end of the intensive course, in line with observations reported by Osterhout and colleagues. Contrary to the McLaughlin et al. study, however, White and colleagues found a significant correlation between behavioral performance on the online task and observed ERP patterns (especially those in response to correct, rather than incorrect, judgments), indicating that learners who perform well behaviorally are more likely to engage in native-like processing mechanisms than learners who perform poorly (also see Tanner, Osterhout & Herschensohn, 2009; 2013). This notion is directly related to the concept of proficiency and, more precisely, "*structure-specific proficiency*" which will be further discussed in *Section 1.5*. Note that the discrepancy from McLaughlin and colleagues' result is likely due to the variability in task (lexical decision vs. grammaticality judgment), type of learning (vocabulary vs. past-tense grammar rule) and participants (different L1-backgrounds).

White and colleagues' findings support the claim that L2 proficiency drives changes in the neurocognitive mechanisms used for L2 processing, and that learners can engage in native-like mechanisms once high levels of proficiency are reached, although they only began to acquire this language in adulthood.

Bowden and colleagues (2014) also tracked the neurocognitive changes associated with L2 learning, in their study with late L2 learners of Spanish in a university context. While semantic violations elicited N400 effects in both a low-intermediate proficiency L2 group and an advanced-L2 group, morphosyntactic word-order violations (modeled after the balanced designs in Steinhauer et al. 2006 and 2012) only elicited a LAN/P600 pattern in the *advanced* L2 group, not in low-proficiency learners. Importantly, the advanced L2 learners were statistically indistinguishable from the Spanish native-speaker controls, suggesting once again that L2 morphosyntactic processing may differ from L1-profiles at initial stages of acquisition, but can eventually converge onto native-like processes with sufficient exposure and proficiency.

In light of the consistency of such findings, Steinhauer, White and Drury (2009) discuss the temporal dynamics of L2 acquisition and how ERP patterns associated with language processing systematically change over the course of late L2 acquisition. Steinhauer and colleagues' main claim is that, with increasing proficiency, late L2 learners' brain activation patterns begin to converge upon those of native speakers, in stages that reflect systematic changes in their reliance on specific neurocognitive strategies during syntactic processing. At the earliest stage in L2 acquisition, ungrammatical sentences are not distinguished from grammatical ones and there are no observable ERP effects. At very low proficiency levels, morphosyntactic violations elicit an N400 response, as these violations are initially perceived as a lexical problem, and the L2 learner relies on declarative rule knowledge and compensatory processing strategies. Between low and intermediate proficiency levels, morphosyntactic violations begin to elicit a small P600, indicating that participants begin to proceduralize or "grammaticalize" morphosyntactic rules (Osterhout et al., 2006; Steinhauer, 2014; see also Ullman 2001; 2005). At intermediate proficiency, late L2 learners begin to show a larger and earlier P600, indicating a late stage of reanalysis and repair which approaches native-like mechanisms. At this stage, if negativities precede the P600, they continue to look like N400s rather than LANs. At higher "near-native" levels of proficiency, L2 learners elicit an anterior negativity which could be bilateral rather than left-anterior, followed by a native-like P600. Finally, at the highest (native-like) proficiency level, late L2 learners show ERP profiles that are indistinguishable from those observed in native-speakers (LAN and/or P600, depending on the type of violation). Steinhauer and colleagues suggest that the relative order of which structures reach higher levels of proficiency may be influenced by factors such as exposure/L2 training or typological similarities between the L1 and L2 (for more details on transfer effects, see Study 3, as well Kasparian, Bourguignon, Drury & Steinhauer, 2010; Sabourin, 2003; Thierry & Wu, 2007; Tokowicz & MacWhinney, 2005; White, Genesee & Steinhauer, 2012).

The developmental trajectory from non-native-like to native-like L2 processing that has been reported in several studies is taken as evidence for the "*Convergence Hypothesis*" (Steinhauer, 2014), according to which the neurocognitive mechanisms underlying L2 acquisition are dynamic and, with increasing practice and language proficiency, converge upon L1 patterns. Thus, there is no evidence that neurobiological maturational constraints prevent L2 learners from ultimately calling upon the same cognitive mechanisms as those involved in L1 processing. Instead, L2 learners are expected to differ from native-controls *only* so long as they are not sufficiently proficient in the L2.

This section reviewed the debate on the effects of AoA and proficiency level on brain mechanisms underlying L2 processing. The picture is obviously mixed, and the question of whether AoA has any impact at all once proficiency level has been controlled is still highly controversial. Discrepancies in findings are due to a number of methodological differences across studies, which clearly influence the ERP results and, naturally, the conclusions drawn by the researchers. For example, studies widely differ in the structures investigated, the L1-background of the L2-learners (and whether certain structures may be easier or more difficult to acquire due to L1-L2 "*transfer effects*"), the design of the stimuli and the nature of the violations, the behavioral tasks employed in conjunction with ERP recording and, as will be discussed in detail in the following section – in how exactly "proficiency" is defined and measured. Another factor potentially accounting for mixed results are the time windows in which ERP effects have been quantified; for example, in the study by Ojima and colleagues (2005) where the P600 was found to be absent in all L2 learners regardless of proficiency level, the researchers had imposed a cut-off of the P600 window at 850 ms, although it has been established in the literature that the P600 often shows variability in its latency and can be observed as much as 900 - 1000 ms after the onset of the target word (Lück, Hahne & Clahsen, 2006; Hahne, 2001; Rossi et al., 2006, as discussed in Steinhauer et al., 2012). Finally, it is also likely that, as a result of participant-related factors that have not been well-controlled, there is a large amount of individual variability in L2 learners' processing strategies as reflected by differing ERP profiles, such that it is impossible to obtain a consistent and statistically-significant group pattern.

1.5. Measuring "proficiency": Inconsistencies in methodology

As mentioned above, one likely reason for the discrepant findings in the literature could be the way in which "proficiency" is defined and measured across studies. First, although many neuroimaging and ERP studies make a point of examining effects of L2-proficiency, they often employ self-report measures, which are not extremely reliable. Other times, researchers use only one measure of overall proficiency, such as a cloze-test (Goad & White, 2008) where participants have to "fill in the blanks", often by choosing a word from a multiple-choice list of options. Other studies have used vocabulary translation tasks, standardized language measures, or grammaticality judgment tasks. Given this wide range of proficiency measures, it is difficult to ensure that an L2 learner who is categorized as having a high level of L2 proficiency in one study would be similarly categorized in another study, using another proficiency measure. Thus, given the methodological inconsistencies across various neuroimaging and neurophysiological studies, it is a challenge to accurately compare the impact of proficiency level on ERP response patterns across studies. Furthermore, apart from the Abrahamsson and Hyltenstam (2009) study described earlier, it is rare that researchers employ a range of proficiency measures and a battery of tasks in order to try and obtain a well-rounded picture of the linguistic abilities of the L2 learners.

Secondly, even within the same group of L2 learners, proficiency levels may vary between different structures (due to facilitation/difficulty resulting from L1-L2 transfer effects, or to relative experience with certain structures as a result of exposure or instruction), and may result in distinct ERP patterns. For example, in Hahne and colleagues' (2006) study with late Russian learners of German, the behavioral data indicated that these L2 learners were more proficient with past participles than with noun plurals. This differential proficiency level was also apparent in the ERP data; whereas the L2 learners showed P600 effects for both inflectional violations involving past participles and noun plurals, they showed an early, automatic LAN effect only in response to past participle violations. This finding indicates that *global proficiency* measures may not predict the ERP response patterns for a particular structure. As briefly discussed in the light of the findings by White, Genesee and Steinhauer (2012), we have shown in studies conducted in our laboratory that *structure-specific proficiency* (i.e., scores on behavioral tasks assessing participants' mastery of the specific structure(s) under investigation, for example a grammaticality judgment task), may be more highly correlated with ERP profiles than global proficiency measures (see Kasparian, Bourguignon, Steinhauer & Drury, 2010; White, Genesee and Steinhauer, 2012; Steinhauer, 2014). In their studies with English learners of German, Tanner and colleagues (2009; 2013) also found the amplitude of the P600 in response to subject-verb agreement violations was positively correlated with learners' performance on an online grammaticality judgment task. Thus, it may be beneficial to (1) to employ a range of proficiency measures, assessing both "global" and "structure-specific" proficiency; and (2) to correlate ERP profiles with proficiency scores, including structure-specific behavioral performance, to explore which proficiency measure(s) may be more or less sensitive in reflecting changes or differences in neurocognitive responses to language.

A third issue concerns the categorization of L2-learners into subgroups of either "high" or "low" proficiency. There is no real consistency in how researchers typically create such proficiency subgroups, and which criteria they base their cut-offs upon (e.g. actual test scores, self-reports, or proficiency indirectly defined by classroom level and months/years of instruction). This variability makes it difficult to compare subgroups of "high proficiency" learners across different studies. An alternative or, in the case of this dissertation, a complementary approach would be to treat proficiency as a *continuum* and to treat proficiency as a continuous variable. To date, only a handful of studies have opted for this method, most of them having focused exclusively on lexical-semantic processing and modulations of the N400 (Moreno & Kutas, 2005; Newman, Tremblay, Nichols, Neville & Ullman, 2012; Ojima et al., 2011). For example, Newman and colleagues (2012) used linear mixed effects (LME) modeling to assess the relative contributions of AoA and proficiency level on N400 responses to lexical-semantic violations (e.g., *The Irishman sipped Todd's *<u>thunder</u> at the party*) in Spanish late-learners of English. The researchers measured proficiency with a standardized test of English proficiency (TOAL-3, Hammill et al., 1994) and created a composite score by summing the scores of the five separate subtests (2 vocabulary, 3 grammar) for each participant. It can be argued that most of the L2 learners were in low-proficiency range (averaging in the 31st percentile on all subtests). Despite the low proficiency scores and minimal overlap between the groups, the key finding was that proficiency as a continuous variable accounted for modulations in N400 amplitude; once the effects of proficiency were accounted for, group status (i.e., L1 vs. L2, or "AoA") no longer predicted differences in N400 amplitude. However, given that researchers in the field have typically advocated for fundamental L1-L2 differences predominantly in the *morphosyntactic* domain and especially for structures of high complexity, it would be useful to adopt this approach in studies of syntactic processing, and to compare how the impact of proficiency level on ERP profiles may change when using one approach (proficiency subgroups) versus another (proficiency as a continuous measure).

Lastly, it has been shown that such proficiency effects are not merely limited to L2 acquisition, although most studies have adopted the erroneous assumption that L1-speakers always perform at ceiling (i.e., have maximal "native" proficiency). It has recently been shown that, even within native monolingual speakers, higher *L1* proficiency levels result in ERP patterns that are typical of native-like profiles, such as more strongly left-lateralized LAN effects and larger amplitudes of P600s, compared to native-speakers with lower L1 proficiency with less-lateralized anterior-negativities and smaller P600s in response to phrase-structure violations (Pakulak et al., 2004; Pakulak & Neville, 2010). Thus, any viable neurocognitive investigation of L2 processing in comparison to L1 processing must consider these issues surrounding the definition and measurement of proficiency, and how proficiency effects may pertain to native-speaker controls as well.

The collection of studies in this dissertation attempt to address the shortcomings highlighted in this section, in order to clarify the role of proficiency on the neurocorrelates of language processing. As outlined in the *General Methods* section, we tested attriters and late L2-learners with a variety of behavioral measures of global proficiency, in speakers' L1 *and* L2: a self-report questionnaire, a written C-test, a written error-detection task, a verbal translation task, and a verbal semantic fluency task. In addition, we investigated structure-specific proficiency by obtaining participants' ratings in response to sentence stimuli in an acceptability/grammaticality judgment task during ERP recording. We chose to use a rating scale (1-5) rather than a binary response (yes/no), as a scale may exhibit a more fine-grained sensitivity to proficiency differences. Crucially, these same measures (except the verbal translation task which was a bilingual exercise) were also administered to both monolingual control groups of native-speakers, thus allowing us to examine potential proficiency differences within native-speakers as well. Furthermore, the present ERP experiments covered a large range of lexical-semantic items and syntactic structures, some of which were of low frequency, high complexity and involved long-distance dependencies, in order to challenge our most highly-proficient speakers and to maximize the potential for differences between groups. Lastly, another important goal of this research was to correlate observed ERP profiles with proficiency level as a continuous variable (either as individual measures or a derived composite measure), in order to (1) compare the subgroup method with a continuous-variable approach, and (2) investigate whether one/some of our proficiency measures was/were most sensitive predictor(s) of native-like ERP profiles.

1.6. The special case of first-language attrition

As explored in detail in the previous section, despite the attention that AoA and proficiency have received in the literature on L2 acquisition, the relative impact of these factors on the nativelikeness of brain mechanisms underpinning L2 processing has not yet been disentangled in a satisfactory, conclusive way.

The optimal scenario in which AoA effects could be effectively disentangled from proficiency differences would be one where AoA and proficiency level would operate in different directions: the bilingual speaker would have reached a high-proficiency level in the late acquired L2, but would have non-native / lower proficiency level in the L1 acquired from birth. Such a scenario would allow us to determine what has a greater impact on the brain's organization and recruitment of resources during language processing (both in the L1 and L2): whether or not a language was acquired within a critical period, or whether the speaker has a high level of command of (and/or exposure to) this language. The proposed study was designed with this optimal scenario in mind. The target group of L2 speakers were "L1-attriters" who, due to immigration and limited exposure to their native language, exhibit increasing linguistic difficulties in their L1, while increasingly developing their proficiency level in the late-learned L2.

1.7. Minority-language speakers and first-language attrition

1.7.1. The unique linguistic situation of "attriters"

In a bilingual or multilingual community, *minority-language speakers* are individuals whose first language and culture is not the *majority language* or culture. Existing definitions of minority-language speakers vary considerably. Generally, the term is used to refer to individuals who immigrated to the host country either in childhood or adulthood, or to their children (second and third generations of immigrants, also called "heritage speakers") who grow up in the host country.

In this dissertation, the population of interest consists of *first-generation immigrants* who had fully acquired their native-language (Italian) in their home country (Italy) prior to moving to the new country (Canada) in adulthood, and who experienced a sudden shift in use and dominance towards the majority language of their new setting (English), as they have very limited opportunities to use their native language in their new environment. This linguistic situation has been described as a form of “*subtractive bilingualism*” where learning and predominantly using the majority language comes at the cost of the minority language (Lambert, 1974). These speakers gradually come to have a greater grammatical proficiency, richer vocabulary and/or greater fluency in the L2, compared to their L1, and may begin to deviate from other native-speakers of their L1 (Genesee, Paradis & Crago, 2003; Polinsky, 2000). This shift in dominance and linguistic ability has been termed "*first-language (L1) attrition*".

First-language attrition has been defined as the deterioration or loss of an individual’s native language that had been acquired and was previously used at a normal, native-like proficiency level. This deterioration is not due to normal aging or to pathological causes such as illness or injury, but is the negative consequence of contact with a second language, in a context where contact with the L1 community is limited or severed (see Köpke, 2002; Köpke & Schmid, 2004; Seliger & Vago, 1991 for a review and definitions). It is important to distinguish between "language attrition" and "language shift/death"; whereas language shift and death typically take place in bilingual communities over several generations, the term “*attrition*” specifically refers to language loss at the individual level and takes place within one generation (De Bot, 2001; Schmid, 2011). Some researchers also make a distinction between "L1 attrition" and "*incomplete L1 acquisition*" (Gürel, 2002; Köpke, 2002; Polinsky, 2000; Schmid, 2002), whereby the term "attrition" is used to refer to only those situations where the L1 had been acquired at a native-speaker level and had remained before immersion into the L2 *after* childhood, contrary to the case of childhood learners (or "heritage learners") of a minority-language, where the shift towards the majority language occurs before the minority-L1 is “fully acquired”. In other words, “attrition” is reserved for situations of sequential acquisition of the L2 after childhood, and not for cases of simultaneous or sequential acquisition of the L2 during childhood.

Behavioral symptoms of L1 attrition have been described as non-nativelike pronunciation (De Leeuw, Schmid & Mennen, 2010; Flege, 1987; Major, 1992), increased lexical retrieval difficulties (De Bot, 1996; Hulsen, 2000; Köpke, 1999; Montrul, 2008; Opitz, 2011; Pavlenko, 2000), increased occurrences of borrowing from the L2 (Pavlenko, 2000), blurred semantic constraints on L1 words and the intrusion of “false friends” from the L2 (Pavlenko, 2000), decreased fluency reflected by frequent hesitations and pauses in speech (Nakuma 1997; Schmid & Fägersten, 2010; Schmid & Köpke, 2009), and difficulties with idiomatic language (Jarvis, 2003) or with pragmatic formulations of requests (Cenoz, 2003). Attrition in the syntactic domain has been described in terms of previously mastered grammatical knowledge becoming “wobbly” (Ammerlaan, 1996) and grammatical performance deviating from native-speakers for a wide range of grammatical structures (Ammerlaan, 1996, Jarvis, 2003; Schaufeli, 1996; Seliger, 1989; Yağmur, 1997; Polinsky, 1997; Schmitt, 2010; Schmid, 2010; Schmid & Köpke, 2011).

1.7.2. Theoretical accounts of attrition

A number of different hypotheses have been proposed to explain the causes of L1 attrition and the different neuropsychological processes involved. Two main views have been put forward: (1) attrition as *forgetting* and *decay*, as a result of disuse, and (2) attrition as *crosslinguistic influence* or *transfer* from the L2 to the L1, as a result of competition and interference between the individual's two languages.

According to generative approaches, attrition entails a decay of the speaker’s underlying linguistic representations (Ecke, 2004). Due to less frequent L1-input, native competence may be altered at a deep syntactic-knowledge level and there may be irrevocable structural changes to the grammar of the native language (Gürel, 2002; 2004; Seliger & Vago, 1991). Seliger and Vago (1991), argue that attrition is either (1) the result of L2-influence at the underlying L1-

competence level, resulting in *erosion of the L1-system*, or (2) the result of the interaction between two *intact* linguistic systems which are simultaneously activated. They also argue that it is only the case where erosion reaches the level of underlying competence "that allows for interesting claims and meaningful insights into the attrition process.”

Other researchers working from a psycholinguistic point of view attribute attrition to a *retrieval* problem rather than a problem of *storage* (i.e., of underlying representations of structures) and claim that observed difficulties in accessibility and retrievability can be explained in terms of an increase in activation thresholds for L1 lexical items and L1 morphosyntactic structures stored in memory, as a consequence of infrequent use (Green, 1986; Paradis, 1997). According to the *Activation Threshold Hypothesis* (Paradis, 1989; 1997), items that are more frequently activated require less stimulation to be reactivated, compared to items that are less frequently activated. Language disuse, therefore, leads to a gradual loss of accessibility or retrievability, resulting in a shift of language dominance and, eventually, in attrition. In line with this theory is the finding that one of the earliest indications of language attrition is not the loss of certain items but an increase in the length of time required for their retrieval (Hansen, 2001). Similarly, it has been reported that relearning an attrited language can be more rapid than acquiring it for the first time, suggesting that the language may not be permanently "lost" (i.e. erased from memory) but inactive (Berman, 1979; Slobin et al., 1993). The *Activation Threshold Hypothesis* also offers an explanation for the selective nature of L1 attrition, as it has been shown that not all aspects of language are susceptible to attrition to the same degree. Rather, the threshold of activation becomes raised for those items which are not frequently activated, causing selective difficulties in the L1. Similarly, according to Paradis (2003; 2007), elements subserved by declarative memory, such as vocabulary, are more vulnerable to attrition than those which rely on procedural memory, such as morphosyntax. This hypothesis has been supported by studies demonstrating that the L1-lexicon is more strongly affected by lack of use than L1-grammar (De Bot, 1996; Hulsen, 2000; Köpke, 2002; Köpke & Schmid, 2004; Opitz, 2011; Schmid & Köpke, 2008).

A related claim argues that L1 attrition is the logical consequence of becoming dominant in the L2. In other words, attrition is an *L2-induced change in the L1*. Language knowledge (in the generative perspective) and/or processing deviates from native-like characteristics as a direct result of *influence* and *interference* from the L2, rather than as a result of L1 memory traces fading away and being "forgotten" or inactive due to disuse. This view would predict, then, that L1 attrition manifests itself as the incorporation of L2 elements into the L1 vocabulary and grammar, and is reflected in a speaker's acceptance of sentences that are semantically or syntactically incorrect in the L1, under the influence of L2 grammar rules (Cook, 2003; Pavlenko, 2000). A large body of empirical evidence has revealed the role of L2 influence in L1 attrition in a number of linguistic domains, specifically where there is competition between the L1 and L2 in vocabulary, syntactic properties or in phonology (see Köpke, 2004 for a review). Linguistic forms that are similar/corresponding across the L1 and L2 are more subject to interference than forms that do not correspond between the two languages (Andersen, 1982; Gürel, 2002; 2004).

It seems likely that L1 attrition is affected by *both* lack of use (and, thus, a decline in accessibility) as well as an increase in competition and cross-linguistic influence of the L2 on the L1 (Cook, 2003). A recent definition given by De Bot incorporates the idea that both of these processes may be interrelated: "L1 attrition is both a decline of retrievability of declarative linguistic knowledge and deproceduralization of linguistic knowledge in L1, and an increase of competition with L2 knowledge". Thus, various neuropsychological processes (such as decay, interference, etc) might result in changes in L1 grammar and an overall decline in L1 proficiency. Still, the relative role played by disuse and crosslinguistic transfer on the attrition process and on the specific structures which are affected remains largely open for investigation.

This dissertation does not seek to determine the *level* at which attrition occurs, i.e. whether the underlying linguistic representation/competence is altered or whether the problem lies with accessibility/retrieval of this knowledge during performance. On the other hand, the different experimental designs used in these studies do allow us to investigate whether attrition may be seen as a global difficulty in the L1 due to infrequent use, versus L2-induced changes in L1 processing.

1.7.3. Disentangling attrition effects from bilingualism effects

An additional note concerns the identification of attrition, despite the general characteristics and symptoms outlined in this chapter. Schmid (2011) has discussed the difficulty involved in distinguishing an attriter from a non-attriter and cautions researchers against interpreting all differences from monolinguals as evidence for attrition. According to Schmid, transfer errors and interference can be indicative of attrition, but also simply a result of "bilingual language use". In order to disentangle attrition effects from bilingualism per se, a second critical group of late L2-learners is needed as a bilingual control group. As argued by Schmid, without this control group of L2-learners, it would not be possible to draw reliable conclusions about effects that are due specifically to attrition, above and beyond effects that are typically observed in other bilinguals.

While I agree with the importance of comparing bilinguals not only to monolinguals but also to other bilingual speakers, I believe that this argument is incomplete, and the simple inclusion of a L2-learner group does not immediately resolve this issue. First, given that attrition is the circumstantial *result* of bilingual language use (i.e., a unique situation of subtractive bilingualism where reduced L1 input/use is the direct consequence of predominantly using the majority-L2), it is impossible to fully disentangle "attrition effects" from "bilingualism effects", even with the inclusion of an L2-learner group; in other words, on a theoretical level, the two concepts are inherently confounded – a bilingualism-induced change to the L1 is the very definition of attrition (whether or not there are actual interference effects observed between the languages).

Secondly, L1 differences between attriters and native-speakers cannot be explained *only* on the basis of how attriters pattern with the L2-learners. For example, if attriters differ in their L1 from native-speakers, but are similar on this language to L2-learners (for whom the language in question is the *L2*), we cannot simply attribute this pattern to a general "bilingualism effect" rather than to attrition, since language status (L1 vs. L2) is confounded in the comparison between L1 attriters and L2 learners. Only by *additionally* assessing the L2-learners on their *L1* (compared to native-speakers of that language), can we attempt to tease apart attrition effects from bilingualism. If L2-learners are indistinguishable from native-speakers on their L1 (although they are bilingual, like the attriters), this would provide extra support that the differences seen between attriters and native-controls on *their* L1 are not attributable to general bilingualism effects. In contrast, if L2-learners differ on their L1 in similar ways from monolingual English controls as attriters differ from monolingual Italian controls, then it could be argued that such effects are possibly due to crosslinguistic transfer common to both bilingual groups. Thus, examining the *L1* of L2 speakers (and not just their L2 as is commonly done in the vast majority of L2 acquisition and attrition studies, including those currently being conducted by Schmid and colleagues (Bergmann et al., 2013; Schmid 2013; also see http://www.rug.nl/staff/m.s.schmid/Project_description.pdf) is vital to make sense of this argument.

Thirdly, if crosslinguistic transfer effects (competition/interference) are observed both in L1-attriters and L2-learners, one cannot ascribe these to "bilingualism effects" without examining whether there are differences in the *direction* of crosslinguistic transfer. Once again, assessing both languages of each group of bilingual speakers becomes necessary. It may be that, while L1-to-L2 transfer is observed in both bilingual groups, transfer in the reverse direction from the L2-to-L1 only occurs in situations of predominant-L2 use and is therefore part of attrition.

In sum, while I agree with Schmid's statement that including a group of late L2-learners is necessary in studies of L1 attrition, I further argue that the inclusion of an L2-group only provides clear answers if *both* languages of the bilingual groups are investigated, relative to their respective monolingual control groups, yielding a design with four main groups. In addition, two bilingual control groups would be beneficial: (1) “attriters” who still show high levels of L1-proficiency (i.e., who are at the top of the attrition group) although they share the same experiential "prerequisites" for experiencing attrition, such as dominant L2-use in the L2-environment with low/no exposure or use of the L1; and (2) non-attriting bilinguals who still live in the L1-environment and continue to be dominant in the L1 though they are highly-proficient in the L2. The first of these two groups would allow us to examine whether L1-attrition manifests itself as "low L1-proficiency effects" rather than an overall bilingualism effect (with high-proficiency "attriters" different from low-proficiency attriters but *not* from native-monolingual controls despite their being bilingual), or whether observed differences from native-speakers may characterize the group as a whole. The inclusion of the latter group would further disentangle whether such effects that characterize the group of attriters as a whole may be due to the L2-dominant environment and to specific circumstances that compose "attrition", in which case they should *not* characterize L1-dominant bilinguals who still live in the L1-environment.

Thus, there is much work to do beyond including late L2-learners as a group and merely examining their L2 relative to the L1 of attriters and native-monolinguals. The series of studies conducted as part of this dissertation adopts this approach, not only testing the L1 and L2 of both attriters and L2 learners relative to two groups of monolingual native-speakers, but also comparing subgroups of higher- and lower-proficiency individuals within each group of bilinguals. The advantage of this approach is that attrition and acquisition can be considered as part of a continuum.

1.7.4. Experiential factors affecting attrition

Several factors have been shown to contribute to the likelihood and the degree of first language loss and have received considerable attention in behavioral studies of attrition, particularly from the socio-linguistic perspective. As described by Schmid (2011), the factors which are most often explored as mediating L1 attrition can be grouped into three main categories: (1) *background factors*, such as age, age of emigration, education, and length of residence, (2) *factors related to language use*, such as amount of L1 and L2 exposure, and (3) *internal and psychological factors*, such as attitudes and emotion (see Köpke & Schmid, 2004; Montrul, 2008).

Age of emigration is considered to be the most crucial factor at play in one's vulnerability to attrition and has also been defined in terms of AoA or *length of residence (LoR)* in the host country (Ammerlaan, 1996; De Bot & Clyne, 1994; Schmid, 2002; 2007). The main assumption regarding age is that attrition should be more severe in young immigrants than in adults. This assumption is directly related to the concept of neuroplasticity and the CPH, predicting that younger immigrants will adapt more rapidly to a radical change in linguistic environment due to greater brain plasticity, leading them to more readily acquire the L2 and, at the same time, to be more susceptible to losing their L1. In older immigrants, on the other hand, reduced brain plasticity would hinder successful adaptation to the L2 environment and should also make them less vulnerable to changes in their (stable) L1 (Köpke, 2007). This hypothesis has been supported by a number of studies comparing pre- and post-puberty immigration (Ammerlaan, 1996; Bylund, 2008; Pelc, 2001; also see Köpke, 2004; Köpke & Schmid, 2004 for a review). Interestingly, behavioral studies on L1 attrition in post-puberty immigrants have failed to find clear effects of age of emigration, and have revealed a rather small range of variability in these speakers' linguistic abilities, with a low incidence of non-native-like L1 use (Köpke, 1999, age = 14-36; Schmid, 2007, age = 17-51). Thus, similar to claims of a "critical period for L2 acquisition", there may be a "critical period for L1 loss" (Montrul, 2008), beyond which actual restructuring/forgetting will be unlikely to occur. This dissertation re-examines these findings in new light by using ERPs in addition to behavioral measures. Is it the case that attrition and non-native-like L1-use *does* turn up in adult immigrants, contrary to much of this research, when we look at various aspects of language-processing with ERPs?

Amount of L1 exposure (or *frequency of L1 use*) is the second factor which is often explored for its predictive value on the occurrence of attrition (De Bot et al., 1991; Köpke, 1999; Schmid, 2007). The assumption is an intuitive one; attrition should be more severe in minority-language speakers who have less input/exposure in their L1, as a result of limited contact with the L1 community. From the generative perspective (Gürel, 2004; 2007; Gürel & Yilmaz, 2011; Tsimpli, 2007), minority-language speakers with limited L1 input will have fewer instances of evidence to support their L1 linguistic representations and may instead use the more readily-available L2 input as evidence, resulting in changes in the underlying grammar. From a processing perspective, infrequent use of the L1 will render the language dormant, and will result in accessing and processing difficulties because of increased cognitive resources needed in order to inhibit the more strongly active and automatic L2 (Green, 1986; Norman & Shallice, 1986). Thus, attriters' L1 performance should be positively correlated with their level of L1 use, and negatively correlated with their level of L2 use.

Research by De Bot and colleagues (1991) and Köpke (1999) has shown that attrition is indeed greater in individuals who have less contact with the L1. However, others (such as Jaspaert & Kroon, 1989) have found no significant correlation between attrition and the amount of L1 input, and there are also a number of studies who have reported evidence of attrition (specifically in terms of L2 interference) in adults who claim to have continued to use their L1, even on a daily basis (Ben Rafael, 2001; Grosjean & Py, 1991; Jarvis, 2003; Major, 1992). As discussed in Köpke and Schmid (2003), these inconclusive results may partly be due to the methodological difficulty of measuring concepts such as frequency of use, or amount of exposure/input/contact, which are typically assessed via self-reports. Moreover, the authors stress how the popular approach of dichotomizing frequency of L1-use into only two subgroups ("high" and "low") does not do justice to the complex picture of which language is used for what purpose, and exactly how often. Thus, despite the relevance of frequency of L1-use in the context of activation models, the impact this factor on the likelihood or severity of attrition is still to be determined.

Furthermore, research has primarily focused on the *quantitative* aspect of L1 contact, rather than considering *qualitative* aspects such as the *context* of L1 exposure/use (e.g. professional setting, communications with family/friends, other expats, churches, community centers, newspapers, etc) or its *quality* (Schmid, 2007). According to Schmid (2011), a more fine-grained approach should be taken when quantifying language exposure (input) and use (output), and three different types of L1 use should be distinguished from each other: (1) *interactive language use* (spoken and written communication with others); (2) *non-interactive exposure* (reading and media); (3) *inner language use* (thoughts, dreams, diary-writing, counting and arithmetic). In fact, L1-use for *professional* purposes is one context factor that has consistently shown a significant negative correlation with L1-attrition, in a number of studies (Leeuw, Schmid & Mennen, 2010; Schmid, 2007; Schmid & Dusseldorp, 2010, as cited in Schmid & Jarvis, 2014). It is conceivable that speakers who use their L1 in a formal context such as work have more practice with inhibiting the L2 and resisting any intrusions onto the L1. Thus, taking qualitative aspects of L1-use into account is important when attempting to predict and explain L1 attrition.

A third dimension to be considered in linguistic analyses of attriters is the affective and motivational dimension (Ben-Rafael & Schmid, 2007; Köpke, 2000; Schmid, 2002). As highlighted by Schmid (2007), leaving one's home country could have far-reaching consequences, either positive or negative. Immigrants often report feeling like a foreigner or an outsider in the host country. In addition to this, there is also the degree to which a minority-language speaker feels motivated to maintain the L1 identity or, instead, to adopt the majority (L2) language and culture. The reasons for immigration (e.g. increased job opportunities, moving for a partner), and developing relationships with native-speakers in the new country are likely to affect these individuals' desire to learn the target language and belong to its culture (Gardner, 1985; Moyer, 2007). Motivation, identity, and attitudes towards the minority and majority languages have been recognized to play a role in language loss (Pavlenko, 2003; Schmid, 2002), although it is still not understood how these factors, individually or in an interaction with each other, affect the likelihood, severity and outcome of attrition.

Studies of L1 attrition find a large degree of variability between individuals in the degree of L1 difficulties. Examining the role of the factors outlined in this section is likely to shed light on these individual differences. This holistic view is the approach I have taken in this dissertation, with the background questionnaires I have adapted (from Schmid, 2011), and the experiential factors considered in correlational analyses.

1.8. Towards a neurocognitive investigation of L1 attrition

To date, research on L1 attrition has barely turned to neuroimaging methods to investigate the neurocognitive mechanisms underlying language loss. Methodologies such as fMRI or ERPs, which have widely been used in investigations of L2 processing, have largely been neglected in L1 attrition research (Schmid, 2009). A neurocognitive investigation of L1-attrition would shed light on many questions that are simply unanswerable by behavioral studies. When the L1 is subject to attrition, is it still processed in the brain like a native-language, engaging similar processing mechanisms as in native-speakers? Could attrition effects be observed in the brain on structures/tasks where we fail to find attrition effects behaviorally? Does attrition manifest itself as low-L1 proficiency effects, L2-to-L1 transfer and/or something more? What is the impact of experiential factors such as AoA, proficiency level, relative exposure to the L1 and L2 and length of residence on the underlying neural mechanisms involved in processing an attriting L1, and is the effect of these factors comparable to the patterns we observe in studies of L2 acquisition?

An fMRI study was conducted by Pallier and colleagues on an extreme case of L1 attrition – the case of internationally-adopted Korean children in France (Pallier et al., 2003; Vantureyra & Pallier, 2004). Participants were adopted between 3 and 10 years of age and they all claimed to have completely forgotten Korean, as French had become their predominant and most fluent language. Adoptees' brain activity was measured using fMRI while they performed a speech recognition task with Korean sentences compared to Japanese and Polish sentences, two languages to which they had never been exposed. Results from both the behavioral tasks and the fMRI task failed to reveal *any* detectable difference when processing Korean, compared to the two other languages which were completely foreign to the adoptees. In all respects, the Korean adoptees behaved in the same way as French controls who had no knowledge of any of the three languages. The authors argued against the CPH or any neurobiological theory claiming that neural connections become fixed during the early years of life, as a result of L1 learning (entrenchment) or due to maturational factors. Theories such as the CPH would have predicted that the adoptees – at least those who arrived in France at older ages – should have displayed some sensitivity to Korean. These results, however, seem to be in line with the behavioral findings of attrition in pre-puberty immigrants described in the previous section.

Different findings were reported in a recent fMRI study by Pierce and colleagues (2014) with internationally-adopted children from China who were exposed exclusively to French after early childhood adoption (on average at 12.8 months) and who had no conscious recollection of Chinese. Neural activation patterns during a lexical tone discrimination task were found to differ significantly between the adoptees and monolingual French speakers with no knowledge of Chinese, and involved the recruitment of additional brain regions such as the right superior temporal gyrus, which are typically implicated in the acoustic processing of complex, non-linguistic auditory stimuli. In contrast, adoptees' activation patterns were similar to those of Chinese-French bilinguals who had maintained their exposure to Chinese since birth, as both groups recruited the left superior temporal gyrus and planum temporale – areas typically associated with top-down processing of learned linguistic categories. Interestingly, activation in the left planum temporale correlated significantly with AoA –children who were adopted at older ages exhibited greater activation in response to lexical tones than children adopted earlier, suggesting that native-language representations are acquired by 6 months of age but increasingly developed with increasing exposure. However, no relationship was found between neural activation patterns and length of residence, suggesting that – once acquired – early-formed representations are stable, regardless of the amount of time that had elapsed without exposure to Chinese. These results, contrary to those reported by Pallier and colleagues', were taken as support for the CPH, as it was shown that new linguistic knowledge (L2) did not replace the neural circuits established by pre-existing knowledge (L1), even if the L1 has become difficult/impossible to access due to discontinued exposure. The studies' discrepant findings could perhaps be reconciled if you consider the difference in task (and, thus, the type of L1-knowledge being assessed); Pierce and colleagues used a phonological discrimination task, tapping into information that had been acquired prior to adoption, while Pallier and colleagues employed a speech recognition task with sentence stimuli containing grammatical elements that would likely not have been acquired during the first year.

As interesting as studies with international adoptees are, these are certainly extreme cases of attrition, where exposure and acquisition of the L1 is abruptly terminated following adoption, as the adoptive families do not speak the language of the children's country of origin.

Furthermore, these are situations of "incomplete L1 acquisition" (Montrul, 2008), as the termination of L1-exposure occurs in early *childhood*, when pro-CPH researchers would argue that neural networks are still plastic. The next logical step is to investigate the neurocorrelates of L1-attrition in speakers who fully acquired the L1 into *adulthood*, but who subsequently experienced a shift in exposure and dominance towards a majority-L2. This is the population in which the strongest evaluation of continued neuroplasticity can be made. The most convincing evidence *against* maturational limits on neuroplasticity would be if the L1 (acquired, "hard-wired" and used in an exclusively monolingual context up until adulthood) were subject to different neural activation patterns than those observed in native-speakers.

Surprisingly, there are currently *no* such fMRI/PET studies and *no published* ERP data on L1 attrition in adult immigrants. An unpublished doctoral dissertation by Datta, Obler and Shafer (2007) used ERPs in conjunction with cross-linguistic and cross-modal priming paradigm to examine whether lexical attrition was the result of low L1-use or L2-interference in a group of Bengali learners of English who had been immersed in English upon immigration to the USA (AoA > 10). Participants were divided into two subgroups (L1-dominant and L2-dominant) based on their performance on a category-fluency task. Only L2-dominant individuals were included in the ERP study. Experimental stimuli consisted of words (presented as a picture followed by an auditory recording) in four "familiarity" conditions (HighBengali-HighEnglish, LowBengali-HighEnglish, HighBengali-LowEnglish, LowBengali-LowEnglish). Participants were asked to implicitly name each picture in whichever language came to mind first. They then heard an auditory recording of the word in either Bengali or English, and were instructed to perform a syllable-counting task (where they had to decide whether the word had one vs. more syllables) while ERP responses were recorded. Datta and colleagues hypothesized that if reduced L1-use led to attrition effects, Bengali words from all four familiarity conditions would elicit longer reaction times and larger N400s than English words. Conversely, if L2-interference were the main cause of attrition effects, all Bengali words except those from the HighBengali-LowEnglish condition (i.e., no interference from L2-English) should elicit longer RTs and larger N400s compared to English words. Behavioral results revealed the longest RTs in Bengali (L1) when the word was highly familiar in both languages (HighBengali-HighEnglish), and the longest RTs in English (L2) in response to HighBengali-LowEnglish words. ERP patterns showed a larger-amplitude N400 in response to Bengali (L1) and English (L2) words in low-

English conditions, regardless of the relative familiarity in Bengali. Datta and colleagues also reported that participants' self-reports of L1-use and L2-proficiency predicted their performance in their L1-Bengali.

From these results, the authors argued that the main cause of L1-attrition in L2-dominant individuals is L2-interference (rather than low L1-use). However, the results may not clearly be attributed to interference (assumed to be entirely related to relative familiarity levels) due to the implicit and uncontrolled nature of the picture-naming task. Given that the size of a *priming effect* (shorter RTs and N400 reduction) would directly depend on the language in which the picture had been implicitly named (due to phonological activation during lexical access), and given that the researchers had no reliable way of knowing which language was activated when the picture was shown, the task introduces a methodological problem that precludes the interpretation of the data as clear effects of familiarity and L2-interference.

In the domain of morphosyntactic processing, the only other research group to be using ERP to investigate L1-attrition and L2-acquisition in adults is the group led by Monika Schmid (see Schmid, 2013; Bergmann et al., 2013). In a large-scale project, Schmid and her colleagues are currently examining the processing of verb form and of grammatical gender in Dutch and German in L1-attriters (living in an English-speaking context), as well as early (AoA < 16) and late (AoA > 16) L2 learners of Dutch and German (with a variety of L1s either with grammatical gender (e.g. Turkish) or without grammatical gender (e.g., Polish)), compared to monolingual native-speakers of Dutch and German. Their auditory stimuli include non-finite verb form violations (e.g., *the rose has *blossom*), and gender agreement violations between determiner-noun (e.g., *the*$_{neut}$ *garden$_{masc}$), or determiner-adjective-noun (*the*$_{masc}$ *fresh* *grass$_{neut}$).To date, their results indicate that L1-attriters perform like native-speakers in their L1 on behavioral tasks and elicit P600 effects that were indistinguishable from native-speakers for all three kinds of violations. In contrast, late L2-learners differ both from the L1-attriters and the monolingual native-speakers – while they elicit a P600 effect in response to finiteness violations (though less robust than the native-speaker groups), the P600 effect is absent in response to the two gender agreement violations. Based on these findings, Schmid and colleagues argue in favor of the CPH, whereby one's L1 is privileged and remains native-like despite limited use over a number of years, whereas learning an L2 is subject to maturational effects and decreased brain plasticity after puberty.

Given that Schmid is one of the forerunners of the field of L1-attrition, it is no surprise that her research program was conceived and developed with thorough consideration of important theoretical and methodological questions that had eluded previous studies. However, in choosing to test gender agreement in languages such as German and Dutch where the noun does not carry much morphosyntactic information in terms of morphological gender markers (in contrast to languages such as Italian or Spanish), it can be argued that what is being tested is primarily speakers' knowledge of lexically-stored knowledge of the idiosyncratic association between a noun and its gender (and, thus, its correct determiner). In their experimental design, many of the words tested are monomorphemic nouns (e.g., *Sonne* (sun), *Blume* (flower)) which provide unreliable cues about the noun's gender. The processing steps involve first retrieving the noun's lexically-stored gender information and feature-checking between the noun and the preceding determiner. However, if the problem lies in the retrieval of the lexicalized arbitrary association between a noun and its gender, this problem is not morphosyntactic in nature. Moreover, although they included a condition where an adjective intervened between the determiner and noun, the adjective is not inflected for gender in German and Dutch, and therefore this manipulation does not allow for testing attriters' and L2 learners' sensitivity to morphosyntactic agreement rules (other than matching a noun to its appropriate determiner based on memorized gender information at the lexical level). Although it is certainly a viable linguistic area to test, it is conceivable that this process is not complex enough to show a breakdown in processing in attriters. It would be interesting to test more complex morphosyntactic properties, especially over longer distances rather than local violations, to determine whether some aspects of morphosyntactic processing do break down in L1-attrition and are subject to difficulties in L2-learners.

Secondly, while these studies assess the contribution of L1-background on native-like L2-processing, they do not systematically investigate proficiency effects in either early or late L2-learners (or among the attriters). The approach taken by Schmid and colleagues is to attempt to match the groups on proficiency level, rather than specifically investigating how proficiency level modulates responses. As discussed in Section 1.5, it is crucial to systematically examine whether proficiency level modulates processing patterns, even if an L2-group as a whole seems to have a relatively high level of proficiency.

In addition, Schmid and colleagues' studies only investigate the *L1* of the attriters (not their L2), and compare their ERP profiles to the *L2* of bilingual (early/late) learners. As discussed in Section 1.7.3, a full picture of attrition effects, transfer effects, proficiency effects and general bilingualism effects can only be obtained if both the L1 and L2 of the bilingual groups are assessed, in comparison to monolingual native-speakers of each language. These are the three main differences between Monika Schmid's ongoing work and my dissertation, and the very aspects which make our ERP data completely novel.

The primary goal of this dissertation is to bridge the gap between existing neurocognitive studies of L2 acquisition and the lack of such investigations in L1-attrition, in an integrative behavioral and ERP investigation that explores highly topical and unresolved questions of neuroplasticity and of the role of exposure, proficiency and dominance in shaping the brain's responses to language.

1.9. The present studies

1.9.1. Aims and research questions

This dissertation comprises a series of behavioral tasks and three distinct ERP studies conducted in two languages – Italian and English – with four groups of speakers: (1) Italian-English first-generation immigrants (i.e. "*attriters*"); (2) English native-speakers who acquired Italian as an L2 in adulthood (i.e., "*late L2 learners" or "bilingual controls"*); (3) Italian native (monolingual) speakers in Italy (i.e. *Italian native controls*); and (4) English native (monolingual) speakers (i.e., *English native controls*).

With respect to **L1 attrition**, this work aims to answer the following research questions:

(1) Is there evidence of L1 attrition in a group of individuals who immigrated to a new country in *adulthood*, or rather, is the native-L1 stable because it was acquired and entrenched within a maturational "critical period" early in life, after which the brain was no longer plastic?

(2) Are attrition effects observed at the level of linguistic *behavior* (proficiency tasks, online acceptability judgment task) and at the level of the *brain* (ERP signatures during real-time language comprehension), or might there be a dissociation between the effects seen in the brain and behavior?

(3) Are there areas of language processing in which attrition effects are more pronounced than others (e.g., vocabulary, grammar, specific areas of crosslinguistic competition, and/or areas where processing demands are high, such as in the case of complex stimuli involving long-distance dependencies, inhibition of contextually-inappropriate meanings, and costs of disconfirmed predictions)?
(4) Are attriters' ERP responses modulated by L1 proficiency level such that lower L1-proficiency scores are associated with a greater degree of deviance from processing patterns in native-controls? In other words, is proficiency as important a factor in predicting neurocognitive processes underlying L1 attrition as it has been shown to be for L2 processing?
(5) Do attriters resemble late L2-learners when they process Italian, such that L1 attrition might be considered along the same proficiency-modulated continuum?
(6) How well do different proficiency tasks, control tasks and experiential factors such as age at immigration, length of residence, amount of L1/L2 use predict attriters' processing patterns?
(7) Does attrition manifest itself as an effect of low L1-proficiency overall, cross-linguistic interference from the L2, and/or as something *more* that differs both from native-monolingual controls and late L2 learners, that is specific to their linguistic situation?

With respect to **L2 acquisition**, the studies aim to answer the following research questions:
(1) To what extent and in which areas/tasks are the two groups of late L2 learners (i.e., attriters in English, L2 learners in Italian) similar to native-speaker controls in their ERP response patterns?
(2) Is nativelikeness in L2-processing more strongly modulated by proficiency or by age-of-acquisition (AoA)? In other words, is there evidence of ongoing neuroplasticity in adulthood, or are L2-processing mechanisms constrained by maturation, regardless of proficiency level?
(3) Which approach(es) to exploring proficiency effects on L2 processing may be most informative (global vs. structure-specific proficiency, online vs. offline tasks, proficiency as a categorical or continuous variable?).

1.9.2. Description of ERP studies

Study 1 investigated morphosyntactic processing of *L1*-Italian, thus comparing the native-Italian controls to the L1-attriters. The aim was to investigate if L1-attriters differed from Italian native-speakers in their automatic detection and online repair (revision, re-analysis) of number-agreement violations. Rather than testing salient violations involving only local mismatches (e.g., determiner-noun, subject-verb), we also assessed sentence processing where number-agreement had to be evaluated over a longer distance in a sentence, across intervening words. Our experimental design manipulated number agreement between three sentence positions: (a) subject, (b) verb and (c) an adjective modifying the subject-noun (e.g., *The workers return from the factory dirty with grease*), with four experimental conditions reflecting all possible combinations of (dis)agreement between these three positions (Molinaro et al., 2011). In assessing ERP correlates on *two* target words within the sentence (verb and modifier), we were able to examine how real-time comprehension unfolds at different points in time within a single sentence, but also to what extent number-agreement errors may be resolved or repaired online before new linguistic information is integrated and a decision regarding sentence-interpretation is made.

A crucial aim was to determine whether processing strategies were modulated by Italian proficiency level even in these two groups of *native* Italian-speakers, and/or whether differences may characterize the attrition group as a whole. With respect to the attriters, we hypothesized that, if long-distance agreement computation and online repair/re-analysis processes in the L1 are vulnerable to changes in adulthood, group differences would be observed on both the verb but perhaps especially the modifier. In addition, we expected that L1-Italian proficiency level would modulate ERP patterns in response to agreement violations in *both* groups of native-speakers, with lower-proficiency speakers exhibiting differences in the presence, amplitude, latency, duration, and/or scalp distribution of ERP components of interest (LAN/N400, frontal positivity/P300, posterior P600).

Our findings revealed group differences between native-controls and L1-attriters in ERP signatures of morphosyntactic processing, on both target words. We explain the differences observed in attriters compared to native-controls (1) as suggestive of crosslinguistic influence/transfer of L2-English morphosyntactic processing patterns when reading in L1-Italian, and (2) as evidence for reduced repair/re-analysis processes during real-time sentence comprehension in attriters, particularly later on in the sentence (i.e., on the modifier). In addition to group differences, ERP effects were also strongly modulated by L1-Italian proficiency level, with lower-proficiency individuals eliciting less robust ERP responses (that also differed in their scalp distribution) relative to high-proficiency individuals, even if all individuals were processing their native-L1. Interestingly, attriters did not differ from native-controls by the end of the sentence, as their offline acceptability judgment ratings of sentences were not statistically different from those of native-controls. However, attriters showed significantly slower reaction times in their responses, further suggesting that processing was less efficient in attriters. These results therefore provide some of the first ERP evidence of attrition effects in morphosyntactic processing, and emphasize that subtle processing differences may exist even at beginning stages of L1 attrition, especially for complex sentences containing long-distance dependencies where agreement has to be verified over several constituents. These processing differences can be ascribed to L1-proficiency effects but also L2-to-L1 transfer, as well as to a more general change in the online elaboration of structural solutions to ungrammatical input.

Study 2 examined lexical-semantic processing of sentences containing confusable Italian words – minimal pairs that differed in their final vowel (e.g., *cappello* vs. *cappella*) but also in their lexical-semantic meaning (*hat* vs. *chapel* respectively). These lexical items are difficult to master for L2 learners due not only to their confusability in form but also the interface with gender, as the final vowel coincides with the morphological marker for gender in Italian. We aimed to examine whether semantically-anomalous sentences where the target noun was substituted with its confusable orthographic neighbor (such that *cappello* (hat) occurred in the *cappella* (chapel) context, and vice versa) would be more likely to be processed as semantically-correct sentences by late L2 learners as well as by L1 attriters, especially for those individuals with lower Italian proficiency levels. We compared N400 and P600 effects in response to sentences where the intended target word was swapped with its minimal pair ("*Swap*" condition), as well as to sentences where an orthographically unrelated word was erroneously inserted into the sentence context ("*mismatch*" condition), compared to semantically-correct sentences (e.g., *The fisherman wears the <u>hat</u>/*<u>chapel</u>/*<u>chin</u> of wool*).

The "*Swap*" condition arguably constitutes a more subtle or difficult to detect lexical-semantic violation compared to outright semantic violations, and we expected group/proficiency differences to be most robust for this experimental condition, even in the case of native-Italian attriters. We sought to determine whether lexical access and integration difficulties in L1 attrition also occur during L1-

comprehension (as the majority of the literature has focused on lexical difficulties in *production*), and whether these difficulties may also occur as *intralinguistic* competition (i.e., *within* the L1) as opposed to crosslinguistic competition. A second aim was to compare L1-attriters to late L2-learners of Italian and to systematically explore proficiency effects on ERP responses in both groups to determine whether a proficiency-based continuum of processing patterns may be observed, regardless of L1/L2 status (i.e., of AoA).

Our results indicated that sentences with "confusable nouns" were processed differently than traditional ("*Mismatch"*) violations in lower-proficiency Italian speakers, who did not elicit significant N400 effects for "*Swap"* errors relative to correct sentences, regardless of whether Italian was the L1 or the L2. Crucially, ERP response profiles followed a continuum of nativelikeness predicted by Italian proficiency scores – high-proficiency Italian attriters and high-proficiency Italian-learners were indistinguishable from native-monolingual controls, while attriters and L2 learners in the *lower* proficiency range elicited significantly reduced N400 effects compared to native-monolinguals. Attriters and late L2-learners did not differ significantly in their ERP patterns when they belonged to the same proficiency subgroup. An additional finding was that attriters showed an enhanced P600 effect in response to both kinds of lexical-semantic anomalies compared to native-Italian monolinguals and late L2-learners of Italian, which we will discuss in the context of increased conflict-monitoring and a more explicit second thought or double-take further downstream from the lexical-semantic anomaly.

These findings therefore provide evidence that even L1-natives who acquired a language since birth and lived in an exclusively monolingual context until adulthood may resemble late L2 learners in their brain's responses to language, and that proficiency is a key factor in predicting native-like neurocognitive profiles, irrespective of whether the language in question was acquired from birth or in adulthood. Study 2 further supported the initial evidence from Study 1 that attrition effects may be *more* than just L1-proficiency modulations and instances of linguistic competition that may be shared with other bilingual speakers of the same language pair, but may further manifest itself as differences at *later* stages of processing (e.g., the P600 and slower reaction times), reflecting more conscious, effortful, elaborated processes or increased attention in this special population of bilinguals.

Study 3 also explored lexical-semantic processing but was conducted in English (i.e., in attriters' L2 and late Italian learners' L1, compared to a group of monolingual English native-speakers) and, in contrast with Study 2, tested *crosslinguistic* competition during lexical access and integration. The aim was to assess whether the Italian lexicon of L1 attriters and L2 learners was automatically co-activated while reading English sentences where the target words were interlingual "false-friend" homographs (*estate* (property vs. summer) or cognates (*music/musica*) of Italian words and whether such co-activation was modulated by proficiency levels, especially Italian-English *relative* proficiency. Our hypothesis was that L1-to-L2 co-activation (cognate facilitation and/or homograph interference) should decrease with advancing L1 attrition (as measured by Italian-English proficiency levels, L1/L2 use and length of residence). Italian-English Attriters were found to elicit native-like N400 effects that were indistinguishable in amplitude, latency and scalp distribution from both monolingual and bilingual native-speakers of English for those conditions which constituted English lexical-semantic violations. On those conditions where we created a conflict between English and Italian readings of the target-word or created a sentence context that was congruent with *both* languages, we observed parallel activation of Italian meanings for both bilingual groups, albeit with differences in the time-course of co-activation (N400 vs. P600).

This is the first ERP study to examine how a shift in L1 dominance and proficiency brought on by changes in linguistic environment and language exposure (i.e., attrition) affect mechanisms of bilingual lexical access during sentence processing. Although we did not explicitly test L1 attrition effects in Study 4 (given that we examined attriters' *L2*), we found that L1 co-activation effects during L2 comprehension were influenced by background factors such as proficiency, language exposure and length of residence, with more English native-like ERP responses (i.e., less co-activation) associated with increased English proficiency, decreased L1 language exposure, and increased length of residence. Interestingly and in contrast with the behavioral literature on bilingual lexical activation, behavioral responses did *not* mirror the processing patterns observed with ERPs, suggesting that our end-of-sentence acceptability judgment task was not sensitive to processing differences between experimental conditions nor between groups. A key behavioral task that was correlated with bilingual individuals' P600 responses was a verbal translation task (production) where attriters and L2-learners were asked to translate false-friend lexical items and matched-control items from Italian to English. This finding showed that structure-specific proficiency (i.e., behavioral performance on a measure closely related to the experimental task) can be an important predictor of ERP responses.

1.9.3. Additional studies not reported in the dissertation[1]

Four additional studies were designed and run to answer the research questions highlighted in *Section 1.9.1*, but are not included in this dissertation for fear that this body of work would exceed five-hundred pages. A brief mention of these studies is important to have a better sense of the structure of our experimental sessions, as well as of the big picture of our investigations.

The first was the verbal translation task mentioned above in the context of Study 3, which was performed by bilingual participants as part of the set of behavioral proficiency measures. Participants were shown single words on a computer screen and had to translate each word aloud, as quickly and accurately as possible, either from Italian to English, or vice-versa (in two separate experimental sessions). Half of the stimuli were "false friends" between the languages (e.g., *'fabbrica'* in Italian = *'factory'* ≠ *'fabric'* in English; whereas *'fabric'* in English = *'tessuto'* ≠ *'fabricca'* in Italian), while the other half were frequency- and length-matched control items. Frequency was also manipulated such that half of the items were low-frequency in the target language and half were high-frequency. Accuracy, reaction time and types of errors were analyzed. The aim of this experiment was to obtain an additional production measure of proficiency for the bilingual participants, as well as to determine whether experimental evidence supports attriters' anecdotal reports of confusing similar words during production, especially of English words intruding into their production of Italian. The preliminary findings of this study were reported in Kasparian, Vespignani and Steinhauer (2013b).

Another pair of experiments not included in this dissertation examined the impact of cross-linguistic differences in the processing of relative clauses in Italian (Exp. 1) and English (Exp. 2). The aim of these studies was to investigate whether L1-attrition effects can also be described as L2-induced changes to the L1, and whether cross-linguistic transfer in attriters may occur in a different direction and/or to a different degree than transfer in other non-attriting bilinguals (i.e., late L2 learners). Certain cross-linguistic differences between Italian and English morphosyntax make the study of relative-clauses highly relevant. Firstly, the languages differ in the kinds of word-orders that are grammatically acceptable; the standard and most-preferred

[1] It was recommended by the internal examiner to remove this section from the dissertation. However, after careful consideration, we felt it necessary to include this contextual information about additional tasks and filler sentences that participants were exposed to during the experimental sessions.

word orders in both Italian and English are V-NP-subject (e.g., *The cat that chases the mice runs in the garden*) and NP-V-object sentence constructions (e.g., *The mice that the cat chases tremble with fear*). However, V-NP-object (e.g., *The mice that chases the cat tremble with fea*r) and NP-V-subject (e.g., *The cat that the mice chases runs in the garden*) orders are also possible in Italian – though less preferred – whereas they are completely ungrammatical in English. Secondly, the languages differ in the linguistic cues that speakers rely on during sentence comprehension; it has been shown that Italian readers rely on semantic cues and subject-verb agreement to resolve ambiguous sentences (due to the relatively free word-order of Italian), whereas English readers rely primarily on word-order (see *Competition Model*, MacWhinney & Bates, 1989).

We tested whether prolonged and predominant L2 use/exposure changes the cues that attriters rely on to interpret sentences in their L1, resulting in non-native (i.e., English-driven) preferences and processing patterns in Italian compared to English. Both experiments were identical in their design and tested the four word-order conditions outlined above. In order to make semantic cues more salient, we created sentences that contained strong agent-patient relationships with a semantically-biasing verb (*e.g. policeman/arrest/thief*) – a cue we expected to be beneficial only in Italian (not English), and only for Italian-natives but not L1 attriters if their processing of Italian L1 grammar was heavily influenced by properties of their dominant English L2. Given that late L2 learners have been shown by some researchers to exhibit difficulties and engage in non-native-like processing patterns for complex morphosyntactic structures that involve long-distance dependencies (see Clahsen & Felser, 2006a, b), it remains to be seen whether these arguments hold true for the processing of relative-clause structures in Italian in high-proficiency L2 learners.

At the time this dissertation was submitted, only the data from L1-attriters relative to Italian monolingual controls had been thoroughly analyzed (see Kasparian, Vespignani & Steinhauer, 2014a, Kasparian & Steinhauer, 2015). While reading in Italian (Exp 1), L1 attriters showed influence from English morphosyntax – L1-attriters were more likely than Italian native-controls to give significantly lower ratings and to elicit ERP patterns consistent with morphosyntactic violation effects in response to those orders which are ungrammatical in English but acceptable in Italian (V-NP-object and NP-V subject sentences). Italian-native controls, on the other hand, only elicited mild N400 effects (V-NP condition) and P3a effects

(NP-V condition) consistent with dispreferred sentence structures and re-interpreting the subject-role using lexical-semantic information. With respect to the English experimental sentences (Exp 2), we expected L1 attriters to be native-like (i.e., similar to English monolingual controls) and to once again process V-NP-object and NP-V subject sentences as word-order violations. We are currently examining the impact of Italian-English proficiency level and/or language exposure on attriters' processing patterns in each language, as well as the data from the late L2 learners of Italian. To date, however, the results of this study corroborate our findings from Study 1 on number agreement processing that L2-to-L1 transfer effects affect morphosyntactic processing patterns in adult L1-attriters.

Finally, we designed and conducted an ERP priming experiment on Italian regular and irregular verbs. Participants were presented with two single words – the prime (a correct or incorrect past participle of a regular or irregular Italian verb), followed by the target (an Italian verb in its correct infinitive form). Their task was to determine whether the two forms of the verb were related in semantic meaning. Four different kinds of verb errors were compared, relative to their respective correct control conditions: (1) *over-regularized irregulars* (e.g., 'correre' (to run) primed by '**corruto*' vs. '*corso*'); (2) *over-irregularized regulars* (e.g., 'preferire' (to prefer) primed by '**preferto*' vs. '*preferito*'); (3) *incorrect verb-class* (e.g., '*credere*' (to believe) primed by '**credito*' instead of '*creduto*'); (4) *semantic mismatch* (e.g., '*accendere*' (to turn on) primed by '**atteso*' (past participle of the verb '*attendere*' (to wait)) instead of '*acceso*'). The aim of the study was to explore whether L1 attriters continue to process regular and irregular verbs like the native-Italian controls, or whether they are more similar to late L2 learners in how they detect (or fail to detect) different types of verb errors. The data from this priming study are not yet analyzed.

To summarize, the general research questions regarding L1 attrition and L2 processing were investigated in a series of studies, three of which are reported in detail in the present dissertation. Study 1 tested *morphosyntactic* processing in *Italian* and compared L1-attriters to Italian monolingual controls. Study 2 and 3 both tested lexical-semantic processing with all three groups of participants – L1-attriters, late L2 learners of Italian and monolingual controls, but Study 2 was conducted in Italian (and investigated lexical-semantic competition or ambiguity *within* Italian) while Study 3 was conducted in English (and explored lexical-semantic competition or ambiguity *between* Italian and English).

2. GENERAL METHODS

The following section details the information about participants, measures and experimental procedures that are common to all three studies in this dissertation, in the interest of reducing repetition in each subsequent paper. Information that is specific to each ERP study (such as experimental design and stimuli creation) will be described in the abridged Methods section of each of the three studies. (Note that the manuscript versions that have been submitted to journals contained a full version of each respective Methods section).

2.1. Participants

2.1.1. Attrition group (attriters)

The target population in our three studies consisted of Italian native-speakers who had immigrated to Canada in adulthood, and had since then become dominant in English. Participants were recruited by placing advertisements in local newspapers and online, but mostly through the collaboration of the *Italian Consulate* and several Italian organizations in Montreal, Canada (such as the *Italian Scientific Community in Canada* and the *Italian Culture Institute of Montreal*).

We explicitly aimed to recruit individuals who (1) were born in Italy and lived there until adulthood (i.e., first generation immigrants), (2) were fully exposed to standard (rather than dialectal) Italian, with minimal exposure to second-languages (especially English) until immigration, (3) became fully immersed in English upon immigration and have since then reached a very advanced English proficiency level, (4) have limited exposure/use of Italian since moving to Canada, and (5) have noticed changes in their Italian fluency over time. It was an important goal for us to attempt to maximize the differences between native-speakers and potential attriters (in order to study the neurocognitive correlates of *attrition*). Thus, we focused on finding individuals who were candidates of L1-attrition, based on their reports of recognizing this phenomenon in themselves or having it pointed out to them by other native-speakers (e.g., family or friends). For this reason – and contrary to the approach taken in some previous studies – we made it clear in our advertisements and in initial correspondences with potential participants that our objective was to investigate how a native-language may (or may not) change after diminished use/exposure to it in a new linguistic and socio-cultural environment.

Twenty-four participants who fit these strict criteria were tested (14 female; *M* age: 36; Range: 25-50). Prior to immigration, these individuals had had some basic exposure to English within the school system in Italy, but reported to have become advanced and dominant in English *only* upon immersion in the L2 environment, in adulthood. In other words, prior to immigration, all participants in our attrition group were monolingual Italian speakers. Individuals' mean age at immigration (also considered their *age-of-acquisition* (AoA) of English) was 28.2 years of age (Range: 18-40). On average, these individuals had spent 9 years in Canada (range: 1-19), although 13 participants in the group had left Italy to live in another English-speaking country (such as the USA or UK), prior to their move to Canada. With this extended period outside of Italy taken into consideration, length of residence (LoR) in an English-speaking country was of 12 years on average (range: 1-26).

Participants in the attrition group unanimously reported a predominant use of English on a daily basis, with minimal use of Italian. Some participants also used French on a daily basis, but use of French was less frequent than the use of English in all but three participants, who reported using English and French equally frequently. As much as we wished to keep French knowledge to a minimum due to its crosslinguistic similarity with Italian, this research was conducted in Montreal, Québec (a Canadian province where French is the official language), therefore it was challenging to recruit attriters who had absolutely no exposure or use of French. However, we did not recruit attriters who were French-dominant, even if their English proficiency level was very high, in order to keep the influence of French as minimal as possible.

Without exception, all participants in the attrition group reported noticing changes to their native-Italian fluency over time, and/or having had these changes pointed out by their Italian family or friends. For many individuals (43%), these changes were primarily described as difficulties in vocabulary (e.g., difficulties in accessing the intended word, semantic intrusions from English and the production of "false friends"), while other participants (9%) reported difficulties only in grammar (e.g., influences of English grammar or awkward non-native Italian constructions). Approximately 48% reported difficulties in *both* grammar and vocabulary. Given the unanimous self-reports of L1-difficulties while living in an L2-environmental context, we refer to these individuals as "*Attriters*".

2.1.2. Italian L2 learners (bilingual controls):

Twenty English native-speakers who were highly-advanced Italian L2 learners were recruited as a bilingual control group (15 female; *M* age: 31.6; Range = 21-51). Individuals in this group were recruited from a number of organizations in the Montreal area: Italian-language programs offered at University level, Italian student associations, language-learning schools offering advanced Italian courses, the Italian Cultural Institute, and an online "*Meetup*" group – a social network consisting of hundreds of Italian speakers of varying levels and linguistic backgrounds who sign up in order to attend events where they could practice with other Italian speakers in Montreal. All recruitment advertisements specified that, in order to participate, individuals had to be (1) learners of Italian as a *second* language, having learned it in adulthood (not in childhood); (2) very advanced in their Italian proficiency level; and (3) native-English speakers.

Participants had learned Italian at an average AoA of 21 years (Range = 15-29), mostly in a classroom-instruction context. Four of the twenty learners had acquired Italian in a more implicit context, either by immersion (e.g. while living and working in Italy), or by frequent contact with Italian friends, colleagues or clients. These individuals did, however, also take at least one Italian course in addition to this implicit exposure. Importantly, none of the learners were heritage speakers who had grown up hearing Italian – or a dialectal variety of it – spoken to them in childhood. As in the case of the Attriters, we recruited L2 learners who were dominant in English though living in a French-speaking province. Despite the potential influence of French on Italian, it can be argued that the two groups of bilinguals were more comparable in their linguistic repertoire than if we had tested English-Italian L2 learners living in Italy.

2.1.3. Italian native-speakers (Italian native-controls)

Thirty Italian native-speakers (17 female; *M* age: 31; Range = 25-54) still residing in Italy were recruited as our Italian control group with the following explicit criteria: (1) born in Italy and living there until adulthood, with no extended periods spent abroad; (2) fully exposed to standard (rather than dialectal) Italian, and (3) with minimal exposure to second-languages (especially English), which we operationally defined as less than five hours per week.

Since we attempted to match the Italian native-controls to the attriters and L2 learners on age and education, it was not possible to limit ourselves to the recruitment of University students. Thus, advertisements were posted online and in public places, in addition to areas of the university that were particularly frequented by graduate and post-doctoral students. Older adult speakers were also recruited by old-fashioned word-of-mouth in true Italian-style, in the small towns of Rovereto and Borgo Sacco. Although it was impossible to match the Italian native-speakers and the attriters on the regions they originated from in Italy, we collected this information in the event that we wished to consider potential regional differences in speakers' preferences of certain sentence constructions.

2.1.4. English native-speakers (English native-controls)

Thirty English native-speakers with minimal exposure to second languages (and, crucially, no knowledge of Italian) were recruited as our English control group (20 female; *M* age: 31; Range = 24-47). Although participants defined themselves as "monolingual English speakers", several of them originating from English-speaking Canadian provinces or the USA, they were exposed to some French in Montreal, as were the other two groups that were tested at McGill University. However, we recruited English native-speakers for whom knowledge of French did not go beyond some basic words used to get around town.

In addition to group-specific inclusion criteria, all participants were required to be right-handed (as determined by the *Oldfield Handedness Inventory*), with normal or corrected vision, and without a history of neurological or reading disorders (e.g. dyslexia). Due to the difficulty of finding individuals who met our stringent language-background and proficiency-level criteria *and* who were also well-matched across the different groups on age and education, we accepted to test a few left-handed individuals who met all other requirements. One attriter, one L2 learner and two English native-controls were left-handed. However, these individuals were not excluded from our analyses, as their ERP response patterns were consistent with those of the rest of the group in terms of both the timing and topography of the ERP components of interest.

The Italian native-controls were recruited and tested by the first author at the *Department of Psychology and Cognitive Science* of the University of Trento (Rovereto, Italy). Participants in the three remaining groups were tested at the *School of Communication Sciences and Disorders* at McGill University (Montreal, Canada). Participants were compensated for each experimental session of approximately three hours. Attriters and L2 learners participated in two sessions each (one in Italian, one in English), while the monolingual control groups participated in only one session (Italian only or English only).

2.2. Behavioral measures

Several behavioral measures were administered in both Italian and English to collect information about all participants' (a) demographic information and language history, (b) proficiency levels in each language in a variety of domains, and (c) working memory and reading performance. In cases where these measures did not previously exist in Italian, they were adapted into Italian by the first author and checked by at least two native-speakers.

2.2.1. Language-background measures

Participants in all four groups were asked to complete a background questionnaire containing questions about their demographic information (age, gender, education level, region of birth), as well as questions about the language(s) they had acquired. This questionnaire was largely identical for all groups, except that the one administered to the attriters contained additional questions pertaining to their immigration history and first-language exposure. This questionnaire was designed and administered online via Survey Monkey (https://www.surveymonkey.com/).

Participants in the attrition group completed an additional *Sociolinguistic and attitude questionnaire* (adapted from Schmid & Dusseldorp, 2010, which in turn was partly based on Dostert, 2009). This questionnaire was also administered online prior to the first lab session, and contained questions pertaining to language maintenance, attitudes towards languages and cultures, identity, and intuitions about their daily language use.

2.2.2. Proficiency measures

Participants also completed four proficiency measures: (1) A *written self-report measure* where they were asked to rate their proficiency level on a scale from 1-7 in listening comprehension, reading comprehension, pronunciation, fluency, vocabulary, and grammatical ability; (2) A *written C-test* (English version: Keijzer, 2007 (Appendix 1a); Italian version: Kras, 2008 (Appendix 1b)), where they were asked to fill in the blanks in 5 short texts in which twenty words in each text had been partially deleted; (3) A *written error-detection task* designed specifically for this study, where participants had to detect and correct a number of errors in two separate texts (Appendix 2a and 2b); and lastly (4) A timed *semantic verbal fluency task* where participants were shown two semantic categories, one at a time, ("animals" and "fruits and vegetables") and were asked to produce as many vocabulary items belonging to that category as possible within one minute.

The *C-test* was originally designed as a measure of overall language proficiency (Grotjahn, 1987), and is argued to be more sensitive to proficiency differences at advanced levels of proficiency than the widely used *cloze-test*, which has been found to be reliable only in low-intermediate proficiency learners, with more advanced speakers scoring at ceiling. Thus, the C-test is an appropriate and amply-used instrument in the study of L1 attrition (see Schmid, 2011). Our C-test results were scored following the scoring guidelines developed by Monika Schmid and her colleagues (http://www.let.rug.nl/languageattrition/codingCTest) and using the answer key for the Italian version provided by Kras (personal communication).

The *Error-detection task* was designed by the first author for the purposes of this dissertation. Given that the *C-test* is a production task (fill in the blanks) whereas the ERP experiment consists of an acceptability judgment task where accurate performance depended on participants' ability to detect anomalies and errors in written sentences, it was believed that an offline error-detection task may be more strongly correlated to participants' performance in the ERP experiment. In each language, two texts of approximately 180 words in length were adapted from published textbooks or online newspapers/magazines. The types of errors that were introduced in the sentences consisted of grammatical, word-choice and word-order errors that were similar to those specifically tested in the separate ERP experiments, and that also tapped into potential sources of difficulty for Italian-English bilinguals which were not explicitly explored in the current studies (such as tense, aspect, definiteness, etc). The purpose of this test was to obtain a better gauge of structure-specific proficiency (targeting potential sources of cross-linguistic transfer), in an exercise that paralleled the task during the EEG session, albeit in an offline, natural reading task.

The *semantic verbal fluency task* was used as a measure of lexical proficiency. This task has been used in previous literature to examine lexical access in both healthy and clinical populations. We chose the two semantic categories that have been used most frequently in investigations of language attrition, so that we may compare our Italian results to studies in other languages (see Schmid 2011). Participants were shown each category on a computer screen for one minute each, and their responses were audio-recorded and transcribed. We scored this task by calculating the average productivity across the two categories. In a given category, we counted all vocabulary words that were intelligibly produced in their entirety (i.e. not counting incomplete words), except for repetitions, intrusions from the other language, singular and plural mentions of the same word (e.g. mice, mouse), or mispronunciations that resulted in a non-word (e.g. "rhinoceront" in English (for "*rinoceronte*" in Italian)).

Attriters and L2 learners performed all these proficiency measures in both their L1 and L2. Semantic categories were identical across the two languages (to permit comparison) but their order was counterbalanced. Participants in the two native-speaker control groups performed all the same proficiency measures as the attriters.

2.2.3. Working memory and reading performance

Participants were also asked to complete two control tasks: (1) A timed *reading fluency task* where they had to silently read and answer as many true-false statements as possible in three minutes (English version by Woodcock et al., 2001 (Appendix 3a); adapted into Italian by first author (Appendix 3b)); and (2) The *letter-number-sequencing task* from the WAIS-IV as a measure of Working Memory (English version: Wechsler, 1997 (Appendix 4a); Italian version: Orsini & Pezzuti, 2013 (Appendix 4b)).

The *Woodcock-Johnson reading fluency task* was used to determine participants' reading fluency and accuracy, especially given that the sentence stimuli in our ERP studies were visually presented. This timed pen-and-paper task consisted of silently reading a list of 98 simple sentences and determining whether each statement was true or false. The sentences contain high-frequency lexical items and simple grammatical structures to ensure that reading fluency is not confounded with language proficiency. The score reflects the total number of correctly answered statements in three minutes, and can also be calculated as a ratio (number of correctly answered statements / total number of answered statements). The Italian version of this test was adapted from English by the first author for the purpose of this dissertation. The sentences in the Italian version are not direct translations of the English sentences, but are similar sentences that have been matched item-wise on grammatical complexity, length (in words), and word frequency with their English equivalents. The Italian version was independently checked by three native Italian speakers and modified accordingly.

As a measure of *Working memory* (WM), participants had to perform a *letter-number-sequencing task* in which they were shown strings of letters and numbers, one character at a time, in the center of a computer screen (e.g. H, 9, 4), and were asked to recall what they were shown but by re-organizing the sequence, such that they would recall the numbers in ascending order, followed by the letters in alphabetical order (i.e., 4, 9, H). The sequences become progressively longer as the task proceeds, and the experimenter must terminate the exercise after the participant has committed an error on three consecutive trials. Participants' performance is scored by counting the number of correctly answered trials, as well as the working memory *span* (the number of characters in the last correctly-answered sequence). The length of the sequences (thus, WM span) ranges from 2 to 8 characters, and there are 3 consecutive trials of each length (except only 2 trials consisting of 2 characters). The task began with 5 practice trials (sequences of either 2 or 3 characters in length), in order to ensure that participants had understood the instructions and recalled the sequences in proper order (numbers first, then letters), without repeating or rehearsing the characters aloud during the actual presentation of the sequences. Contrary to typical applications of this test, we chose to administer it on a computer screen rather than orally dictated by the experimenter, for two reasons: (1) The presentation of the sequences one character at a time at the center of the screen was highly comparable to the rapid-serial-visual presentation of the words in our experimental sentences during the ERP session; (2) This method allowed for the presentation rate to be consistent across all participants, and removed any experimenter-based variability (e.g., extra cues or distractions) that might have affected participants' performance. When adapting the Italian version into a computer-based task, we ensured that its structure and length was identical to the English version (i.e., same number of trials per memory span), but we did not modify the sequences from the original Italian paper-version. The sequences tested in both languages, however, were not identical. We chose this WM task rather than a sentence-repetition task (Newcomer & Hammill, 1982), in order for our measure to be as little confounded with language proficiency as possible, while still being considered a *verbal* WM measure.

As with the proficiency measures described above, attriters and L2 learners performed these tasks in both their L1 and L2, while participants in the native-speaker control groups performed these tasks in their (only) language. Group means on behavioral/proficiency measures are reported in each study, depending on the language tested (Italian in Study 1 and 2; English in Study 3) and the groups compared (Italian-controls and Attriters in Study 1; all groups in Study 2 and 3).

2.3. **Proficiency**

As one of the main goals of this dissertation was to systematically explore the effects of L1 and L2 proficiency level on the neurocognitive correlates of L1 and L2 processing, considerable attention was dedicated to analyzing proficiency levels. Three methodological approaches were taken: (1) creating participant subgroups of "high" and "low" proficiency for each language on the basis of a composite measure derived from the 3 proficiency tasks (not-including self-reports); (2) treating proficiency as a continuous measure and correlating proficiency scores to ERP profiles; and (3) considering acceptability ratings in response to experimental stimuli as a form of "structure-specific proficiency" and correlating this continuous variable with ERP profiles.

First, for the "subgroup approach", participants in each of the four groups were categorized as "high" or "low" on each individual task (C-test, error detection test, semantic verbal fluency), on the basis of a median split of their scores. After performing a median split of participants' scores on each individual task, we categorized each participant as either having high or low *overall* proficiency by considering all 3 tests together, as a kind of composite measure. If participants were "high" on 2 out of the 3 proficiency tests, they were categorized as being "high" overall. Conversely, if participants were "low" on 2 measures but high on only one, they were assigned to the "low proficiency" subgroup. Often, participants' categorization was consistent across the 3 tasks (H, H, H or L, L, L).

Note that, in our median split, we opted for the *group median* on each task, rather than the median of all scores collapsed across all four groups. After careful consideration, we decided to use the group-specific median on each task because our objective with this approach was to end up with a categorization *within* each group, in order to have a clear idea, for example, of who the most "attrited" individuals were (i.e., lowest Italian proficiency within the attrition group), who the strongest L2 learners were (i.e., highest Italian proficiency within the L2 group, highest English proficiency within the attriters), and who the most "native" native-speakers were among the monolingual controls. The goal was to obtain subgroups of high and low proficiency-levels within each group that we could directly contrast in our visual inspections of the ERP data. By treating each group separately, we were able to compare the highest-proficiency attriters to the lowest-proficiency attriters in the group, and the same for the L2 learners.

In contrast, our correlational analyses treated proficiency as a continuous variable (across groups) and allowed us to separately assess the impact of individual tests in predicting ERP response patterns. For the bilingual groups (L1-attriters and L2 learners), we additionally assessed how "relative-proficiency" (Italian minus English) predicted processing patterns, especially in the context of Study 3 (crosslinguistic competition / co-activation of Italian while reading in English) as well as for the study on relative clauses not reported here. In each language, we took the 3 proficiency tests into consideration and derived standardized (Z) scores for each proficiency task separately, with participants from both bilingual groups combined. Thus, each participant ended up with a standardized score on the C-test, the error-detection task, and the semantic verbal fluency task. We then added up the three standardized scores for each participant and obtained a standardized "composite" score. We computed a composite score in each language (Italian and English) and derived a standardized "relative-proficiency" score by subtracting English proficiency from Italian proficiency for all participants.

In addition to correlating proficiency measures to ERP patterns, we also examined to what extent participants' acceptability ratings (at the end of each experimental sentence) predicted the amplitude of ERP effects of interest. The behavioral task is not timed per se, but participants were instructed to respond as quickly and intuitively as possible, immediately following the prompt at the end of a given sentence. Given that the acceptability judgment task was performed on the same stimuli in response to which ERP effects were analyzed, participants' ratings may serve as an additional, structure-specific proficiency measure that may prove to be more sensitive than overall proficiency measures. On the other hand, we expected that end-of-sentence acceptability ratings may be insensitive to certain types of online processes that take place as comprehension unfolds (such as revision or repair), such that we may not always see a correlation between brain responses and behavioral performance.

2.4. Procedure

Once recruited and informally screened to ensure that all inclusion criteria were met, participants were sent an electronic copy of the consent form, which was specific to each group. They were then emailed the links to the background questionnaires on SurveyMonkey and were asked to complete them before arriving at the lab. These questionnaires served as an additional screening process, and participants who did not meet the requirements were immediately notified.

Upon arrival at the lab, participants signed a hard-copy of the consent form. Participants then completed the behavioral portion of the experiment (verbal and written tasks). Both in Italy and in Canada, the sequence of the behavioral tasks during the testing session was identical (semantic verbal fluency, WM, written error-detection test, written C-test, reading fluency task). In cases where individuals in the L2 learner group were not deemed advanced enough in Italian to qualify for the study based on their performance on the behavioral tasks, they were compensated only for the behavioral portion and did not take part in the EEG portion of the study.

Participants were then shown the EEG equipment, and the procedure of fitting the cap was visually explained to them. While the cap was being fitted, participants read a handout that provided additional instructions as well as some examples of the kinds of sentences they would see during the experiment. They were informed that their task would be to rate the "acceptability" of each English sentence on a scale of 1 to 5, where 1 is used for a sentence that is completely unacceptable to an English speaker (severely ungrammatical or does not make sense), and where 5 is an absolutely perfect sentence. Participants were encouraged to use the entire rating scale, rather than just a categorical judgment of "unacceptable" and "acceptable" using only 1 and 5. Participants had to decide subjectively what types of errors they considered more severe than others, and were encouraged to decide as quickly and intuitively as possible after reading each sentence. Note that all instructions were given in the language of the testing session (Italian or English) in order to situate bilingual participants in the appropriate language mode.

Both in Montreal and in Rovereto, participants were seated in comfortable chair in a dimly-lit, sound-attenuated booth, at approximately 80 cm from the computer monitor with a Cedrus seven-button RB-740 response box placed in front of them (Cedrus Corporation, San

Pedro, CA, USA). A short demonstration was given to show them how eye-movement, blinks and muscle movement create artifacts in the signal, and participants were encouraged to blink only when they saw an image of an eye appear on the screen. They were also instructed to carefully read each sentence until the end, given that there could be subtle variations in acceptability between similar-seeming sentences. They were instructed to provide their response as quickly but as accurately as possible once the prompt ("???") appeared. A practice block of twenty sentences, representative of those used in the actual experiment, familiarized participants with the procedure and ensured that they had understood the task. If participants rated all sentences as perfect ("5"), the experimenter(s) attempted to determine whether they had noticed the errors. Two participants (one L2, one English-control) who were unable to perform accurately during the practice session were excluded from the study and compensated proportionally for their time.

Words were presented in white 40-font Arial characters, at the center of a black background. The first character of the first word appeared in uppercase, and the rest of the words were in lowercase. The sentence-final word was presented along with the period. Each trial began with the presentation of a white fixation cross for 500 ms, followed for 200 ms by a blank screen (ISI). Each word then appeared one at a time for 300 ms (+ 200 ms ISI). A visual prompt (“???”) followed the offset of the sentence-final word, indicating the onset of the response interval. The prompt remained on the screen until participants pressed a button from 1 to 5. Immediately after a response was provided, the image of the blue eye appeared at the center of the screen for a 2000 ms interval, encouraging participants to blink their eyes between trials. The next trial began after the blinking interval, with the presentation of another fixation cross. Each session lasted approximately 3 hours, including setup, short breaks and cap removal. For the bilingual participants, the second test session took place at least one week after the first session (but for the majority of participants, the sessions were several weeks apart). All consent forms, materials and procedures were fully approved by the Ethics Review Board of each institution (Faculty of Medicine, McGill University (#A06-B30-11A) and Ethical Committee for Human Research, University of Trento (#2013-003) for the duration of the study.

2.5. EEG recording and analysis

The EEG was recorded continuously from 25 Ag/AgCl electrodes, 19 of which were electrodes mounted on a standard electro-cap according to the 10-20 system (Jasper, 1958), and 6 of which were external electrodes: 4 electro-oculogram (EOG) channels placed above and below the left eye (EOGV), and at the outer canthus of each eye (EOGH), as well as 2 reference electrodes placed on the mastoids (A1 and A2). All electrodes were referenced online to the left mastoid (A1). Impedances were kept strictly below 5 kΩ for scalp and reference electrodes, and below 10 kΩ for EOG electrodes. Signals were amplified using NeuroScan (Canada) and BrainVision (Italy) and filtered online with a band-pass filter of 0.1 to 100 Hz a sampling rate of 500 Hz. Data pre-processing and analyses were carried out using EEProbe (ANT, Enschede, Netherlands).

Offline, EEG recordings were re-referenced to the average activity of the left mastoid (except when otherwise specified) and filtered with a phase-true 0.3-40 Hz band-pass filter. A subset of participants who exhibited severe random drifts that affected random noise-levels in the grand-average data were filtered with a 0.5-30 Hz band-pass filter, in order to avoid excluding difficult-to-recruit Attriters from our sample. With the awareness that such a band-pass filter could affect slow-going waves such as the P600, single-subject data as well as grand-average data of filtered participants were compared to the original data with the standard filter, and the pattern of results was consistent. Trials containing artifacts due to blinks, eye-movements and excessive muscle activity were rejected prior to averaging, using a moving-window (400 ms) standard deviation of 30 microvolts.

Subsequent details about data analysis (quantification of ERPs into time windows, baseline intervals, and exclusion of participants due to low quantity or quality of trials) will be reported in the Methods section of each study.

3. MANUSCRIPT 1

First-language attrition induces changes in online morphosyntactic processing and re-analysis: An ERP study of number agreement in complex Italian sentences

Kristina Kasparian[1,2], Francesco Vespignani[3] and Karsten Steinhauer[1,2]

[1] School of Communication Sciences & Disorders, McGill University, Montreal, Canada
[2] Centre for Research on Brain, Language and Music (CRBLM), Montreal, Canada
[3] Dipartimento di Psicologia e Scienze Cognitive, Università degli studi di Trento, Italy

This manuscript has been submitted to Cortex for publication and is currently under review

ABSTRACT

In the domain of language, it remains an open question whether the neurocognitive mechanisms underlying real-time language-processing are constrained by maturational limits on neuroplasticity, or whether proficiency is the crucial factor modulating the brain's responses to language. "First-language (L1) attrition" offers new light on this debate, particularly when "attriters" lived in an exclusively monolingual context until adulthood and, due to immigration, experienced a shift in exposure and dominance (and eventually, proficiency) toward the late-acquired second-language (L2). To date, the neurocognitive correlates of L1 attrition are largely unexplored. Using event-related-potentials (ERPs), we examined L1-Italian grammatical processing in 24 "attriters" and 30 Italian native-controls. In an experiment based on Molinaro, Vespignani, Zamparelli and Job (2011), we assessed whether (1) attriters differed from non-attriting native-speakers in their online-detection and repair/re-analysis of number-agreement violations, and whether (2) differences in processing were modulated by L1-proficiency. Rather than testing salient violations involving only local mismatches, we manipulated agreement between three inflected constituents and examined ERP responses on two of these (subject, verb, modifier). Our main findings revealed group differences in the amplitude, scalp distribution and/or duration of LAN/N400 + P600 effects (but not frontal positivities between 550-750 ms). Interestingly, on both target-words, the P600 was longer-lasting in native-controls than attriters, reflecting more elaborated re-analysis processes. In addition, L1-Italian proficiency modulated ERP effects on the modifier: lower-proficiency individuals elicited a smaller, less frontal and longer-lasting N400 and a smaller P600 than high-proficiency individuals. Proficiency-level modulated the P600 between 650-900 ms, whereas the late-P600 (beyond 900 ms) depended on group-membership and on amount of L1 exposure in attriters. Our results provide the first ERP evidence of L1 attrition in morphosyntactic processing, thus challenging claims of decreased neuroplasticity beyond a maturationally-delimited "critical-period". Our results emphasize that proficiency-level predicts language-processing profiles, and that the P600 is not a monolithic component.

Keywords: *neuroplasticity, morphosyntactic processing, first-language attrition, P600, event-related potentials (ERP)*

1. INTRODUCTION

For more than half a century, a highly controversial question in second language (L2) research has centered on the existence of a neurobiological "critical period" for language-learning. The claim held by many is that maturational limits constrain L2 acquisition, such that an L2 acquired in late childhood or adulthood must rely on different neurocognitive substrates and processes than those used for the first language (L1), as a result of decreased neuroplasticity (Lenneberg, 1967; Penfield & Roberts, 1959; see also Bley-Vroman, 1989; Ullman, 2001). Conversely, one's L1 has a privileged status and remains stable, as a result of having been hard-wired or "entrenched" in the brain during this early critical period (Marchman, 1993; Penfield, 1965).

A number of neuroimaging and neurophysiological studies of L2 acquisition have corroborated the behavioral evidence in support of this theory – L2 learners with a late *age-of-acquisition* (AoA), who began to learn the L2 after puberty or in adulthood, have been shown to deviate from native-speakers and early L2 learners in the neurocognitive mechanisms underlying language-processing (e.g., Johnson & Newport, 1996; Hahne & Friederici, 2001; Kim et al., 1997; Weber-Fox & Neville, 1996). However, this seemingly straightforward picture was complicated by a factor whose role was initially overlooked; in much of this research, late-AoA was confounded with a low L2 *proficiency level*, rendering it difficult to determine whether any differences observed at the brain level in late L2 learners were indeed due to the advanced age at which the language was learned, or whether insufficient exposure and a low level of language mastery leads the brain to engage in non-native-like, compensatory processing strategies.

Since this shortcoming was pointed out, research has been directed towards exploring the effect of proficiency level on the neurocognitive processes underlying a late-acquired language. Many studies have revealed that L2 learners at high levels of proficiency (a) do exist and (b) show processing patterns that are indistinguishable from those of native-speakers, even if this language was acquired in adulthood, thus casting doubt on a maturationally-delimited critical period (Bowden et al., 2014; Friederici, Steinhauer & Pfeifer, 2002; Osterhout et al., 2006; Rossi et al., 2006; Steinhauer, White & Drury, 2009; White, Genesee & Steinhauer, 2012). The crucial role of proficiency level in modulating the brain's responses to language was further emphasized in studies conducted with *monolingual* native-speakers (Pakulak & Neville, 2010; Prat, 2011). Despite these findings, however, others still argue that the role of AoA cannot be ignored (Clahsen & Felser, 2006; Hahne, 2001; Ojima et al., 2005; Pakulak & Neville, 2011), that the effects of both factors are independent (e.g., Moreno et al., 2005; Wartenburger et al., 2003), that proficiency level has not been adequately measured (e.g., Newman et al., 2012), and that the linguistic structures or tasks being investigated are far too easy to reveal subtle AoA-related differences (e.g., Abrahamsson & Hyltenstam, 2009). In sum, although recent research has focused on clarifying the relative role of AoA and proficiency on neurocognitive mechanisms involved in language, the controversial question of what exactly determines "native-like" language processing is still open to empirical debate.

The present work contributes to this controversial question by turning to a unique L2-learning situation – the case of "*first-language (L1) attrition*" – where AoA and proficiency-level operate in opposite directions. In our studies, first-language *attriters* are operationally defined as first-generation immigrants who move to a new country in adulthood and who, due to full immersion into the majority-L2 environment and limited use of their native-language, experience a shift in dominance from the L1 to the L2. These individuals anecdotally report that they have experienced a gradual decline in automaticity, fluency and, eventually, proficiency in their native-language (a phenomenon defined as "*attrition*"), whereas their proficiency level in the late-learned L2 continues to strengthen.

This situation where individuals experience non-pathological difficulties in an L1 they had fully acquired and used as a native-speaker up until adulthood provides a unique test of the critical period hypothesis, and sheds new light on the controversial question. If attrition results in changes in online language-processing mechanisms in these individuals' L1, this would constitute evidence of ongoing neuroplasticity for language, even in adulthood, and would indicate that one's L1 is not as stable and hard-wired as advocates of the critical period claim would assume. Furthermore, if L1 processing patterns in attriters were found to be modulated by proficiency level, such that individuals with lower levels of proficiency differed most from non-attriting native-speakers (although the groups do not differ on AoA), such a finding would lend convincing support that the neurocognitive substrates underlying language are strongly determined by proficiency, whether the language in question is the L1 or the L2. The study of L1 attrition constitutes, therefore, a logical and necessary bridge between L1 and L2 acquisition research, and may reveal that attrition and acquisition lie on the same continuum, with proficiency level modulating the brain's response patterns to language.

1.1. First language attrition and the brain

The phenomenon of L1 attrition has been extensively documented in behavioral research, revealing non-native linguistic performance in a number of linguistic domains, from pronunciation and fluency in production (e.g., De Leeuw, Schmid & Mennen, 2010; Schmid & Beers Fägersten, 2010), to difficulties with vocabulary (e.g., de Bot, 1996; Köpke, 1999; Montrul, 2008; Opitz, 2011), as well as with morphosyntax (e.g., Ammerlaan, 1996; Schmid, 2010; Schmid & Köpke, 2011). Due to the prevalence of attrition effects in the lexical-semantic domain, it has been argued that linguistic elements served by declarative memory – such as vocabulary – are more susceptible to attrition than those which rely on procedural memory – such as morphosyntax (Paradis, 2003; 2007). However, given that behavioral evidence investigating this claim is mixed, it remains an open question whether morphosyntactic processing is indeed more resistant to attrition.

To date, researchers have barely begun to turn to neuroimaging approaches such as fMRI or event-related-potentials (ERP) to investigate the neurocognitive correlates of language attrition, although these methods have been widely used in investigations of L2 acquisition. A neurocognitive investigation of L1 attrition in adults is key to advancing our knowledge of the phenomenon, and would allow researchers to address many questions that are simply unanswerable by behavioral studies. For example: (1) When an individual shows behavioral "symptoms" of attrition, do L1 processing patterns in the brain deviate from non-attriting native-speakers? (2) Could attrition effects be observed at the brain-level in linguistic domains or certain structures/tasks where we fail to find attrition effects behaviorally? (3) What is the impact of proficiency level (and other experiential factors such as length of residence and language exposure) on the neurocognitive processes underlying an attriting L1?

Although a few studies have looked at the extreme case of language attrition in young international adoptees, it can be argued that these are situations of "incomplete L1 acquisition" (Montrul, 2008), as the termination of L1-exposure occurs in early childhood, when pro-CPH researchers would argue that neural networks are still plastic. Evidence of attrition in these individuals would therefore not challenge the CPH. The next logical step would be to turn to speakers who fully acquired the L1 into adulthood before experiencing a shift in linguistic environment, as this is the population in which the strongest evaluation of the CPH can be made. The most compelling evidence in favor of ongoing neuroplasticity would be if the L1 – acquired, "hard-wired" and used in an exclusively monolingual context up until adulthood – were subject to different neurocognitive processes than those observed in native-speakers. Astonishingly, there are currently no such fMRI studies and no published ERP data investigating L1 attrition in adult immigrants (but see dissertation by Datta, Obler & Schafer (2007) for ERP evidence of lexical attrition; Schmid et al., forthcoming).

One research group led by Monika Schmid and colleagues (Bergmann, Berends, Brouwer, Meulman, Seton, Sprenger and Stowe) is currently conducting studies on L1 attrition using ERPs in German and Dutch attriters living in an English-speaking context, compared to monolingual native-speakers of German and Dutch, as well as different groups of early/late L2-learners with different L1-backgrounds. Their auditory stimuli include non-finite verb form violations (e.g., *the rose has *blossom*), and gender agreement violations between determiner-noun (e.g., *the*$_{neut}$ **garden*$_{masc}$), or determiner-adjective-noun (*the*$_{masc}$ *fresh *grass*$_{neut}$).To date, their results indicate that L1-attriters perform like native-speakers on behavioral tasks in their L1 and elicit P600 effects of similar amplitude, distribution and latency compared to native-speakers, for all three kinds of violations, whereas late-L2 learners deviate most from native-speakers (Bergmann et al., 2013; Schmid, 2013). These results have been interpreted as favoring the critical period stance that the L1 is robust and remains native-like even despite a shift in dominance towards the L2.

Although our work developed in parallel and independently from the work by Schmid and colleagues, two interesting distinctions between their study and ours are worth noting. First, in choosing to test gender agreement in languages such as German and Dutch, it can be argued that Schmid and colleagues are not investigating morphosyntactic processing per se, but rather speakers' *lexicalized* knowledge of idiosyncratic associations between a noun and its gender (and thus, the correct determiner). Given that an intervening adjective between a gender-marked determiner and a gender-marked noun is not inflected for gender in German or Dutch, it is not possible to test attriters' and L2 learners' sensitivity to morphosyntactic (agreement) rules. In contrast, the current study examines morphosyntactic processing in attriters by investigating number agreement across three separate constituents within a given sentence, all of which are inflected for number in Italian. Furthermore, considering reports that complex sentences and long-distance dependencies are a more reliable source of differences between late L2 learners and native-speakers compared to salient morphosyntactic violations (e.g., Clahsen & Felser, 2006), we opted to test morphosyntactic agreement processing in sentences with multiple inflected constituents, where evaluating grammaticality involved a long span within the sentence, rather than the detection and resolution of local mismatches that involved only lexicalized agreement information.

A second distinction is that a major aim of our study was to systematically examine the effect of proficiency level on ERP responses elicited during L1 morphosyntactic processing in attriters and native-speakers, both by clustering individuals into subgroups of high(er) and low(er) proficiency, as well as by using proficiency as a continuous variable in predicting ERP effects. In contrast to the majority of ERP studies on L1/L2 processing, we used a range of written and oral measures (not limited to self-report scales) to determine proficiency and, crucially, these same measures were administered to our control group of native monolingual speakers. Our behavioral task during ERP recording was also novel, in that we asked participants to rate sentences from 1-5 rather than providing a binary (yes/no) acceptability judgment; we expected this task might tap into more subtle differences between our groups (and might ensure that sentences were read attentively until the end, rather than until the first violation).

1.2. Number agreement processing

Using ERPs and an online acceptability judgment task, our goal was to investigate whether L1 attriters differed from Italian native-speakers in their automatic detection and online repair/re-analysis of number-agreement violations. Number-agreement processing has been widely studied in monolingual native-speakers of different languages (Angrilli et al., 2002; Barber & Carreiras, 2005; De Vincenzi et al., 2003; Hagoort & Brown, 2000; Leinonen, Brattico, Jarvenpaa & Krause, 2008; Molinaro et al., 2008; Osterhout and Mobley, 1995; Roehm, Bornkessel, Haider & Schlesewsky, 2005; Silva-Pereyra & Carreiras, 2007). Given that agreement patterns are subject to crosslinguistic variation, it is of interest to extend this research to bilinguals whose two linguistic systems differ in their expression of number morphology. Italian is a language with a relatively free word-order and a rich morphological marking system where number-agreement is salient and can often constrain the identification of a subject (Bates et al., 1982; MacWhinney & Bates, 1989). In contrast, number-agreement in English is poorly signaled due to a less-detailed system of morphological markers, and speakers instead rely on word-order for sentence interpretation. It is of interest to examine what processing strategies bilingual speakers rely on when their languages differ cross-linguistically on agreement properties, particularly in the special case of attrition.

Number agreement studies conducted with late L2 learners have been suggestive of cross-linguistic influence from the L1 onto the L2, revealing non-native-like processing profiles (i.e., missing LAN and/or P600) in cases where the L2 agreement properties in question did not exist in the speakers' L1 (e.g., Chinese-English: Chen et al., 2007; Japanese-English: Ojima et al., 2005; English-French: Osterhout et al., 2004; English-Spanish: Tokowicz & MacWhinney, 2005). These findings have been taken as support for the critical period hypothesis, whereby speakers' "entrenched" L1 constrains the neurocognitive processes involved in the L2. However, several other studies have highlighted proficiency-based processing differences and have emphasized ongoing neuroplasticity, by showing that L2 learners converge on native-monolinguals' processing patterns with continued learning and high proficiency levels (e.g., Hopp, 2010; Osterhout et al., 2006; 2008; Rossi et al., 2006).

To date, online number agreement processing has not been investigated in L1 attrition. Advocates of the critical period hypothesis, in its strict form, would claim that L1 processing strategies should remain unchanged due to reduced plasticity for language in adulthood. In this view, transfer/influence from the L2 onto the L1 would not be easily accommodated. If instead neuroplasticity were maintained into adulthood, it would be conceivable to expect English (L2) influences on Italian (L1) in attriters who are dominant in the late-acquired L2 rather than the L1. Due to their excellent temporal resolution, ERPs are particularly useful for investigating real-time processing of agreement patterns during sentence comprehension. Three components have generally been associated with the processing of number-agreement violations across languages: (1) **a *left-anterior negativity* (LAN)** elicited between 300-500 ms, reflecting the early detection of a morphosyntactic violation (e.g., Kaan, 2002; Molinaro et al., 2011; Osterhout & Mobley, 1995); (2) **an *early frontal positivity*** between 500-700 ms, argued to reflect difficulties integrating the mismatching constituent with the previous sentence context, particularly in ambiguous or complex sentences (Barber & Carreiras, 2005; Friederici, Hahne & Saddy, 2002; Kaan & Swaab, 2003; Molinaro, Vespignani et al., 2008); and (3) **a *posterior P600*** between 700-1000 ms, indexing morphosyntactic re-analysis and repair once the anomaly has been diagnosed, with larger and more prolonged P600s reflecting costlier repair (Carreiras, Salillas & Barber, 2004; Hagoort & Brown, 2000; Mancini, Vespignani, Molinaro, Laudanna & Rizzi, 2009; Molinaro, Vespignani et al., 2008; Silva-Pereyra & Carreiras, 2007).

Although these components have been reported in the agreement literature, it is not the case that all three effects are reliably elicited in response to any number agreement violation, nor have they been quantified in the same way across studies. The most robust effect elicited in response to number-agreement violations is a P600, which is sometimes preceded by a negativity and/or a predominantly frontal positivity, and sometimes not. LAN-like negativities have been reported for subject-verb number agreement violations in English (Kutas & Hillyard, 1983; Molinaro et al., 2008; Osterhout & Mobley, 1995), German (Roehm et al., 2005), Spanish (Barber & Carreiras, 2005; Silva-Pereyra & Carreiras, 2007) and Italian (Molinaro et al., 2011). Gender and person agreement violations, on the other hand, have been shown to elicit a more broadly-distributed N400 (e.g., Barber et al., 2004; Deutsch & Bentin, 2001; Mancini et al., 2011a, b; Molinaro et al., 2008). Researchers have explained this LAN/N400 dissociation by positing that number is a morphosyntactic feature that signals structural relations within the sentence, whereas computing gender or person agreement information relies more heavily on access to lexical-semantic or discourse-level information (for a review, see Molinaro, Barber & Carreiras, 2011). Other number studies, however, did not find a LAN (e.g., Balconi & Pozzoli, 2005; Hagoort & Brown, 2000; Hagoort et al., 1993; Nevins, Dillen, Malhotra & Phillips 2007, Osterhout et al., 1996), found a negativity with a bilateral-anterior focus (Kaan, 2002; Leinonen et al., 2008) or a broad distribution more akin to an N400 (Coulson et al., 1998; Mancini et al., 2009), or the effect only reached significance when the statistical analysis (often a t-test) was conducted on a small cluster of anterior electrodes (e.g. T3, C3, F3, F7), as was the case in several Italian studies of subject-verb number agreement (Angrilli et al., 2002; De Vincenzi et al., 2003; Mancini et al., 2009; Molinaro et al., 2011). Thus, it is not necessarily the case that a LAN in its pure definition has been reliably elicited across number agreement studies, once methodological details are taken into consideration.

Similarly, the distinction between the prototypical (posterior) P600 and an earlier (more fronto-central) positivity has not consistently been made, given that the majority of number agreement studies quantified their P600 *only* in the earlier time-window between 500-750 ms where others quantified the frontal positivity (Angrilli et al., 2002; De Vincenzi et al., 2003; Hagoort et al., 1993; Hagoort, 2003; Osterhout & Mobley, 1995; Osterhout et al., 1996; Roehm et al., 2005). Only a few studies actually examined separate positivity windows, describing the effect between 500-700 ms as an early phase of the P600 with a more central (if not primarily frontal) distribution, in contrast to a later P600 phase extending from 700 ms until about 900 or 1000 ms and limited to posterior areas (Barber & Carreiras, 2005; Hagoort & Brown, 2000; Kaan et al., 2000; Kaan, 2002; Kaan & Swaab, 2003; Molinaro et al., 2008 and 2011; Silva-Pereyra & Carreiras, 2007). Advocates of the claim that the P600 is not a monolithic component have taken such modulations in scalp distribution and timing as evidence that different positivities reflect distinct neurocognitive processes. In the agreement literature, the early/frontal positivity has been argued to represent the diagnosis of the incongruence while accessing non-syntactic, discourse-level information to detect the source of the error[2] (see Molinaro et al., 2011). The late P600, in contrast, reflects mechanisms of re-analysis and repair that are necessary to establish a well-formed sentence and arrive at a plausible interpretation (see related *'Diagnosis & Repair'* theory by Fodor & Inoue, 1998, discussed in Friederici et al., 2001 for garden-path sentences). The finding of larger "late P600s" in sentence contexts involving costlier repair supports this claim (Barber & Carreiras, 2005; Molinaro, Vespignani et al., 2008; Silva-Pereyra & Carreiras, 2007). Given that agreement studies have not uniformly investigated late positivities as potentially distinct processing stages, it is still not clear what factors may modulate these early/late positivities, and how consistent these ERP effects even are for different kinds of number agreement violations across languages.

A final point to make about expected ERP responses concerns the related notions of component overlap and individual differences, which have been argued to contribute to the inconsistency in observed ERP effects across studies. First, Osterhout has questioned the reliability of the LAN as an index of morphosyntactic violation detection that is distinct from the N400. Instead, he has argued that the LAN is merely what is "left over" after a right-lateralized positivity (P600) cancels out the broadly-distributed negativity (N400) in a time-window where the two effects overlap. A second issue that has been highlighted by Osterhout (1997, 2004) and colleagues Tanner, Inoue and Van Hell (2014) is that the variability in the presence or robustness of distinct ERP components is largely due to individual variability in response patterns within a

---

[2] It should be noted, however, that the early (fronto-central) stage of the positivity is not limited to ambiguous/complex sentences or to agreement errors, but has been reported in response to other morphosyntactic violations such as phrase structure violations (e.g., Bowden & Steinhauer, 2012). The early frontal positivity has also been described as a member of the P300 component family (P3a: Squires, Squires & Hillyard, 1975; in Italian morphosyntactic violations: Mueller, Oberecker & Friederici, 2009; see review in Osterhout & Hagoort, 1999 and Polich, 2007), although many studies have termed it a "frontal P600" (e.g., Kaan & Swaab, 2003). As the discussion of this complex debate is beyond the scope of our current paper, we refer to this effect in our data (and in similar agreement data) as a "frontal positivity", although we also discuss the effect in terms of a P3a in our Discussion.

group of speakers. These authors have shown that a biphasic LAN+P600 pattern at the group level is a byproduct of individual differences in participants' inherent processing strategies: while some individuals within a group show negativity-dominant response patterns (typically consistent with an N400), others reliably elicit a positivity-dominant response, and the averaging process yields a biphasic pattern that is unrepresentative of most individual-participant data (Nieuwland & Van Berkum, 2008; Osterhout, 1997; Tanner et al., 2013; Tanner & Van Hell, 2014; Tanner, Inoue & Osterhout, 2014).

Tanner and Van Hell (2014) provided support for this view by demonstrating a strong *negative* correlation between the amplitudes of the N400 and P600 in their participants, further emphasizing that individuals who were likely to show one effect were not likely to show the other. However, these studies have treated the P600 as a monolithic component, typically quantified in a large time-window (500-1000 ms) or *only* in the first portion of the positivity (500-800 ms), in both cases immediately adjacent to the time-window of the negativity. As discussed above, it is of interest to explore different time-frames of the P600 elicited in response to agreement violations, and to study the relationship between negative/positive effects in *non-adjacent* time-windows where a group-level superimposition of the N400 and the subsequent positivity should no longer play a role (e.g., early negativity and *late* posterior positivity after 800 ms). If a biphasic pattern does exist at the individual level, and a negative correlation is still found between an early negativity and a late positivity, it may well be that the costliness of the re-analysis/repair process depends on how automatically the error was initially detected. If a morphosyntactic anomaly is detected at an early stage (i.e., larger LAN/N400), it may trigger less elaborated processing and repair at a later stage of processing (i.e., smaller P600). Alternatively, it may be that the two effects are *positively* correlated, such that the degree of ungrammaticality predicts the magnitude of both the error diagnosis and the re-analysis/repair.

Furthermore, in their discussions, Tanner, Osterhout and colleagues (Osterhout, 1997; Tanner et al., 2013; Tanner & Van Hell, 2014; Tanner, Inoue & Osterhout, 2014) have focused on the parietal P600 effect which is typically posterior and right-lateralized, and have not specifically discussed frontal positivities and their relationship with preceding negativities, either in terms of distributional overlap or in terms of individual differences in responses.

In sum, it is necessary to investigate how reliably and how independently the three effects reported in the morphosyntactic agreement literature (negativity, early frontal positivity, late posterior P600s) are elicited in response to number agreement violations, and how they may be correlated with one another, in order to achieve a better understanding of the processes involved in detecting and repairing agreement mismatches during online comprehension.

1.3. The present study

Our study was based on a previous experiment conducted with Italian monolinguals by Molinaro, Vespignani, Zamparelli and Job (2011; Experiment 1). Number agreement was manipulated between 3 sentence positions: (a) subject, (b) verb and (c) an adjective modifying the subject-noun (e.g., *I lavoratori tornano dalla fabbrica sporchi di grasso* / *The workers$^{(pl)}$ return$^{(pl)}$ from the factory dirty$^{(pl)}$ with grease*). Four experimental conditions, reflecting the four possible combinations of (dis-)agreement between the 3 sentence positions, were compared: (1) *Correct* ("xxx"); (2) *Inconsistent verb* ("xyx"); (3) *Inconsistent noun* ("xyy"); and (4) *Inconsistent modifier* ("xxy"). As was done in the Molinaro et al. study (but contrary to the majority of agreement studies), ERP correlates of morphosyntactic processing were examined on *two* target words: the *verb* and the *modifier.*

In order to facilitate readers' comprehension of Molinaro and colleagues' findings as well as our experimental predictions, our conditions and sentence examples are summarized in **Table 1**, which we refer back to in our Methods section.

Table 1 about here

The original study by Molinaro and colleagues only tested the 3 first conditions. On the verb, Italian native-speakers showed a LAN (350-450 ms) followed by an early (500-800 ms) and late (800-100 ms) P600 in response to subject-verb number mismatches. The LAN was only significant in the authors' direct comparison of condition pairs (*xyy* vs. *xxx* and *xyx* vs. *xxx*) rather than in a global ANOVA. Crucially, on the modifier, the researchers' aim was to test their "*Repair hypothesis*". According to this theory, the easiest and most cost-effective way for the cognitive system to process a morphosyntactic mismatch occurring early in a sentence is to repair this mismatch based on the constituent on which it is detected (in this case, the verb), and to pursue the repaired/grammatical interpretation for the remainder of the sentence. In support of this hypothesis, Molinaro and colleagues found that the modifier elicited a long-lasting P600 (with no preceding LAN) in response to *xyx* violations but not *xyy* violations. These results suggested that Italian native-speakers repaired the subject-verb number mismatch on the basis of the number of the verb before integrating the upcoming modifier, which was either congruent with the revised version of the sentence (*xyy*) or clashed with it (*xyx*). The number agreement mismatch between subject-noun and modifier (*xyy*) could no longer be detected, given that the internal representation of the subject had been revised to match the number of the verb[3]. Thus, rather than maintaining the opposing number values of the two mismatching constituents (subject and verb) in working memory until further incoming information disambiguated the sentence, native-speakers were found to immediately revise/repair the number mismatch and to project the number of the verb onto the rest of the sentence. For both the verb and the modifier, the P600 effect showed a different scalp distribution depending on the processing stage; in the earlier time window (500-800 ms), the positivity was larger at fronto-central than posterior sites, whereas the later P600 (800-1000 ms) was mainly posterior.

Based on these findings as well as on much of the number agreement literature, we expected Italian native-speakers to show typical morphosyntactic violation effects (LAN and/or early (frontal) P600 and/or late P600) when they reached the verb, for those conditions where the verb disagreed with the subject (*xyx* and *xyy*, but not *xxy*). On the modifier, we anticipated the largest violation effects for *xxy* where the modifier disagreed in number with both elements that preceded it, as the number of the subject would have been confirmed and reinforced by the verb. However, a strong version of the "Repair hypothesis" may instead predict that *xxy* should not differ from *xyx* – given the violation on the verb in the *xyx* condition, the theory would state that the internal representation of the subject-noun is repaired to match the number of the verb, and therefore *xyx* violations (having become *yyx* after repair) would be identical to *xxy* violations. Given that the original study did not include the *xxy* condition, this hypothesis was not tested with the present set of stimuli (but see English study by Molinaro, Kim & Vespignani, 2008). If we were to replicate Molinaro and colleagues' original findings, we should find that Italian native-speakers show morphosyntactic processing difficulties on the modifier for condition

[3] One might argue that, instead of being repaired, *xyy* sentences do not elicit morphosyntactic-violation effects because the number of the modifier is simply "checked" relative to the number of the most recent constituent (the verb) rather than to its antecedent (subject). However, Molinaro and colleagues (2011) ruled out this "Recency hypothesis" in Experiment 2 of the same paper with an additional condition that allowed them to tease apart the predictions of the two accounts.

xyx but not *xyy*, in spite of condition *xyy* being ungrammatical due to the mismatch between the modifier and its antecedent (subject-noun). Predictions for native Italian monolinguals are summarized in **Table 2**.

Table 2 about here

The novelty of the present study centers on our extension of this paradigm to the realm of bilingualism and L1 attrition. Not only is this one of the earliest ERP investigations of L1 attrition, but it is the first experiment to investigate online morphosyntactic processing in attriters at multiple points in a given sentence, in an attempt to determine whether attriters' L1 linguistic system detects and recovers from erroneous analyses in the same way as non-attriting Italian individuals (i.e., monolinguals and higher-proficiency attriters). Our goals were (1) to examine the potential breakdown in the error detection/diagnosis and strategies of online repair/re-analysis in attriters' L1 grammar, and (2) to determine whether "native-like" patterns in number agreement processing were predicted by L1 proficiency-level. We expected that, if long-distance agreement computation and online repair/re-analysis processes are vulnerable to attrition, group differences would be observed on both the verb but perhaps especially the modifier. In addition, we expected that Italian proficiency level would modulate ERP patterns in response to agreement violations in *both* groups of native-speakers, with lower-proficiency speakers exhibiting differences in the presence, amplitude, latency, duration, and/or scalp distribution of the components of interest[4].

2. METHODS

2.1. Participants

Two groups of native-Italian speakers were tested (see *General Methods*, p. 47-50 for demographic information).

[4] Our study also included a third participant group of high-proficiency late L2 learners of Italian, given the relevance of comparing attriters' to L2 learners to further explore questions of (1) neuroplasticity for language in adulthood; (2) proficiency effects on L1 and L2 processing; (3) cross-linguistic transfer and its relation to attrition. However, we decided to defer the discussion of the L2 data to a separate (forthcoming) paper, for two reasons. First, the primary goal of the current paper is to compare groups on their L1 (not the L2) and explore processing differences related to attrition and L1 proficiency level. Second, including a third group in a study with many statistical analyses on two target words would have rendered the Results and Discussion sections much too heavy and complex.

2.2. Behavioral measures

Participants completed background questionnaires and several behavioral measures (see *General Methods*, p. 50-54). Group means are provided in **Table 3**. Although Attriters scored numerically lower on all four proficiency measures, they did not differ significantly from Controls ($p > 0.1$). Subgroups of "*high*" and "*low*" proficiency were derived by median split (see *General Methods*, p. 54-56 for details). Note that, for the sake of brevity, we refer to individuals in the lower range as "low proficiency" individuals, but it is obvious that their proficiency level is not "low" in the conventional sense of the word.

Table 3 about here

High and *low* proficiency subgroups differed significantly on all measures except WM span, but especially on the written measures (**Table 4**). High and low proficiency Attriters differed significantly on all measures (including WM) whereas high and low proficiency Controls only differed on the two written proficiency tests. Numerically, low proficiency Attriters scored lower than low proficiency Controls on all proficiency measures, although this difference was not significant ($ps > 0.1$).

Table 4 about here

2.3. Stimuli

Sentence examples from all four experimental conditions are provided in Table 1. The experimental stimuli consisted of eight-word sentences containing two target words: (1) a lexical *verb* (in 3rd position), and (2) a *modifier* (in 6th position). Each sentence began with a masculine, animate subject (composed by a noun preceded by its correct definite article), and number was counterbalanced such that half of the subject-nouns were plural, and half singular. The modifier was separated from verb by two constituents – a function word and an inanimate noun. This time lag was necessary to provide enough time for a possible structure re-analysis and to allow for slow-going ERP waves such as the P600 elicited by the verb to return to baseline prior to the presentation of the modifier. The intervening noun was inanimate and feminine in gender to eliminate ambiguity that would lead readers to attach the modifier to the intervening noun rather than to the subject-noun. In cases where the verb was intransitive (52.5%), the intervening phrase was a prepositional phrase, while for transitive sentences (47.5%), the intervening words consisted of a noun phrase (determiner + direct object noun). Sentences always ended with a prepositional phrase, in order for sentence wrap-up effects not to be confounded with effects on the modifier. Each target word contributed to each condition, thus ruling out that effects were driven by contextual or lexical (frequency, length) differences between conditions. There were no repetitions of subject nouns, verbs or modifiers across items.

The stimuli were adapted from those used by Molinaro, Vespignani, Zamparelli and Job (2011; Exp.1). However, several modifications were made which resulted in the creation of a number of new sentences. First, we balanced singular and plural versions of each sentence whereas Molinaro and colleagues only presented participants with sentences where the subject-noun was plural[5], thus any singular forms of verbs and modifiers would automatically indicate a violation. In order to minimize the predictability of our agreement combinations and the possibility of identifying violations on a superficial level, we decided to counterbalance the number of the sentence subject. Second, we changed the tense of the verbs from the remote past ("*passato remoto*") to the present tense, as the remote past tense is subject to regional differences throughout Italy and is used somewhat infrequently in some regions. We additionally balanced transitivity, as the original set of stimuli contained an uneven proportion of intransitive (61%) and transitive (39%) sentence constructions. We also substantially reduced repetitions of non-target segments (such as the intervening and sentence-final phrases). Finally, we replaced several nouns, verbs and modifiers that exceeded 10 letters in length, in order to ensure that words (and their agreement inflections) could be read in full when presented in rapid serial visual mode (especially by late L2 learners), without saccadic artifacts. Two Italian native-speakers (including one of the authors of the original study) checked and modified the stimuli until all sentences were deemed unproblematic.

---

[5] Molinaro et al. 2011 used only plural forms in their study as their main goal was to compare the processing of morphosyntactic plurals and coordinate plurals (joined by "and")

A set of 120 different sentences were constructed and realized in each of the eight conditions (four main conditions x singular/plural). Eight experimental lists were created such that, across lists, each sentence contributed equally to each condition, while no sentence was repeated within any of the experimental lists. Each participant also saw 204 filler sentences, which were part of the larger study (testing Italian lexical-semantic processing and relative clause sentences) and will be reported in forthcoming papers (see Kasparian, Vespignani & Steinhauer, 2013a,b; 2014a,b). Out of the total of 324 pseudorandomized stimuli (120 experimental and 204 fillers) per participant, 146 sentences (approx. 45%) were acceptable (grammatically and semantically), while 178 were expected to receive a rating of 3 or lower on a five-point rating scale (approx. 55%).

2.4. Procedure

The experimental procedure unfolded as described in the *General Methods* (p. 56-57).

2.5. EEG recording and analysis

For details on EEG recording, see *General Methods* (p. 57). Offline, EEG recordings were re-referenced to the average activity of the two mastoids and filtered with a phase-true 0.3-40 Hz band-pass filter. A methodological note regards the choice of offline reference; following an observation made by Molinaro, Barber & Carreiras (2011) in a review paper on agreement processing that LAN effects were more robustly reported in studies using average mastoids as the offline reference rather than the left mastoid, we decided to empirically test this hypothesis in our own agreement data by running all our analyses with each reference choice (left mastoid or average mastoids). At least for the present data with these two groups, the pattern of results was identical whether we used the left mastoid or the average mastoids as our offline reference.

Trials containing artifacts due to blinks, eye-movements and excessive muscle activity were rejected prior to averaging, using a moving-window (400 ms) standard deviation of 30 microvolts. On average, participants contributed 27/30 artifact-free trials per condition (range: 54-100%), with no differences across conditions for either target word (ps > 0.1).

ERPs were analyzed separately on the verb (-200-1200 ms) and the modifier (-200-1600 ms), and were time-locked to the onset of each target word with a baseline correction from -200-

200 ms. Our original baseline of -200-0 ms was adjusted in order to compensate for early differences triggered by the subject-noun in some Attriters but also in L2 learners; given that our data from L2 learners will be reported in a separate paper but ultimately compared to current data from Controls and Attriters, maintaining a consistent baseline interval was considered important. Crucially, all ERP effects reported here were consistent with either baseline correction (see an example in Appendix a).

ERPs were quantified in time-windows corresponding to each component of interest, based on previous agreement studies and on visual inspection of the grand average data for each participant group. On the verb, the time-windows were: (1) 300-500 (LAN/N400); (2) 550-650 (early frontal positivity); (5) 650-1000 (P600); (6) 1000-1200 (late P600). On the modifier, slightly different time-windows were selected based on visual inspection, especially to ensure that the negativity and positivity did not overlap in a given time-window: (1) 300-500 (LAN/N400); (2) 500-600 (intermediate window); (3) 600-900 (P600); (4) 1000-1300 (late P600). Note that the previous study by Molinaro et al (2011) did not include a late P600 window extending beyond 1000 ms.

For the ERP analyses, the four (dis-)agreement conditions were collapsed into two factors: *Agreement 1* (= Ag1), describing (dis-)agreement between the *first* two sentence positions (i.e., subject-noun and verb) and *Agreement 2* (= Ag2), describing (dis-)agreement between the *last* two sentence positions (i.e., verb and modifier), each with two levels (*correct* and *violation*). Thus, conditions *xxx* and *xxy* were collapsed into *Ag1-correct* sentences, *xyx* and *xyy* were *Ag1-violation* sentences, *xxx* and *xyy* were *Ag2-correct* sentences, and finally *xyx* and *xxy* were collapsed into *Ag2-violation* sentences. Although, on the verb, only *Ag1* is meaningful (as the third target word has not yet been encountered), *Ag2* was included as a factor in the global ANOVA for the verb position in order to confirm that modulations in Ag2 had no effect. This also allowed us to conduct identical ANOVAs on both target words. On the modifier, when interactions in the global ANOVA motivated follow-up comparisons by condition pairs (e.g. *xxy* vs. *xxx*), the factor "*Condition*" was used to describe the contrast (e.g., 2 levels: *xxy*, *xxx*).

Repeated-measures ANOVAs were performed separately for 4 midline electrodes (Fz, Cz, Pz, Oz) and 12 lateral electrodes (F3/4, C3/4, P3/4 and F7/8, T3/4, T5/6). Global ANOVAs for the midline sites included within-subject factors *Ag1* (correct, violation), *Ag2* (correct, violation), *Ant-Post* (anterior, central, parietal, occipital), whereas lateral ANOVAs additionally included factors *Hemisphere* (left, right) and *Laterality* (lateral, medial). For all ANOVAs, *Group* (Controls, Attriters) and *Proficiency* (High, Low) were between-subjects factors. Where appropriate, Greenhouse-Geisser correction was applied to analyses with more than two levels (e.g., *Ant-Post*). In these cases, the corrected p values but original degrees of freedom are reported. As a default, reported analyses are restricted to the midline only, except in cases where the lateral ANOVAs revealed additional effects (e.g., LAN).

3. RESULTS

3.1. Acceptability judgments

Acceptability ratings (on a scale from 1-5) for each sentence condition are shown in **Fig 1a**. A repeated-measures ANOVA with within-subjects factor *Condition* (xxx, xyx, xyy, xxy) and between-subjects factor *Group* (Controls, Attriters) revealed a significant main effect of *Condition* ($F(3,156) = 146.99$, $p < 0.001$ after Greenhouse-Geisser correction), but no effects or interactions with *Group* (ps > 0.10). Follow-up analyses indicated that the correct condition *xxx* received a significantly higher rating than violation conditions *xyx* ($F(1,52) = 165.17$, $p < 0.01$), *xyy* (($F(1,52) = 143.51$, $p < 0.01$), and *xxy* (($F(1,52) = 162.54$, $p < 0.01$), but that the violation conditions did not differ significantly from each other ($p > 0.05$).

Figure 1a about here

Correlational analyses indicated that overall-proficiency scores were positively correlated with participants' ratings in response to correct control sentences only ($r = 0.350$, $p < 0.01$). There were no significant correlations between working-memory measures and acceptability ratings ($p > 0.1$).

3.2. Reaction times

Reaction times between the onset of the prompt and participants' button-press are depicted in **Fig 1b**. A repeated-measures ANOVA with within-subjects factor *Condition* (xxx, xyx, xyy, xxy) and between-subjects factor *Group* (Controls, Attriters) revealed a significant main effect of *Condition* ($F(3,156) = 6.958$, $p < 0.01$), as well as a significant main effect of *Group* ($F(3,156) = 6.263$, $p < 0.05$), but no significant interaction between *Condition* x *Group* ($p > 0.1$), indicating that Attriters took longer to respond overall than the Controls. Follow-up analyses indicated that response times to the correct *xxx* condition were significantly longer than response times for *xyx* violations ($F(1,52) = 10.03$, $p < 0.01$) and *xyy* violations ($F(1,52) = 10.03$, $p < 0.05$) but not significantly different from response times for *xxy* violations ($p > 0.1$), suggesting that participants took longer to respond when the first two constituents agreed in number, compared to sentences where a violation was already present on the verb. The *xxy* condition only marginally differed from *xyx* ($F(1,52) = 3.36$, $p = 0.07$) and *xyy* violations ($F(1,52) = 3.47$, $p = 0.07$).

Figure 1b about here

Correlational analyses indicated that reading fluency scores were positively correlated with participants' reaction time in response to correct control sentences ($r = 0.306$, $p < 0.01$), but not to violation sentences. There were no significant correlations between reaction times and proficiency or working-memory measures ($p > 0.1$).

3.3. ERPs elicited at the Verb position

Grand average ERP waveforms for *Ag1* conditions time-locked to the verb are presented in **Fig. 2a** (Controls) and **Fig. 2b** (Attriters). In Controls, *Ag1* (subject-verb) violations elicited a small left-temporal negativity localized primarily at T5 between 300-500 ms (**Fig. 3a**), followed by a frontal positivity between 550-650 ms and a large posterior P600 lasting until 1200 ms. Attriters showed a prominent negativity between 300-500 ms which was primarily distributed over left and midline sites (**Fig. 3b**), followed by a frontal positivity between 550-650 ms, and a large P600 which appeared to be shorter in duration (lasting until 1000 ms) and less focal (less posterior) than in Controls.

Figures 2a and 2b about here

Figure 3a and 3b about here

Visual inspection of ERP patterns by proficiency level (**Fig. 4**) suggests that the amplitude of the negativity (300-500 ms) and of the P600 (650-1000) was largest in individuals with higher Italian proficiency. Moreover, the difference waves illustrated in **Fig. 5** suggest that the amplitude of the P600 in the standard time-window between 650-1000 ms was modulated by proficiency (larger P600 in high-proficiency), whereas the P600 amplitude in the late time-window between 1000-1200 ms depended on group membership rather than proficiency (present in Controls but absent in Attriters).

Figure 4 about here

Figure 5 about here

3.3.1. Negativity between 300-500 ms

The global ANOVA in the 300-500 ms time-window for midline electrodes revealed a significant main effect of *Ag1* ($F(1,50) = 16.63$, $p < 0.001$) and a significant *Ag1 x Group* interaction ($F(1,50) = 5.27$, $p < 0.05$). No interactions with *Ant-Post* or *Proficiency* reached significance (*p*s > 0.1). Follow-ups by *Group* demonstrated a significant main effect of *Ag1* in Attriters ($F(1,23) = 18.54$, $p < 0.001$) but not in Controls ($p > 0.1$), confirming that only Attriters elicited a broadly-distributed negativity on the midline when processing a verb that mismatched in number with the preceding subject-noun.

The lateral global ANOVA revealed a significant main effect of *Ag1* ($F(1,50) = 15.31$, $p < 0.0001$) and a significant *Ag1* x *Hemi* interaction ($F(1,50) = 7.95$, $p < 0.01$), reflecting a stronger negativity over left sites ($F(1,50) = 21.38$, $p < 0.0001$) rather than right ($F(1,50) = 5.83$, $p < 0.05$). Unlike at the midline, the interaction between *Ag1* x *Group* was only marginal ($F(1,50) = 2.89$, $p = 0.09$), as was the interaction between *Ag1 x Group x Laterality* ($F(1,50) = 3.45$, $p = 0.07$). However, follow-up analyses by *Laterality* supported the group differences observed during visual inspection: at medial sites, a significant main effect of *Ag1* ($F(1,50) = 15.93$, $p < 0.001$) was qualified by a significant *Ag1 x Group* interaction ($F(1,50) = 5.12$, $p < 0.05$), where Attriters elicited a significant negativity (Ag1: $F(1,23) = 14.78$, $p < 0.001$) but Controls did not ($p > 0.1$). At lateral sites, however, the ANOVA revealed only a significant main effect of *Ag1* ($F(1,50) = 10.96$, $p < 0.01$) and no significant interaction with *Group* ($p > 0.1$), indicating that the negativity was shared by both groups at lateral electrodes. No significant differences were found between *Proficiency* subgroups (ps > 0.1)

The amplitude of the negativity was not correlated with proficiency or working memory measures, nor with participants' acceptability judgment ratings in response to *Ag1* violations ($p > 0.1$).

3.3.2. Frontal positivity between 550-650 ms

Between 550-650 ms, the positivity elicited by *Ag1* violation sentences relative to correct sentences reached statistical significance across groups (*Ag1*: $F(1,50) = 12.31$, $p < 0.001$) and was qualified by a significant *Ag1 x Ant-Post* interaction ($F(3,150) = 7.34$, $p < 0.005$), reflecting that the positivity was most robust at Fz ($F(1,50) = 19.24$, $p < 0.0001$) than at more posterior electrodes (Cz: $p < 0.005$; Pz: $p < 0.01$; Oz: $p > 0.1$). The frontal positivity did not statistically differ by *Group* or *Proficiency* (ps > 0.1).

Correlational analyses did not yield any significant correlations between the amplitude of the frontal positivity and any behavioral measures (ps > 0.1).

3.3.3. P600 between 650-1000 ms

A highly significant main effect of *Ag1* ($F(1,50) = 36.26$, $p < 0.0001$) was qualified by a significant interaction with factor *Ant-Post* ($F(3,150) = 17.56$, $p < 0.0001$), reflecting the posterior distribution of the P600 (Fz: $p = 0.05$; Cz: $p < 0.0001$; Pz: $p < 0.0001$; Oz: $p < 0.0001$). Surprisingly, the expected interaction with factor *Proficiency* did not reach significance (ps > 0.1) in the global ANOVA (but see correlational analyses below). The lack of a significant *Group* interaction points to a P600 effect of similar amplitude and scalp distribution for Controls and Attriters in this time-window.

Correlations revealed that P600 amplitude was positively correlated with overall-proficiency scores (at Pz: $r = 0.274$, $p < 0.01$), as well as with individual proficiency measures such as the C-test (at Pz: $r = 0.320$, $p < 0.05$) and verbal semantic fluency (at Pz: $r = 0.345$, $p < 0.01$), such that individuals with higher Italian proficiency scores elicited a larger P600 effect in response to subject-verb number agreement mismatches. There was no significant correlation between end-of-sentence acceptability ratings and the P600 effect on the verb (ps > 0.1).

3.3.4. Late P600 between 1000-1200 ms

In this late interval, *Ag1* violations elicited a posterior P600 (*Ag1*: $F(1,50) = 27.44$, $p < 0.0001$; *Ag1 x Ant-Post*: $F(3,150) = 26.63$, $p < 0.0001$) compared to correct sentences. A significant interaction with factor *Group* ($F(1,50) = 10.33$, $p < 0.005$) indicated that this effect was present in the Controls ($F(1,29) = 21.79$, $p < 0.0001$) but not the Attriters ($p > 0.1$). *Proficiency* did not prove to be a meaningful factor in this late P600 time-window (ps > 0.1).

Correlational analyses confirmed that proficiency scores did not modulate the P600 in this late interval (ps > 0.1), contrary to the earlier P600 window between 650-1000 ms.

3.4. **ERPs elicited at the Modifier position**

On the modifier, condition *xxy* was expected to elicit a large negativity followed by a large P600, as it constituted the most salient number-agreement violation out of the four conditions. In order to replicate Molinaro and colleagues' reports that subject-noun number mismatches are repaired on the basis of verb number before the modifier is subsequently integrated into the sentence, we would expect that condition *xyy* would not differ from condition *xxx* on the modifier, but that condition *xyx* would elicit violation effects. Finally, we expected ERP patterns to be affected by group membership and/or by Italian proficiency.

ERP waveforms for the *xxy* condition versus *xxx* for each group (**Fig. 6**) illustrate that Controls and Attriters show a similar pattern, namely a large, broadly-distributed N400-like negativity (most prominent at medial electrodes) followed by a large parietal P600. In the attriters, the P600 is also present at frontal sites, while no frontally-distributed positivity is discernible in Controls. The scalp distribution and duration of the negativity seem to be influenced by proficiency (**Fig. 7**), with lower proficiency individuals eliciting a less left-lateralized and longer-lasting negativity (until 600 ms).

Figure 6 about here

Figure 7 about here

The P600 appears to be differentially modulated both proficiency level (600-900 ms) and by group (1000-1300 ms), with lower proficiency individuals eliciting a less focal P600 of smaller amplitude in the earlier time-window, and Attriters eliciting a shorter-lived P600 than Controls. This differential proficiency/group effect on different slices of the P600 is further illustrated with difference waves in **Fig. 8** and is reminiscent of the pattern we had observed on the verb (**Fig. 5**).

Figure 8 about here

Comparisons of violation conditions *xyx* and *xyy* with the correct condition *xxx* (**Fig. 9)** show that, in Controls, the *xyx* condition elicits a P600 starting around 650 ms, whereas the *xyy* condition largely overlaps with the correct control condition in the P600 interval. In contrast, Attriters show a clear P600 effect for both violation conditions, relative to correct sentences.

Figure 9 about here

3.4.1. Negativity (LAN/N400) between 300-500 ms

The global ANOVA in the 300-500 ms time-window for midline electrodes revealed a significant main effect of *Ag2* (i.e., agreement between verb and modifier; $F(1,50) = 11.92$, $p < 0.001$) as well as a main effect of *Ag1* (i.e., agreement between subject and verb; $F(1,50) = 6.14$, $p < 0.05$), which were qualified by a significant interaction between *Ag1 x Ag2* ($F(1,50) = 17.57$, $p < 0.0001$). *Ag1* x *Ant-Post* was also significant ($F(3,150) = 4.74$, $p < 0.005$), as was *Ag2* x *Ant-Post* x *Proficiency* ($F(3,150) = 2.68$, $p < 0.05$) and *Ag2* x *Ag1* x *Ant-Post* ($F(3,150) = 3.80$, $p < 0.05$). Interactions with *Group* did not reach significance, indicating that the negativity was shared between Controls and Attriters. Given that our predictions as well as visual inspection of the data pointed to the negativity being primarily driven by the *xxy* condition, we proceeded directly to investigating follow-up analyses by condition pairs (which was motivated by significant *Ag2* x *Ag1* interactions). The only comparison that revealed significant effects in the negativity time-window was *xxy* vs. *xxx. Condition* was found to be highly significant ($F(1,50) = 25.31$, $p < 0.0001$), as was *Condition* x *Ant-Post* ($F(1,50) = 4.58$, $p < 0.005$). These effects were qualified by a significant *Condition* x *Ant-Post* x *Proficiency* interaction ($F(3,150) = 3.14$, $p < 0.05$). Follow-up analyses by *Proficiency* indicated that the negativity was present in both subgroups (High: *Cond*: $F(1,26) = 10.01$, $p < 0.005$; *Cond x Ant-Post*: $F(3,78) = 3.30$, $p < 0.05$; Low: *Cond*: $F(1,26) = 15.67$, $p < 0.001$; *Cond x Ant-Post*: $F(3,78) = 4.84$, $p < 0.001$) but that it differed in its scalp distribution. While the effect was frontally-predominant in the higher proficiency subgroup (Fz: $p < 0.05$; Cz: $p = 0.05$; Pz and Oz: $p > 0.1$), it was predominant at Pz in lower proficiency individuals (Fz: $p > 0.05$; Cz: $p < 0.05$; Pz: $p < 0.001$; Oz: $p < 0.01$).

Correlational analyses supported the ANOVA results that that the scalp distribution of the negativity in the *xxy* condition was influenced by Italian proficiency. Participants' C-test scores positively (rather than negatively) correlated with *xxy* vs. *xxx* amplitude at Pz ($r = 0.293$, $p < 0.01$), indicating that participants with higher scores were less likely to exhibit a significant negativity at Pz. In contrast, proficiency was negatively correlated with the effect at Fz (Error-detection test: $r = -0.210$, $p < 0.05$), demonstrating that the negativity was more enhanced at frontal sites in higher-proficiency individuals. There were no significant correlations between the negativity elicited by the modifier and end-of-sentence acceptability ratings or reaction times (*p*s > 0.1).

3.4.2. Intermediate time-window between 500-600 ms

This intermediate time-window was selected to corroborate the grand-average data depicting an ongoing negativity only for lower-proficiency individuals. The midline ANOVA revealed a significant main effect of *Ag1* ($F(1,50) = 5.61$, $p < 0.05$) but not of *Ag2* ($p > 0.1$), as well as a significant interaction between *Ag2 x Ag1* ($F(1,50) = 7.38$, $p < 0.01$) and *Ag1 x Ant-Post* ($F(3,150) = 5.26$, $p < 0.01$). The interaction between factors *Ag2* x *Ant-Post* x *Proficiency* reached significance ($F(3,150) = 3.57$, $p < 0.05$), while the interaction between *Ag2* x *Proficiency* was marginal ($F(1,50) = 4.04$, $p = 0.05$).

Follow-up analyses by *Condition* pairs were then performed, motivated by the significant *Ag2* x *Ag1* interaction. As expected based on visual inspection of the data, none of the condition pairs revealed any significant effects except *xxy* vs. *xxx*, where the expected Proficiency interaction was found (*Condition*: $F(1,50) = 5.33$, $p < 0.05$; *Condition* x *Proficiency*: $F(1,50) = 3.15$, $p = 0.08$; *Condition* x *Ant-Post* x *Proficiency*: $F(3,150) = 2.28$, $p = 0.08$). The interaction with *Ant-Post* and *Proficiency* reflected High and Low proficiency subgroups' shared effect at frontal electrodes, while at central-posterior electrodes, only the low proficiency subgroup showed a negativity. Despite marginal interactions with factor *Proficiency*, follow-up analyses within each proficiency subgroup clearly supported the trend seen in the data: the negativity persisted from 500-600 ms for lower-proficiency (Condition: $F(1,26) = 15.47$, $p < 0.001$) but not higher-proficiency individuals ($p > 0.1$).

Correlational analyses confirmed this trend and revealed that individuals with higher proficiency scores, especially on the C-test, elicited an effect with a positive amplitude for *xxy* vs. *xxx* in the 500-600 ms range (C-test and amplitude at Cz: $r = 0.270$, $p < 0.05$; C-test and amplitude at Pz: r 0.337, $p < 0.01$; Error-test and amplitude at Pz: $r = 0.274$, $p < 0.05$; Overall proficiency and amplitude at Pz: $r = 0.190$; $p = 0.08$). Thus, the negativity in this time-window persisted only for lower-proficiency individuals.

3.4.3. P600 between 600-900 ms

On the midline in the prototypical P600 window, a significant main effect of *Ag1* ($F(1,50) = 10.95$, $p < 0.005$) and of *Ag2* ($F(1,50) = 30.86$, $p < 0.0001$) were qualified by a significant interaction between the two factors ($F(1,50) = 11.72$, $p < 0.005$). Interactions with *Ant-Post* were also significant (*Ag2* x *Ant-Post*: $F(3,150) = 11.90$, $p < 0.0001$; *Ag2* x *Ag1* x *Ant-Post*: $F(3,150) = 12.08$, $p < 0.0001$), reflecting the posterior prominence of the positivity. *Ag2* significantly interacted with *Proficiency* (*Ag2* x *Proficiency*: $F(1,50) = 5.77$, $p < 0.05$; *Ag2* x *Ant-Post* x *Proficiency*: $F(3,150) = 5.12$, $p < 0.05$), but the interaction between *Ag2* x *Ag1* x *Ant-Post* x *Proficiency* was even more significant ($F(3,150) = 5.34$, $p < 0.005$; *Ag2* x *Ag1* x *Proficiency* was marginal ($F(1,50) = 3.81$, $p = 0.06$)). No interactions with factor *Group* were statistically significant (*p*s > 0.1), suggesting that the P600 in this time-window was modulated by Italian proficiency level irrespective of group membership.

Follow-up analyses by *Condition* (motivated by the Ag2 x Ag1 interactions) revealed the most significant difference to be between *xxy* and *xxx* conditions (*Condition*: $F(1,50) = 29.77$, $p < 0.0001$; *Condition* x *Ant-Post*: $F(3,150) = 16.01$, $p < 0.0001$). Interactions with proficiency also reached significance (*Condition* x *Proficiency*: $F(1,50) = 6.98$, $p < 0.01$; *Condition* x *Ant-Post* x *Proficiency*: $F(3,150) = 7.10$, $p < 0.0005$). ANOVAs within each proficiency subgroup indicated a significant P600 effect only in the higher proficiency individuals (*Condition*: $F(1,26) = 35.67$, $p < 0.0001$; *Condition* x *Ant-Post*: $F(3,78) = 16.20$, $p < 0.0001$; significant at Cz, Pz, Oz at $p < 0.0001$). In lower proficiency individuals, the P600 effect was only marginally significant (*Condition*: $F(1,26) = 3.39$, $p = 0.08$; *Condition* x *Ant-Post*: $F(3,78) = 2.28$, $p = 0.08$, only at *Pz*: $p < 0.05$). Correlational analyses provided further support that a higher level of overall Italian proficiency was associated with a larger P600 amplitude (at Pz: r = 0.349, p < 0.01). Individual proficiency tests were also correlated with P600 amplitude at Pz (<u>C-test</u>: $r = 0.381$, $p < 0.01$; <u>Error-test</u>: $r = 0.358$, $p < 0.01$; <u>Semantic fluency</u>: $r = 0.239$, $p < 0.05$).

The comparison between *xyx* and *xxx* conditions yielded a significant main effect of *Condition* ($F(1,50) = 5.19$, $p < 0.05$) and an interaction with factor *Ant-Post* ($F(3,150) = 5.03$, $p < 0.05$) as well as a marginal three-way interaction between *Condition* x *Ant-Post* x *Proficiency* ($F(3,150) = 2.47$, $p = 0.06$). Despite the marginal significance, follow-up analyses within each proficiency subgroup supported the trend, such that only higher proficiency individuals showed a significant P600 for the *xyx* condition (*Condition*: $F(1,26) = 4.54$, $p < 0.05$; *Condition* x *Ant-*

Post: $F(3,78) = 6.09$, $p < 0.01$) which was significant at Pz ($p < 0.005$) and Oz ($p < 0.05$). In lower proficiency individuals, no effects approached significance (*p*s > 0.10).

The comparison between *xyy* and *xxx* also revealed a significant *Condition* x *Ant-Post* interaction ($F(3,150) = 4.49$, $p < 0.05$), reflecting the posterior distribution of the P600. Contrary to the repair hypothesis, the *xyy* condition did not overlap with the correct control condition. Surprisingly, the difference between the two conditions was neither modulated by *Group* nor *Proficiency* (*p*s > 0.1) in this P600 time-window. No correlations between P600 amplitude and behavioral measures reached significance for condition *xyy* vs. *xxx* (*p*s > 0.1)

Finally, comparing *xyy* and *xyx* violation conditions revealed a marginal main effect of *Condition* ($F(1,50) = 3.88$, $p = 0.06$) but no interactions with *Ant-Post* nor with between-subject factors such as *Group* nor *Proficiency* (*p*s > 0.1).

3.4.4. Late P600 between 1000-1300 ms

The global ANOVA on the midline confirmed the pattern observed in the data (**Fig. 9**), namely that *Group* (but not *Proficiency*) was the meaningful factor that modulated P600 effects in this very late time-window. The interaction between *Ag2* x *Group* was significant ($F(1,50) = 5.91$, $p < 0.01$), as were the interactions with factor *Ant-Post* (*Ag2* x *Ant-Post*: $F(3,150) = 13.31$, $p < 0.0001$; *Ag2* x *Ag1* x *Ant-Post*: $F(3,150) = 13.95$, $p < 0.0001$), as the effect was visibly strongest at posterior electrodes.

Group interactions (but not *Proficiency* interactions) were also found in follow-up analyses performed by *Condition* pairs. The comparison between *xxy* vs. *xxx* yielded a significant *Condition* x *Group* interaction ($F(1,50) = 8.75$, $p < 0.005$), which, when followed-up within each group, revealed that Controls showed a significant P600 effect in response to *xxy* violations ($F(1,29) = 8.39$, $p < 0.01$) but Attriters did not ($p > 0.1$). In the comparison between condition *xyx* vs. *xxx,* the *Condition* x *Group* interaction also reached significance ($F(1,50) = 5.21$, $p < 0.05$), once again reflecting the presence of the P600 effect in Controls ($F(1,29) = 6.93$, $p < 0.05$) but not in Attriters (*p*s > 0.1). Contrasting conditions *xyy* vs. *xxx* yielded a significant *Condition* x *Ant-Post* interaction ($F(3,150) = 3.19$, $p < 0.05$), but violation conditions *xyy* vs. *xyx* were not statistically different from one another in this late P600 window (*p*s > 0.1).

3.5. Experiential factors and ERP patterns in Attriters

We also assessed the role of background factors (such as age, education, age at immigration, and length of residence) as well as factors related to language use (such as amount of L1 and L2 exposure) on proficiency scores and ERP patterns.

Length of residence was found to negatively correlate with scores on the written Error-detection test only ($r = -0.49$, $p < 0.01$). Amount of L1 exposure (in terms of hours/week) was positively correlated with Attriters' overall proficiency scores ($r = 0.43$, $p < 0.01$) as well as with their performance on the semantic fluency task ($r = 0.35$, $p < 0.05$).

With respect to ERP patterns, amount of daily L1 exposure (% relative to L2 exposure) was positively correlated with the late P600 elicited by the modifier in the *xxy* conditions, both in the 900-1000 ms time-window ($r = 0.40$, $p < 0.005$) as well as the 1000-1300 ms window ($r = 0.45$, $p < 0.000$). Thus attriters with more L1 exposure were more similar to native-controls in showing a late P600 effect. As expected given the relationship between L1 and L2 exposure, amount of L2 exposure was *negatively* correlated with the late P600 for the same two late time-windows (900-1000 ms: $r = -0.35$, $p < 0.005$; 1000-1300 ms: $r = -0.36$, $p < 0.005$). There was no relationship between ERP responses to *xyy* or *xyx* violations and experiential factors (ps > 0.1). In line with our predictions that attrition effects might be more visible on the modifier than on the verb, no significant correlations were found between Attriters' experiential characteristics and ERP responses elicited by the verb (ps > 0.1).

Finally, no correlations were found between ERP patterns and attriters' age, education level, length of residence or age at immigration (ps > 0.1).

4. DISCUSSION

Our study compared Attriters and non-attriting native-speakers in their online detection and resolution of number agreement violations in their native-L1 (Italian), with two main goals in mind: (1) to determine whether long-distance agreement computation and online error detection/re-analysis mechanisms are vulnerable to attrition, behaviorally and/or at the neurocognitive level, and (2) to explore whether processing differences are driven by proficiency level, even in two groups of *native* Italian speakers.

4.1. Main findings on the verb

4.1.2. Negativity: LAN, N400 or left-temporal negativity?

On the verb, we showed that Attriters and Controls differed in the early negativity (300-500 ms) elicited by subject-verb number mismatches, both in terms of its amplitude and scalp distribution. While Attriters showed a robust, broadly distributed negativity that extended from midline to lateral sites (though larger over the left-hemisphere), Controls showed a much weaker negativity which was focused at left-temporal sites (T3, T5), and thus only reached significance in the global ANOVA that included the most lateral electrodes. Therefore, although the negativity in both groups was more prominent over the left-hemisphere than the right-hemisphere, the effect was more consistent with a left-temporal negativity (LTN) in Controls, but a left-lateralized N400 in Attriters. LTNs have previously been reported to occur in response to morphosyntactic violations instead of a LAN in reading studies (see Steinhauer et al., 2009; 2010; Neville et al., 1991; Newman et al., 2007; Weber-Fox & Neville, 1996).

One possible view is that the negativity was shared by both groups but that the distributional differences were merely caused by overlap with the subsequent positivity. According to Tanner and colleagues (Tanner et al., 2013; Tanner & Van Hell, 2014; Tanner, Inoue & Osterhout, 2014), extending previous suggestions by Osterhout and colleagues (Osterhout, 1997; Osterhout & Mobley, 1995), left-lateralized negativities are N400s modulated by P600 components. In line with this argument, a number of previous studies have demonstrated that the topography and the duration of negativities preceding the P600 can be substantially altered by component overlap, both in reading (e.g., Tanner & Van Hell, 2014) as well as in the auditory domain (Steinhauer & Drury, 2012). When applied to our data, the assumption would be that both native-Controls and Attriters in our data show an N400 effect in response to subject-verb agreement violations, but the scalp distribution of the N400 is affected by the onset of the following positivity, which cancels out the negativity at sites where both effects overlap in time. However, we question the ability of this account to explain our data, given that the frontal positivity (550-650 ms) was shared across both groups, and that neither its amplitude nor scalp distribution differed significantly between Controls and Attriters. In fact, the frontal positivity was *numerically larger* in the Attriters (**Fig 2b**), i.e. in the group with the larger negativity. It is not conceivable that a larger negativity would survive when followed by a numerically larger positivity in an adjacent time-window. Nonetheless, even if component overlap were to explain the resulting difference in scalp distribution of an otherwise shared N400 across groups, Controls and Attriters would have to have differed on the frontal positivity. Thus, our data would have still revealed a group difference.

In terms of the functional significance of the negativity, we initially expected subject-verb number agreement violations to elicit a LAN in native-Controls, in accordance with much of the cross-linguistic literature on morphosyntactic agreement (Barber & Carreiras, 2005; Carreiras, Salillas & Barber, 2004; Hagoort & Brown, 2000; Osterhout & Mobley, 1995; Mancini et al., 2009; Molinaro et al., 2008; 2011; Silva-Pereyra & Carreiras, 2007) and compatible with traditional accounts of LAN-like negativities reflecting the (quasi-automatic) detection of morphosyntactic violations (e.g., Friederici, 2002). A recent review of ERPs elicited by morphosyntactic agreement violations (Molinaro, Barber & Carreiras, 2011) found LANs to be the most reliable kind of negativity in response to number agreement violations (in contrast to gender or person agreement violations), especially when both subject noun and verb were morphologically marked for number. The authors argued that the morphosyntactic marker on the "trigger element" (subject) results in predictions regarding the number-marking on the "target element" (verb), which – when violated – elicit a LAN.

As discussed in introduction, however, a "typical LAN" has not been consistently reported across *all* studies (e.g., Balconi & Pozzoli, 2005; Hagoort & Brown, 2000; Hagoort et al., 1993; Nevins, Dillen, Malhotra & Phillips 2007, Osterhout et al., 1996). We also considered the possibility raised in Molinaro and colleagues' review paper that the lack of a robust LAN may be due to the choice of offline reference but, as discussed in the Methods section, neither averaging to the left mastoid (A1) nor to the average of the two mastoids (A1, A2; our reported results) affected the prominence of the LAN in our data (also see Tanner, in press in Cortex). The absence of a strong LAN in monolingual Italian native-speakers (in contrast to English studies) is not an entirely surprising finding, given that several Italian studies have previously struggled to detect a significant LAN in overall statistical analyses (Angrilli et al., 2002; De Vincenzi et al., 2003; Mancini et al., 2009; Molinaro et al., 2011). It has been argued that, due to the relatively free word-order of Italian and the grammaticality of post-verbal subject constructions, the expectation of agreement between a sentence-initial noun and a subsequent verb is weaker in Italian than in languages such as English, where post-verbal subjects are never acceptable and the verb must therefore agree with its preceding noun (Molinaro, Barber & Carreiras, 2011).

Given that it is syntactically acceptable for the subject to follow the verb in Italian, a mismatch in number agreement detected on the verb does not necessarily signal a grammatical violation and thus may fail to elicit a robust LAN compared to, for example, determiner-noun number agreement violations (as shown in Vespignani, Molinaro & Job, in preparation).

If a biphasic LAN + P600 pattern is more readily reported in English than in Italian morphosyntactic processing, one might expect Attriters to show the English processing pattern if they are affected by their English dominance (i.e., high degree of English exposure/use). Attriters' reliance on word-order cues (subject precedes verb) rather than entertaining the possibility of a post-verbal subject-noun may have led them to be more sensitive to number agreement mismatches on the verb, and to elicit a stronger negativity than non-attriting native-speakers. It follows that, possibly due to interference from English and/or to less elaborated processing routines, Attriters show a reduced capacity to immediately explore all possible structural solutions to the agreement problem (e.g., the possibility of a post-verbal subject in Italian).

The fact that the left-lateralized negativity in Attriters can be interpreted as either an N400 or a LAN may be in line with arguments by Osterhout, Tanner and colleagues that all negativities are in fact N400s. Interestingly, in terms of its functional significance, Tanner and colleagues provide a similar account for the N400 as Molinaro, Barber & Carreiras (2011) had given for the LAN, namely the violation of "word- or morphological-form-based predictions" (p. 298). A plausible assumption in our study would be that participants had strong expectancies that the verb would have the same number as the subject-noun phrase, either morphosyntactically or conceptually, or both. The N400 would therefore reflect a mismatch of this expectation. To some extent, this interpretation is similar to the traditional LAN interpretation as a component reflecting an early morphosyntactic mismatch. However, it adds a possible conceptual level (i.e., is the sentence about *one* worker or *multiple* workers?). The N400 in our data may reflect Attriters' accessing of lexical-semantic or discourse-level information when processing the number mismatch on the verb, similar to what has been reported in processing gender and person agreement errors (see Molinaro, Barber & Carreiras, 2011 for a review).

4.1.3. Early frontal and late posterior positivities

The second ERP response elicited by the verb when it mismatched in number with the preceding subject-noun was a frontal positivity (550-650 ms) similar to the effect reported in the previous study by Molinaro and colleagues (2011). As discussed above, the amplitude and scalp distribution of the frontal positivity was not statistically different between Controls and Attriters. The functional significance of the frontal positivity will be further discussed in Section 4.3 below.

As expected, subject-verb number mismatches also elicited a large posterior P600 (as of 650 ms) relative to correct control sentences. This posterior positivity was divided into two distinct phases – a first phase (650-900 ms), shared by Controls and Attriters but modulated by L1 proficiency (larger P600 in higher-proficiency individuals), and a second phase (1000-1200 ms) where only the Controls showed a late, ongoing P600, whereas the P600 in Attriters returned to baseline by 1000 ms. This late portion of the posterior P600 was not modulated by Italian proficiency level. As later stages of the P600 have been associated with re-analysis and repair processes, we interpret these results as suggesting that Controls engage in more extensive/elaborated repair than Attriters do for the same number-agreement violations (Carreiras, Salillas & Barber, 2004; Hagoort & Brown, 2000; Mancini, Vespignani, Molinaro, Laudanna & Rizzi, 2009; Molinaro, Vespignani et al., 2008; Silva-Pereyra & Carreiras, 2007). Interestingly, this two-stage posterior P600 pattern and the differential effect of proficiency vs. group on its amplitude was strikingly similar to the ERP pattern elicited by the modifier (in response to *xxy* violations).

In examining the different time "slices" of P600 effects elicited in response to number agreement violations, we also sought to explore the relationship between these positivities and their preceding early-negativities. On the verb, we found a significant negative correlation between the LTN/N400 from 300-500 ms and the frontal positivity between 550-650 ms, similar to what has been reported by Tanner and colleagues, although they have not explicitly investigated *frontal* positivities ($r = 0.337$, $p < 0.01$). However, there was no correlation between the N400 and the prototypical P600 elicited between 650-1000 ($p > 0.1$). Interestingly, though, we discovered a significant negative correlation between the LTN/N400 and the late P600 from 1000-1200 ms ($r = 0.315$, $p < 0.05$), indicating that larger negativities in the 300-500 ms range were associated with a smaller positivity in the very late P600 time-window between 1000-1200 ms. Given the absence of a negative correlation between the negativity and the prototypical parietal P600 in the 650-1000 ms time-window, we argue that our ERP results on the verb are not entirely in line with Tanner and colleagues' predictions.

In sum, ERP analyses on the first target word in the sentence (verb) revealed that Attriters and Controls differed in the amplitude and scalp distribution of the early negativity (300-500 ms) and the later stage of the posterior P600 (1000-1200 ms) elicited in response to subject-verb number violations. In contrast, the two groups shared the frontal positivity (550-650 ms) which followed the negativity, as well as the early phase of the posterior P600 (650-1000 ms), although the latter effect was modulated by L1-proficiency level, with higher Italian proficiency scores associated with larger P600 effects.

4.2. Main findings on the modifier

On the modifier, we sought to examine how number agreement violations detected on the verb may affect the online integration of subsequent information. The processing of the modifier was expected to highlight differences in the repair / re-analysis strategies that Controls and Attriters engage in during online comprehension.

4.2.1. A biphasic pattern in response to xxy violations

Condition *xxy* was a condition that the previous Italian study by Molinaro and colleagues (2011) had not tested with similar stimuli. Given that, in these sentences, the modifier marks the first point of violation after both the subject-noun and the verb emphasize the (consistent) grammatical number value of the sentence, we expected *xxy* violations to elicit a large negativity, followed by a large P600. In contrast, for the other two violation conditions xyx and xyy, we expected the morphosyntactic violation effects to be smaller than in condition xxy. In order to replicate Molinaro and colleagues' (2011) "*Repair hypothesis*" which argues that subject-noun number mismatches are repaired on the basis of verb number before the modifier is integrated into the sentence, we expected that condition *xyy* would not differ from the correct control condition (*xxx*) on the modifier, but that condition *xyx* would elicit a P600. Finally, we expected ERP patterns to be affected by group membership and/or by L1-proficiency.

In response to *xxy* violations, we showed that Attriters and Controls elicited a similar N400 effect (300-500 ms) followed by a P600 (600-900 ms) *without* a preceding frontal positivity. However, the groups differed once again on the duration of the P600, which persisted into the 1000-1300 ms time-window in Controls but not in Attriters. Moreover, L1 proficiency level modulated ERP responses to *xxy* violations relative to the correct *xxx* condition – individuals with higher Italian proficiency scores showed an anterior negativity followed by a larger P600, whereas individuals with lower L1 proficiency scores showed a smaller, more posterior and longer-lasting (until 600 ms) negativity, followed by a P600 of a smaller amplitude. Thus, the negativity and the P600 elicited in response to *xxy* violations were modulated by L1 proficiency level, in terms of the amplitude, scalp distribution as well as the latency of the effects. Crucially and similar to the pattern observed on the verb, proficiency only modulated the P600 amplitude in the time-window between 650-1000 ms, whereas the P600 effect extending beyond 1000 ms was dependent on group rather than on proficiency level, with the P600 persisting only in Controls.

Although the voltage maps in Figure 7 suggest that the negativity in the high-proficiency subgroup resembles a LAN, while the negativity in the low-proficiency subgroup has an N400-like distribution, the lateral ANOVA did not reveal a significant *Hemisphere* x *Group* interaction. Proponents of the component overlap view proposed by Tanner and colleagues (Osterhout, 1997; Tanner et al., 2013; Tanner & Van Hell, 2014; Tanner, Inoue & Osterhout, 2014) would argue that both subgroups show an N400 effect, but that the left-anterior distribution of the negativity in the high-proficiency individuals is the result of the larger and earlier-onsetting parietal P600 that cancels out the N400 at electrodes sites where the two effects overlap. To investigate this possibility, we visually examined the scalp distribution of the negativity with a 50 ms moving window and determined that the negativity in high-proficiency individuals appeared at left-frontal electrodes as early as 250 ms. Thus, the steepness of the P600 in the high-proficiency group is not the reason for the frontal distribution of the negativity in these individuals. Given that, based on the literature, a LAN + P600 pattern may be the expected response to number agreement violations, it seems intuitive for the negativity to be more frontally-localized (and more left-lateralized, at least qualitatively) in high-proficiency than lower-proficiency individuals. However, the presence of a large negativity in both proficiency subgroups (as well as a shared negativity in Controls and Attriters) indicates that individuals from both groups and proficiency subgroups made predictions about the number of the modifier and were able to detect the number mismatch in a similar way.

The finding of an N400-like negativity is consistent with the view that computing certain types of agreement information requires access to lexical-semantic or discourse-level information (Barber et al., 2004; Deutsch & Bentin, 2001; Molinaro et al., 2008; see Molinaro, Barber & Carreiras, 2011), given that, at this point in the sentence, readers must determine the antecedent of the modifier and the subject of the sentence. An alternate possibility is that readers may have initially tried to attach the modifier to the immediately preceding word – the intervening noun which was inanimate, feminine and singular in all trials. Once this attachment failed (N400), they may have tried to associate the modifier with the subject-noun. A larger N400 may therefore reflect a better detection of this Gender+Number+Plausibility mismatch (in higher Italian proficiency individuals). However, the mismatch in gender and plausibility between the modifier and the intervening noun should have affected all conditions equally, and not only condition *xxy* where the only large N400 effect was observed. Moreover, if the N400 were driven by the number mismatch between the modifier and the intervening noun, we would expect to see differences between singular and plural trials, as the intervening noun was consistently singular. The clash between the modifier and its immediately preceding noun should have been strongest in subconditions where the modifier was plural. We investigated this possibility but did not find any systematic differences between singular/plural subconditions that supported this interpretation. Thus, we believe that the integration of the modifier into the previous sentence context and its long-distance attachment to the sentence-initial subject-noun is the reason we found a negativity that resembled both an N400 and a LAN (only in high-proficiency individuals).

With respect to the relationship between the N400 and different stages of the P600 on the modifier, we found a strong negative correlation between the N400 (300-500 ms) and each interval of the P600, such that larger negativities tended to be followed by smaller positivities. However, the strength of the correlation increased the further apart the time-windows became (600-900ms: $r = 0.401$, $p < 0.01$; 900-1000ms: $r = 0.500$, $p < 0.01$; 1000-1300ms: $r = 0.547$, $p < 0.01$), similar to the correlations we had observed on the verb.

Note that these kinds of correlations between negativities and subsequent positivities can be due to two distinct (although not exclusive) mechanisms. First, large P600s may temporarily overlap with preceding negativities and partly cancel out their amplitude, especially in later portions of the negativity. Secondly, the negativity and the positivity may reflect complementary (or even competing) cognitive processes or processing styles, such that one can functionally replace the other. In other words, when the process reflected by a negativity (e.g., mismatch detection based on lexical predictions) is strong, there may be no need for the process reflected by the positivity (e.g., the post-hoc processing of linguistic anomalies violating "combinatorial morphosyntactic processing constraints"; cf. Tanner & Van Hell, 2014, p. 298). According to Tanner and colleagues, participants often engage in one of these two processes and elicit either an N400 or P600, such that a biphasic pattern in the group average is unrepresentative of most individual-participant data (Nieuwland & Van Berkum, 2008; Osterhout, 1997; Tanner et al., 2013; Tanner & Van Hell, 2014; Tanner, Inoue & Osterhout, 2014). It must be noted, however, that the majority of our participants showed a biphasic N400 + P600 pattern in response to *xxy* violations.

The two arguments outlined above are essential ingredients in Tanner's recent publications regarding the relationship between N400s, P600s and LANs. However, in order to support their claims that the N400 and P600 reflect functionally complementary processing mechanisms, Tanner and colleagues typically correlate N400 and P600 amplitudes in immediately adjacent time-windows, where the effects are likely to be influenced by component overlap. The potential problem with this approach is that correlations due to such overlap are trivial, as one may simply correlate the earliest parts of the P600 (overlapping with the N400 and thereby reducing its amplitude at the point in time and space where the two intersect), with later parts of the same P600. As long as the underlying assumption is that the P600 is a monolithic component, this approach may be less of a problem. However, our (and other) data strongly suggest that this assumption may be wrong (see Section 4.3. below).

By correlating distant time-windows of the N400 and the late P600, one can examine whether larger positivities are indeed associated with weaker negativities, even across time-windows where these two effects cannot overlap in time and space. Given that Tanner, Osterhout and colleagues treat the P600 as a monolithic component, typically quantifying it in a large time-window (500-1000 ms) or *only* in the first portion of the positivity (500-800 ms), in both cases immediately adjacent to the time-window of the negativity, our current analyses extend their approach.

Our results preliminarily suggest that early negativities and very late posterior P600s may be complementary psycholinguistic processes in certain types of morphosyntactic violations (and more so than preceding portions of the P600). The late stage of the P600 has been associated with costly, elaborated re-analysis/repair processes. If a morphosyntactic anomaly is detected at an early stage (i.e., larger LAN/N400), it may require less elaborated processing and repair at a later stage of processing (i.e., smaller P600). Although deciphering the relationship between negativities and positivities was not a primary goal of our current study, we have reported these findings with the hope that future studies will begin to address these questions in more detail.

4.2.2. The 'Repair hypothesis' and xyy/xyx violation conditions

When comparing the two other violation conditions (*xyy* and *xyx*) relative to correct *xxx* sentences in order to evaluate Molinaro and colleagues' "*Repair Hypothesis*", we found that Attriters elicited a P600 for both violation conditions, while in Controls condition *xyy* seemed to overlap with the correct *xxx* condition, similar to the finding reported in the previous study (as well as in Molinaro, Kim, Vespignani & Job, 2008). This pattern was in the direction of our hypotheses, namely that Attriters would differ from Controls in their online repair/re-analysis strategies and may not effectively revise the interpretation (number) of the sentence based on the violation detected on the verb. In contrast, Attriters elicited a P600 effect on the modifier in response to *xyy* violations in addition to the P600 effect they elicited on the preceding verb when it clashed with the subject-noun, demonstrating that a repair process had not occurred by the time they reached the modifier. This idea of Attriters (and lower-proficiency individuals) engaging in less thorough revision/repair processes is also consistent with the finding that (a) P600 amplitudes were smaller in low-proficiency individuals and (b) late P600s were absent in Attriters relative to native-Controls. Despite the visible graded P600 pattern (*xxy* > *xyx* > *xyy*) in the grand average ERP data, however, group and/or proficiency differences for the *xyy* violation condition did not reach statistical significance. Narrowing the time-window to 600-750 ms based on post-hoc visual inspection did not alter the pattern of results. Our results, therefore, do not allow us to fully replicate Molinaro and colleagues' findings that number agreement mismatches are repaired on the verb prior to real-time integration of the upcoming modifier.

A possible reason for the discrepancy between our findings and those of Molinaro et al. (2011) is that, in our study, counterbalancing number across conditions may have better concealed predictabilities in sentences in a way to avoid strategy formation, thus allowing violation effects to continue to be elicited in response to processing difficulties. Secondly, task differences are likely to have also played a role. In previous studies supporting the "Repair hypothesis" (Molinaro et al., 2011; Molinaro, Kim, Vespignani & Job, 2008; Vespignani et al., in preparation), the task involved reading for comprehension rather than rating acceptability. It is possible that repair is not mandatory or not fully pursued in an acceptability judgment task as compared to reading for comprehension.

The absence of any negativity in the *xyy* and *xyx* conditions (as was the case in the original study by Molinaro et al., 2011) may elucidate the nature of the negativity observed in the xxy condition – given that, in xxy sentences, the modifier marked the first (and only) point of number agreement violation, it is conceivable that a biphasic N400/LAN + P600 pattern was elicited, similar to the biphasic pattern elicited on the verb when a first violation was detected between the number of the verb and the preceding subject-noun. The finding that the negativity is more robust for the modifier than on the verb may be in line with the notion that subject-verb number mismatches in Italian are less likely to elicit robust LAN/N400 effects due to the flexible word-order of the language and the possibility of a post-verbal subject noun, whereas encountering a modifier referring back to its previously-encountered antecedent noun must unambiguously agree with it in number. This idea of how ERP patterns may differ at the first point of violation vs. in subsequent violations in a given sentence is further elaborated below, in relation to the P600.

4.3. Different positivities reflect different underlying processes

An important contribution of our study was the finding of three distinct time "slices" of the P600 that were differentially affected by group membership and proficiency level, strongly supporting the notion that positivities typically referred to as P600s cannot be viewed as a monolithic component. The majority of morphosyntactic number agreement studies had not extended the time-window of the P600 beyond 900 or 1000 ms. To our knowledge, the present data are also among the first to demonstrate that separate time-windows of the P600 are *differentially* modulated by participant background factors.

In accordance with several number agreement studies including the original study by Molinaro and colleagues (2011), we replicated the early positivity that was prominent on fronto-central areas of the scalp (Barber & Carreiras, 2005; Hagoort & Brown, 2000; Kaan et al., 2000; Kaan, 2002; Kaan & Swaab, 2003; Molinaro et al., 2008 and 2011; Silva-Pereyra & Carreiras, 2007). In our study, we quantified this effect between 550-650 ms in order to minimize overlap with both the preceding negativity and the following posterior P600. The effect was elicited by the verb when it clashed in number with the sentence-initial subject-noun, and was similar for both Attriters and Controls. Some previous reports have attributed this effect to increased difficulty in integrating the mismatching constituent with the previous sentence fragment, and having to override the preferred structural representation of the sentence (Hagoort, Brown & Osterhout, 1999). However, one should also consider the large literature of similar frontal positivities that have been elicited in a variety of paradigms (e.g., in attended oddballs following the MMN) and are referred to as P3a components. The P3a is often driven by surprise (see Polich, 2007; Squires, Squires & Hillyard, 1975) and is viewed as part of an orientation response allocating special attention to the stimulus in question (e.g., Näätänen & Galliard, 1983). This interpretation would explain why the early frontal positivity in our study was present for violations realized on the verb (i.e., early in the sentence), but not on the modifier (i.e., where violations were more predictable), and was larger for trials in the first half of the experiment, compared to the second half (see Appendix b)[6]. In this vein, we argue that the early frontal positivity is a P3a, driven by a violation that occurs early on in a sentence without much context, and is absent when the violation occurs at a later point in the sentence (irrespective of whether it is the first or the second violation).

We then demonstrated a prototypical, posterior P600 effect in the window between 650-900 ms, elicited in response to subject-verb number violations, as well as in response to violations where the modifier disagreed with the verb (*xxy* and *xyx*). Importantly, this posterior P600 effect was modulated by L1 proficiency, irrespective of group membership. From 900 ms onwards, however, group became the meaningful factor in modulating the presence/absence of the P600, as only Controls continued to show a long-lasting positivity which only returned to baseline at 1200-1300 ms. This pattern was strikingly similar for both target words.

In terms of their functional significance, late parietal positivities in language studies typically referred to as P600 components have been shown to comprise subcomponents reflecting distinct

[6] Note that similar P3a effects have been found for other types of syntactic violations in early sentence positions, such as word category violations (e.g., Steinhauer et al., 2009). However, it is unclear if P3a components are restricted to reading studies or may also occur in auditory studies.

cognitive processes (e.g., Friederici et al., 2001). The P600 has been discussed either as a reflection of structural re-analysis or repair (Friederici, 2002; Hagoort et al., 1993), or in terms of monitoring or engagement of combinatorial processing streams (van Herten et al., 2005; Tanner & Van Hell, 2014). Given that substantial parts of the P600 amplitude have been shown to depend on task, with larger P600s elicited in studies employing grammaticality or acceptability judgments (see Molinaro, Barber & Carreiras, 2011, and Royle et al., 2013 for recent discussions), other researchers have argued that the P600 is related to metalinguistic "well-formedness" categorizations in judgment tasks and can, therefore, be interpreted as a late non-linguistic P300 (or P3b) component, reflecting the metalinguistic decision that as sentence is ungrammatical (Bornkessel-Schlesewsky et al., 2011; Coulson, King & Kutas, 1998; Friederici et al., 2001; but see Frisch & Kotz, 2003, Osterhout & Hagoort, 1999, and Steinhauer et al., 1997 for arguments as to why P600s cannot be reduced to 'just another P300').

Our results showed large P600 responses to subject-noun number violations on the verb, as well as to the (first and only) *xxy* violation on the modifier, whereas the P600 effects were significantly smaller in *xyx* and *xyy* conditions where the modifier marked the second point of violation within the sentence. Importantly, this graded P600 pattern was strikingly similar to the P600s reported in the original study by Molinaro and colleagues (2011) as well as in an English study with a similar experimental design by Molinaro, Kim, Vespignani, and Job (2008). This consistent pattern supports interpretations of the P600 in terms of a delayed P300 reflecting "well-formedness" judgments (e.g., Coulson et al, 1998; Bornkessel-Schlesewsky et al., 2011). Given the well-known sensitivity of P300 components to task relevance, one might expect a gradient P600 pattern as a function of task, with largest amplitudes in binary (yes/no) acceptability tasks, slightly reduced amplitudes in judgment tasks using Likert scales (as in our study), and rather small amplitudes for studies not requiring acceptability judgments at all (e.g., employing comprehension questions, probe word verification, or no task at all).

Interestingly and unexpectedly, the finding of a larger P600 for a first violation relative to a subsequent violation in a given sentence was shown to be uninfluenced by task – despite our use of an acceptability task on a five-point scale, our graded P600 effects were comparable in amplitude (as well as in latency and scalp distribution) to the pattern of P600 effects reported by Molinaro and colleagues (2011), although their task involved reading for comprehension. Based on these findings, we argue that the posterior P600 is partly influenced by "well-formedness" judgments (and, therefore, is larger for the first violation in a sentence and smaller for subsequent violations), but that this pattern is not entirely dependent on an explicit judgment task. This hypothesis should be further evaluated in future experiments.

The finding that the P600 between 650-900 ms was significantly larger in more proficient participants (in both groups) is in line with a number of previous studies reporting reduced (and sometimes delayed) P600s in less proficient L2 learners (e.g., Rossi et al., 2006). This observation can either be interpreted as a more confident categorization of sentences as "ungrammatical" in more proficient language users (eliciting a larger P3b component), or it may reflect a second cognitive process that also contributes to the overall P600 amplitude in this time interval. Since various studies have demonstrated that the P600 cannot be reduced to "just a P300" (e.g., Friederici et al., 2001; Osterhout & Hagoort, 1999) and that even certain semantic anomalies can elicit P600s – but only when subjects overtly detected (and presumably revisited) the semantic problem (Sanford et al., 2011) – we believe that these P600 proficiency effects are more likely to reflect psycholinguistic processes associated with structural re-analysis or combinatorial integration of different streams of information (e.g., semantic versus morphosyntactic information for number mismatches). The earlier phase of the P600 has also been characterized as reflecting an initial step of 'diagnosis' of the linguistic anomaly (e.g., Friederici et al., 2001), whereas later portions of the P600 have been associated with the actual revision or repair.

This leads us to the last phase of the P600 in our data (after 900 ms). In the literature, a longer-lasting P600 has been associated with costlier repair – a more elaborated effort to regularize an anomalous sentence in order to interpret its meaning (Molinaro et al., 2011). In this vein, we may interpret the shorter-lived P600 effects in Attriters as indicative of lesser elaborated repair. This is consistent with the finding of a P600 effect on the modifier in the *xyy* condition for Attriters, contrary to Controls in both our study (at least qualitatively) and in the previous experiment by Molinaro and colleagues (2011) that ours aimed to replicate. The lack of a long-lasting P600 is likely a sign of the reduced capacity of Attriters to pursue an in-depth online analysis of the error, in terms of re-analysis/repair processes to recover what the correct input should have been. Importantly, this processing difference was found to be a distinct feature of Attriters, beyond proficiency effects that were shared with the native-Controls. It appears, then, that attrition may differentially influence stages of error detection and re-analysis/repair and that, at least in early stages of attrition, online re-analysis/repair mechanisms are affected first.

It may be worth investigating whether amplitude differences observed between Attriters and Controls in the late P600 time-window eventually become amplitude differences in the earlier time-window with advancing attrition, or whether attrition continues to manifest itself differently from effects of overall proficiency that are shared with non-attriting individuals. In our current data, low-proficiency attriters exhibited P600 effects of smaller amplitudes, although this proficiency modulation was also true of the native-Controls. Similar to what has been done for L2 acquisition and *L2 attrition* (White et al., 2010; Tanner et al., unpublished data), it is of value to study L1 attrition longitudinally, to explore how advancing stages of L1 attrition may impact ERP processing patterns.

Overall, our results support the notion that different positivities reflect distinct underlying neurocognitive processes, which appear to be differentially impacted by L1 attrition and by proficiency level. We therefore encourage future studies to adopt the view that the P600 is not a monolithic component, and to systematically explore potential proficiency- or group-differences in separate, non-overlapping slices of the positivity.

4.4. Attrition as more than just proficiency variation

Our study revealed that L1 proficiency modulated the timing, amplitude and scalp distribution of certain ERP responses to ungrammaticality, namely the N400 elicited on the modifier and the P600 effects elicited in the first phase of the posterior P600 (between 650-900 ms) by both target words. These results are in line with similar accounts of proficiency effects on L1 morphosyntactic processing (Pakulak et al., 2014; Pakulak & Neville, 2010). Crucially, however, we also found *group* differences in the amplitude, timing and scalp distribution of ERP effects between Attriters and native Controls, demonstrating that a normal degree of proficiency variation within native-speakers cannot be the whole story. Thus, attrition does not seem to manifest itself as a generalized reduction in proficiency, neither behaviorally (acceptability judgments and proficiency measures) nor in terms of ERP effects that appear to be proficiency-dependent – at least not in the case of number agreement processing in such sentences. Instead, the differences appear to be rather subtle – at least in initial stages of attrition – and to affect specific processing routines during real-time comprehension. We examined four distinct processing phases of morphosyntactic number agreement (i.e., early negativity, frontal positivity/P3a, P600, late P600), and found additive effects of proficiency and attrition on these different phases of processing.

Our general pattern of results suggests that Attriters are less efficient or less thorough processors in their L1. Although they arrived at similar acceptability judgment ratings as native-Controls (albeit at significantly slower rates), they exhibited differences at separate stages of morphosyntactic processing that were qualitatively distinct from general proficiency effects. While Attriters did detect number agreement violations both on the verb as well as on the modifier, they differed in how they computed syntactic relations during online comprehension as well as how they *recovered* from erroneous input.

We also found interesting correlations between factors thought to modulate L1 attrition, such as length of residence and amount of L1/L2 exposure (see Köpke & Schmid, 2004; Schmid, 2011). Our analyses revealed that proficiency scores (overall proficiency and/or particular sub-tasks) were lower in individuals with a longer length of residence and with less L1 exposure. In further support of attrition effects, ERP patterns on the modifier were found to be modulated by amount of L1 exposure (on a daily basis, relative to L2 exposure). Crucially, this correlation was observed on the late P600 effect where group differences between Controls and Attriters were most pronounced; Attriters with higher L1 exposure were more similar to native-controls (in eliciting a late P600 effect for *xxy* violations) than Attriters with lower levels of L1 exposure.

The next necessary step is to include a second group of bilingual speakers of Italian and English, in order to determine whether patterns in Attriters are characteristic of "attrition" (and/or special socio-linguistic circumstances surrounding attrition) or whether at least part of these patterns are shared with other bilingual speakers of the same two languages (e.g., influence of English morphosyntax on Italian processing). As mentioned in a footnote at the end of the introduction, we chose to focus this first paper on Attriters and non-attriting native-speakers' processing of the L1, in order to first make the most compelling case for any group differences, rather than including the late learners for whom Italian is the L2. Our next goal is to determine whether, in this rather difficult experimental paradigm, late L2 learners of Italian show similar ERP responses as the native-Italian controls, and whether Italian proficiency level modulates their processing profiles in a similar way as in High/Low proficiency L1 speakers. Crucially, we aim to clarify whether the difference in the negativity elicited on the verb by Attriters and/or the durational differences in the P600 are specific to Attriters or are shared by L2 learners.

Although including L2 learners is important in order to clarify the "big picture" of attrition and neuroplasticity, we also believe that future research should pinpoint other "bilingual control groups" that may be a more suitable comparison, depending on the specific research question (e.g., comparing L2 to L1 processing vs. focusing on L1 processing). A future study could compare Italian late-learners of English (i.e., similar to the L1 attriters) who do not experience attrition in their L1 – for example, individuals who live in Italy and are considered highly proficient in English, but without the shift in dominance characteristic of attriters. In other words, it may be of value to include a group of Controls who are similar to our current controls but bilingual rather than monolingual. In both groups, Italian would still be the L1, but only the Attriters would be dominant in the L2. This could be a viable future direction to disentangle attrition effects from general bilingualism effects (e.g., transfer/interference or co-activation of the two languages), while *still* comparing individuals on their L1. In sum, the present study is only the first step on a potentially exciting road towards understanding the neurocognitive correlates underlying L1 attrition.

5. CONCLUSIONS

In one of the first ERP investigations of L1 attrition in morphosyntactic processing, we compared Attriters to non-attriting Italian native-speakers in their real-time detection and revision/repair of number agreement errors in their native-L1. In order to maximize the potential for differences between native-speakers and attriters, we opted to test potential processing difficulties in demanding sentences where agreement computation involved multiple inflected constituents over a longer span, rather than only local mismatches. Our results revealed interesting differences between attriters and non-attriting individuals in the presence, amplitude, latency and scalp distribution of ERP effects previously associated with the processing of morphosyntactic agreement violations during real-time comprehension. Interestingly, processing differences were not mirrored by behavioral performance, suggesting that attrition effects may begin to show at the brain-level before they are robust enough to be detected behaviorally, or that certain tasks are not sensitive enough to tap into subtle attrition effects. Previous studies claiming that morphosyntax is resistant to attrition may have tested morphosyntactic properties that were too undemanding to pose a challenge for attriters and to reveal group/proficiency differences.

We also showed that, irrespective of group, some processing differences were driven by Italian proficiency level, even within native-speakers. Proficiency-based differences in the neurocognitive mechanisms underlying language processing have been reported in studies of L2 processing as well as a handful of studies with monolingual speakers that attempted to elucidate the importance of this factor in determining the *nativelikeness* of language processes in the brain. This study therefore begins to bridge the gap between issues that have been studied in L2 processing literature and the study of L1 attrition. In this vein, we also addressed a number of theoretical and methodological questions that are imperative for advancing the field and that could be considered as avenues for impending research (e.g., the functional significance of different types of positivities). These findings are novel and highly topical, as they provide the first ERP evidence of attrition effects in the L1 grammar of adult first-generation immigrants, but also extend L2-processing research in a number of ways, given our exploration of issues such as neuroplasticity, dominance and proficiency.

By illustrating differences between groups of Italian *native*-speakers who had acquired and used Italian in an exclusively monolingual context up until adulthood, our findings cast doubt on a neurobiological 'critical period hypothesis' or similar claims that one's L1 is stable and has a privileged status in the brain. We instead advocate for ongoing neuroplasticity for language, even in adulthood, and emphasize that proficiency is a key factor in modulating native-like neurocognitive responses, regardless of whether the language being processed is the L2 or the L1.

ACKNOWLEDGEMENTS

K.K was supported by a Vanier Canada Graduate Scholarship from the Canadian Institutes of Health Research, as well as a Richard Tomlinson Doctoral Fellowship awarded by the Faculty of Medicine of McGill University. K.K's research stay in Italy for data collection was made possible by a Michael Smith Foreign Study Supplement (MSFSS-CGS). This research was supported by grants awarded to K.S. by the Canada Research Chair program and the Canada Foundation for Innovation (CRC/CFI; project # 201876), the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPGP 312835-05 and RGPIN 402678-11), the Social

Sciences and Humanities Research Council of Canada (SSHRC; # 435-2013-2052), and the Fonds de Recherche Société et Culture, Québec (FQRSC; # 2010-SE-103727). The authors thank Nicola Molinaro and Eleonora Rossi for valuable feedback on the experimental design and stimuli, as well as John E. Drury and Stefanie Nickels for comments on a draft of this manuscript. We are also indebted to research assistants Linna Jin, Kristina Maiorino, Filippo Vicari and Paolo Zandomeneghi for their invaluable help with data collection.

Study 1: Tables and figures

Table 1. *Experimental stimuli by condition. Number was counterbalanced such that the subject noun was either singular or plural. English translations are presented in italics. Target words (verb, modifier) are underlined. The asterisk marks the point of first violation.*

CONDITION	Subject-noun	Verb	Intervening phrase	Modifier	Prepositional phrase
xxx: Correct	**x**	**x**		**x**	
Singular	Il lavoratore	torna	dalla fabbrica	sporco	di grasso.
	The worker(sg)	*returns*(sg)	*from the factory*	*dirty*(sg)	*with grease.*
Plural	I lavoratori	tornano	dalla fabbrica	sporchi	di grasso.
	The workers(pl)	*return*(pl)	*from the factory*	*dirty*(pl)	*with grease.*
xyx: Inconsistent verb	**x**	**y**		**x**	
Singular	Il lavoratore	*tornano	dalla fabbrica	sporco	di grasso.
	The worker(sg)	**return*(pl)	*from the factory*	*dirty*(sg)	*with grease.*
Plural	I lavoratori	*torna	dalla fabbrica	sporchi	di grasso.
	The workers(pl)	**returns*(sg)	*from the factory*	*dirty*(pl)	*with grease.*
xyy: Inconsistent noun	**x**	**y**		**y**	
Singular	Il lavoratore	*tornano	dalla fabbrica	sporchi	di grasso.
	The worker(sg)	**return*(pl)	*from the factory*	*dirty*(pl)	*with grease.*
Plural	I lavoratori	*torna	dalla fabbrica	sporco	di grasso.
	The workers(pl)	**returns*(sg)	*from the factory*	*dirty*(sg)	*with grease.*
xxy: Inconsistent modifier	**x**	**x**		**y**	
Singular	Il lavoratore	torna	dalla fabbrica	*sporchi	di grasso.
	The worker(sg)	*returns*(sg)	*from the factory*	**dirty*(pl)	*with grease.*
Plural	I lavoratori	tornano	dalla fabbrica	*sporco	di grasso.
	The workers(pl)	*return*(pl)	*from the factory*	**dirty*(sg)	*with grease.*

Table 2. *Predictions/hypotheses by condition based on findings by Molinaro, Vespignani, Zamparelli & Job (2011).*

CONDITION	Predicted ERP effects on VERB	Predicted ERP effects on MODIFIER
xxx: Correct	--	--
xyx: Inconsistent verb	LAN/N400 and/or Frontal positivity and/or P600	LAN/N400 and/or frontal positivity and/or P600 If "Repair hypothesis" fully true, then *xyx* = *xxy* (as subject number changed to match verb)
xyy: Inconsistent noun	LAN/N400 and/or Frontal positivity and/or P600	If "repaired" on verb number (2011 results), then *xyy* = *xxx* If mismatch perceived between modifier and subject (antecedent), then LAN/N4 and/or frontal positivity and/or P600
xxy: Inconsistent modifier	--	LAN/N400 and/or frontal positivity and/or P600 If "Repair hypothesis" fully true, then *xxy* = *xyx* If *xxy* most salient violation, then largest violation effects

Table 3. *Group means (standard deviation) for proficiency and control tasks (ps > 0.1).*

BEHAVIORAL MEASURES	CONTROLS (n = 30)	ATTRITERS (n = 24)
Self-report of proficiency (7 point scale)	7 (0)	6.87 (0.2)
Listening comprehension	7 (0)	7 (0)
Reading comprehension	7 (0)	7 (0)
Pronunciation	7 (0)	6.96 (0.2)
Fluency	7 (0)	6.79 (0.6)
Vocabulary	7 (0)	6.63 (0.7)
Grammar	7 (0)	6.83 (0.4)
C-test (%)	96.3 (4.4)	95.2 (4.6)
Error-detection test (%)	90.0 (5.1)	89.5 (5.9)
Verbal semantic fluency (average of 2 categories)	23.4 (5.5)	21.5 (3.9)
Reading fluency (# correct in 3 minutes)	71.6 (13.0)	75.3 (15.0)
Working memory		
Correct	11.2 (2.7)	11.9 (2.6)
Span	5.4 (1.1)	5.7 (1.1)

Table 4. *Proficiency subgroup means (standard deviation) for proficiency and control tasks.*
** p < 0.05. ** p < 0.01. *** p < 0.005. ns = not significant (p > 0.1)*

BEHAVIORAL MEASURES	HP CONTROLS		LP CONTROLS	HP ATTRITERS		LP ATTRITERS	HP ALL		LP ALL
Self-report of proficiency (7pt.scale)	7 (0)	--	7 (0)	6.97 (0.1)		6.74 (0.3)	6.9 (0.1)	ns	6.8 (0.2)
Listening comprehension	7 (0)	--	7 (0)	7 (0)	--	7 (0)	7 (0)	--	7 (0)
Reading comprehension	7 (0)	--	7 (0)	7 (0)	--	7 (0)	7 (0)	--	7 (0)
Pronunciation	7 (0)	--	7 (0)	7 (0)	ns	6.9	7 (0)	ns	6.9 (0.2)
Fluency	7 (0)	--	7 (0)	7 (0)	ns	6.5	7 (0)	*	6.8 (0.6)
Vocabulary	7 (0)	--	7 (0)	6.9	*	6.3	6.9 (0.2)	*	6.7 (0.6)
Grammar	7 (0)	--	7 (0)	6.9	ns	6.7	6.9 (0.2)	ns	6.8 (0.3)
C-test (%)	98.8 (1.1)	***	93.6 (2.8)	97.6 (2.7)	***	92.3 (4.7)	98.2 (2.1)	**	93.0 (3.7)
Error-detection test (%)	94.6 (3.4)	***	86.2 (4.3)	93.9 (3.6)	***	84.3 (3.4)	94.3 (3.6)	**	85.4 (3.9)
Verbal semantic fluency (average)	23.5 (6.9)	ns	21.6 (5.6)	23.8 (3.0)	***	18.9 (3.3)	23.6 (5.3)	*	20.5 (4.9)
Reading fluency (# correct)	73.4 (9.1)	ns	70.0 (15.8)	82.4 (12.4)	**	66.8 (13.9)	77.7 (68.7)	*	68.7 (14.8)
Working memory									
Correct	11.5 (2.4)	ns	10.9 (2.9)	13.2 (2.7)	**	10.5 (1.5)	12.3 (2.7)	*	10.7 (2.4)
Span	5.4 (1.1)	ns	5.3 (1.3)	6.2 (1.2)	*	5.1 (0.7)	5.8 (1.2)	ns	5.2 (1.1)

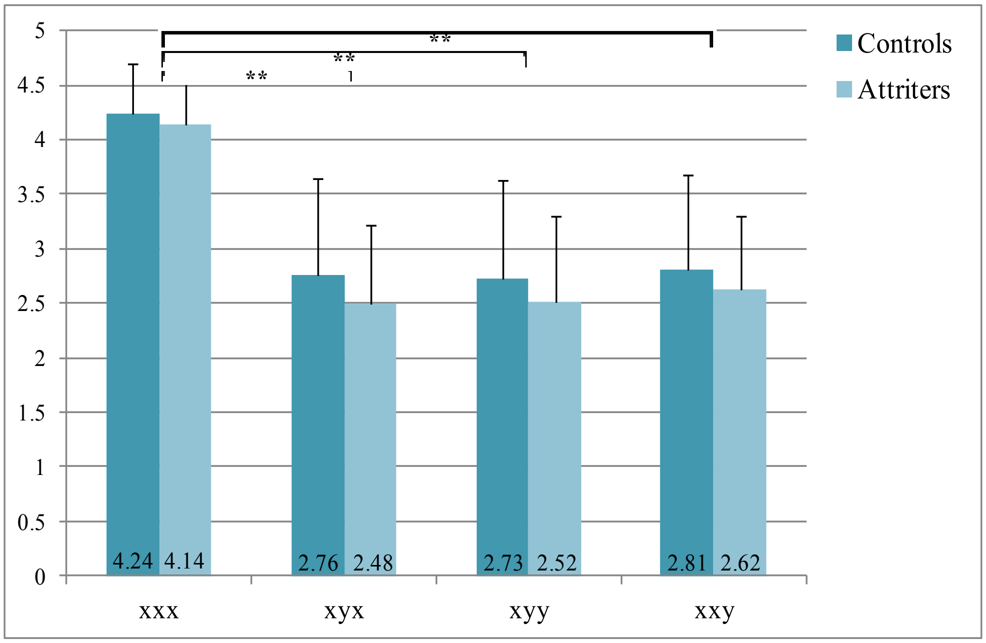


Figure 1a. Group acceptability ratings on a scale from 1 (completely unacceptable) to 5 (perfect) by condition. Attriters do not differ overall from Controls. * $p < 0.05$. ** $p < 0.01$. Error bars represent standard deviation.

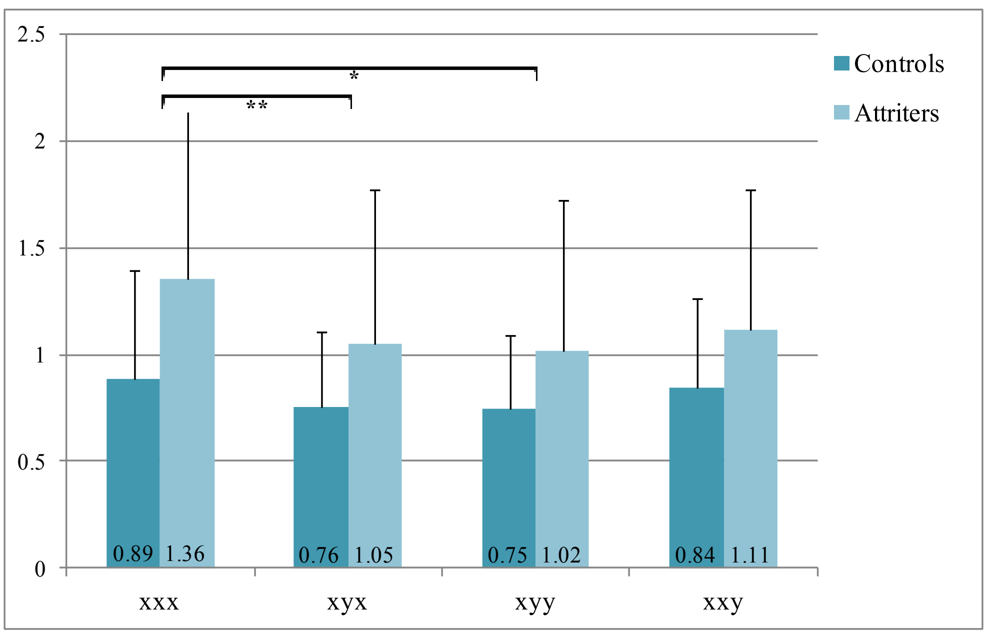
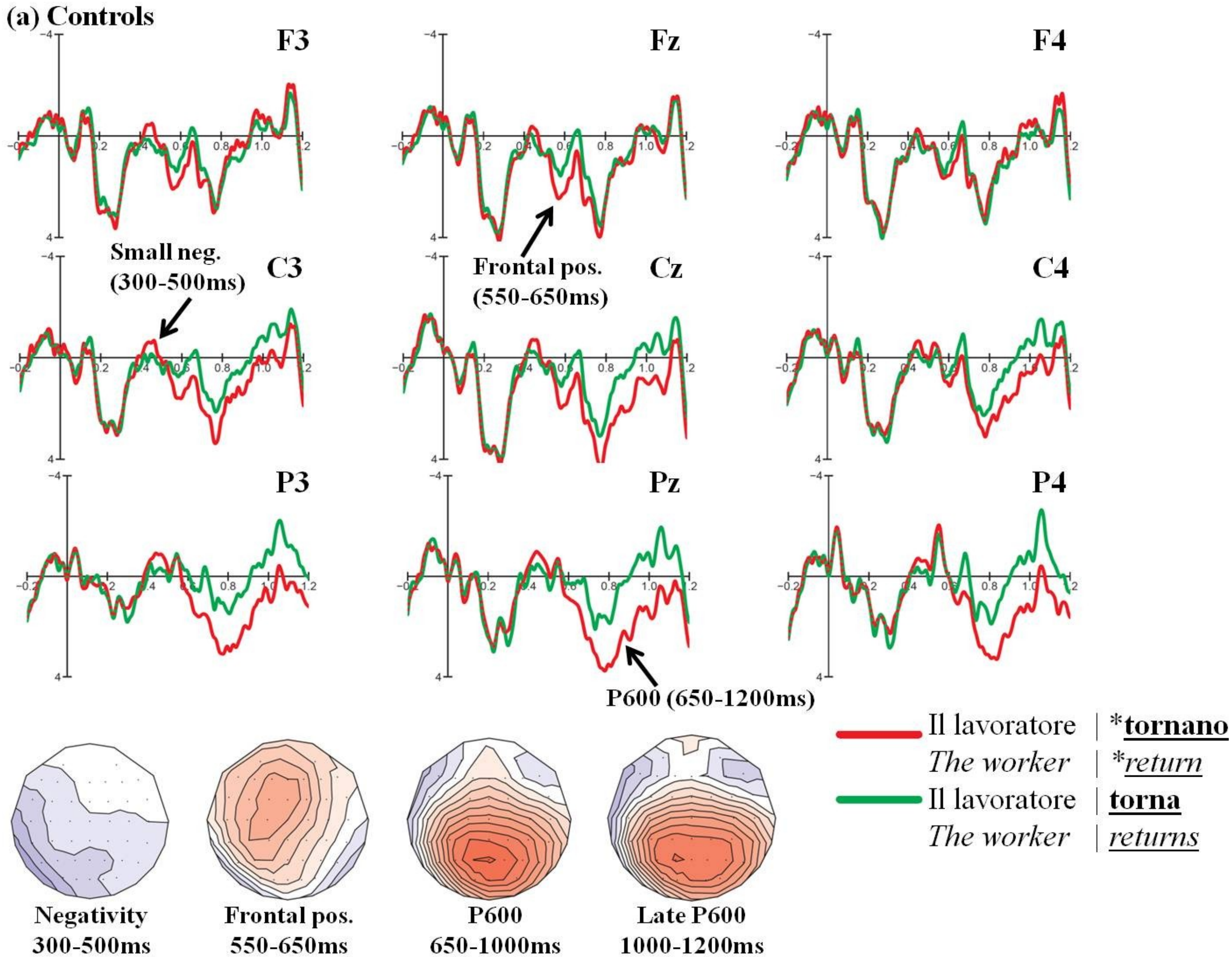


Figure 1b. Group reaction times (in seconds) by condition. Attriters were consistently slower than Controls ($p < 0.05$). * $p < 0.05$. ** $p < 0.01$. Error bars represent standard deviation.


(a) Controls
F3
Fz
F4
Small neg.
(300-500ms)
C3
Frontal pos.
(550-650ms)
Cz
C4
P3
Pz
P4
P600 (650-1200ms)
Il lavoratore | *tornano
The worker | *return
Il lavoratore | torna
The worker | returns
Negativity
300-500ms
Frontal pos.
550-650ms
P600
650-1000ms
Late P600
1000-1200ms

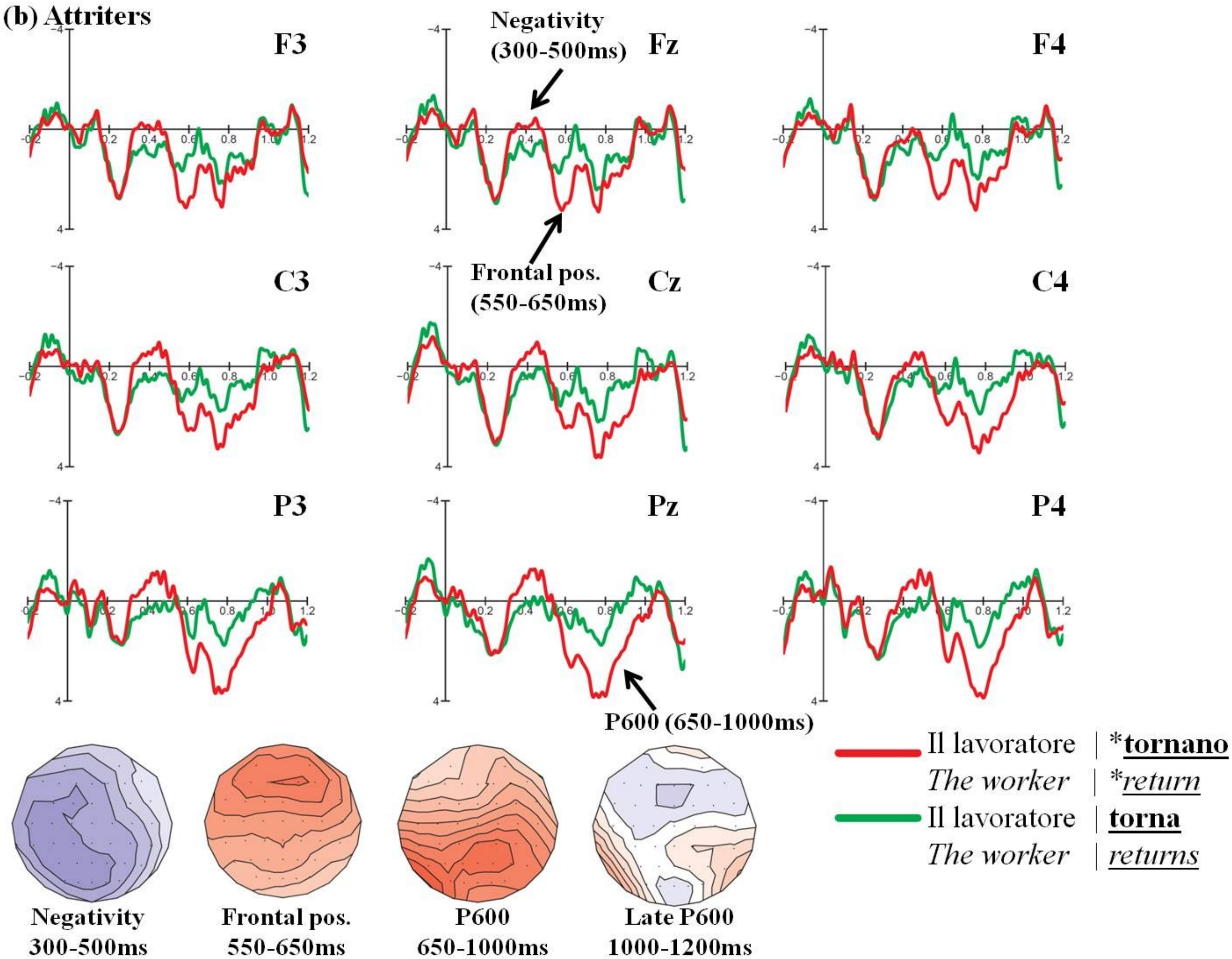


Figure 2. ERPs elicited by the verb in response to Ag1 violations (red) compared to Ag1 correct (green) in Controls **(a)** and Attriters **(b)**. Time ranges (in milliseconds) depicted on the x-axis are relative to the onset of the verb (0 ms). Negative values are plotted up. Voltage maps illustrate the scalp distribution of the effects observed for the time-windows of interest.

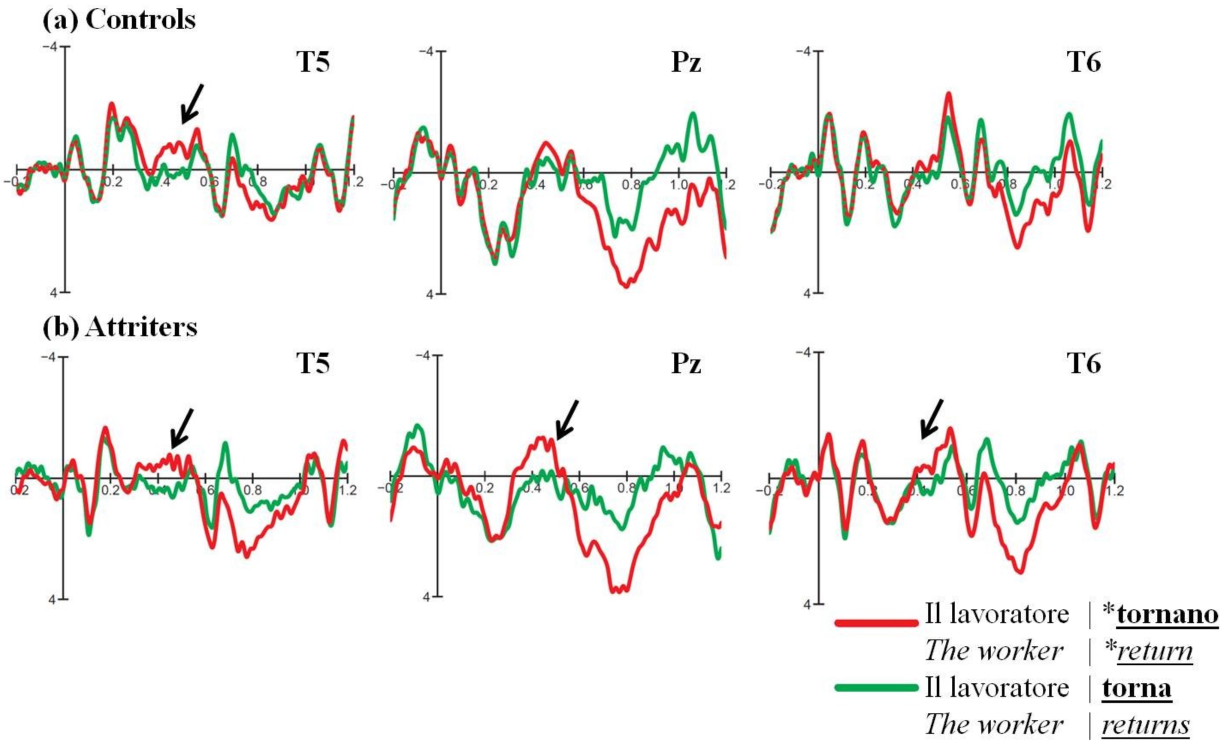


Figure 3. A comparison of electrodes T5 (left), Pz (midline) and T6 (right) for Ag1 violations (red) relative to Ag1 correct (green) in Controls and Attriters. Time ranges (in milliseconds) depicted on the x-axis are relative to the onset of the verb (0 ms). Negative values are plotted up. The negativity elicited in Controls is focused at T5, while it is broadly distributed for Attriters.

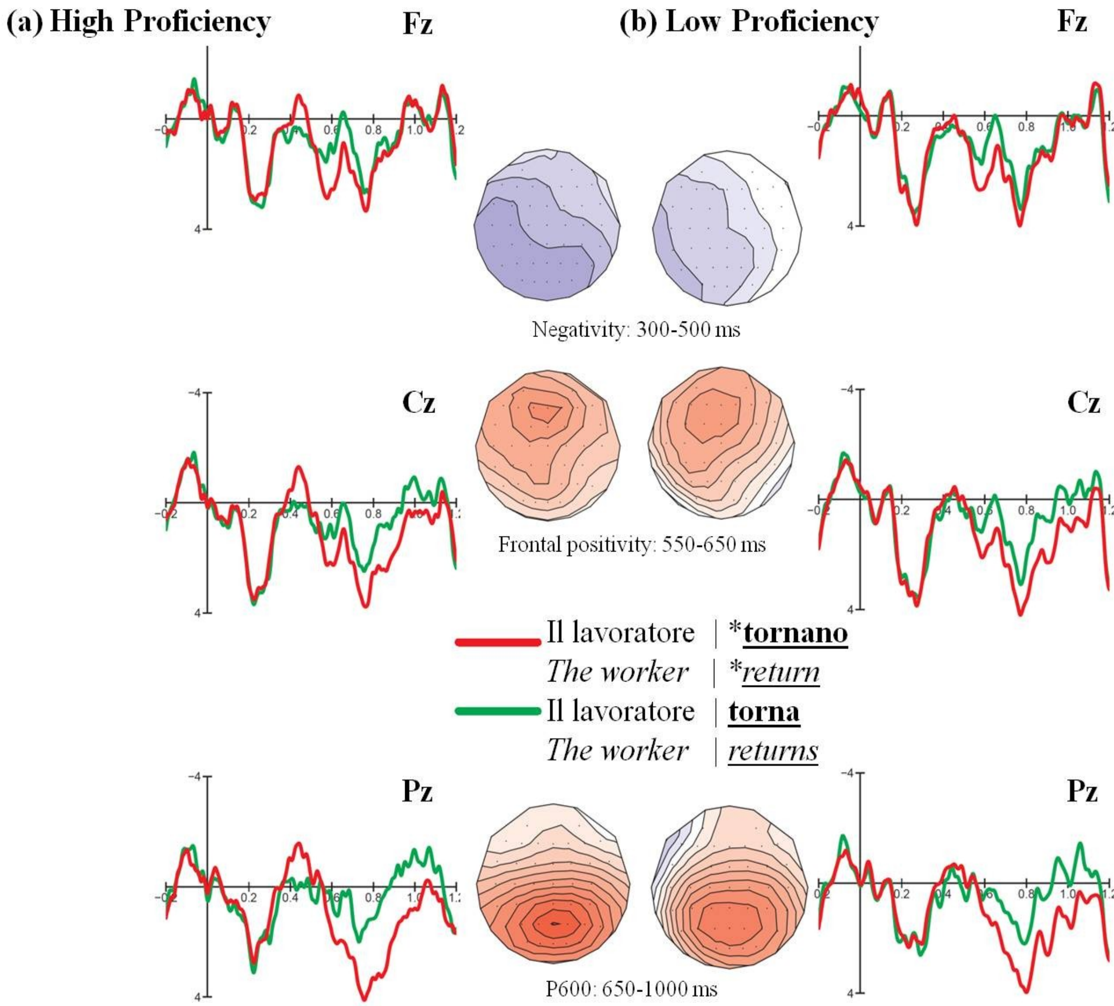


Figure 4. ERPs elicited by the verb in response to Ag1 violations (red) compared to Ag1 correct (green) in High Proficiency (a) versus Low Proficiency (b) individuals. Proficiency appears to modulate the amplitude of the negativity (300-500 ms) as well as the P600 (650-1000).

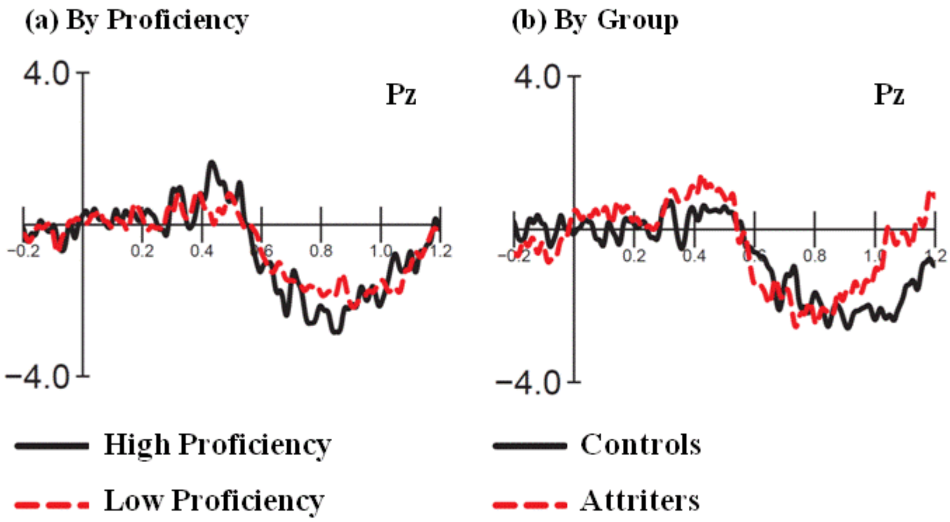


Figure 5. P600 difference waves (Ag1 violation – correct) at Pz emphasizing proficiency differences (LP in red < HP in black) in the early time window (a) but group differences (Attriters in red < Controls in black) in the late time window (b).

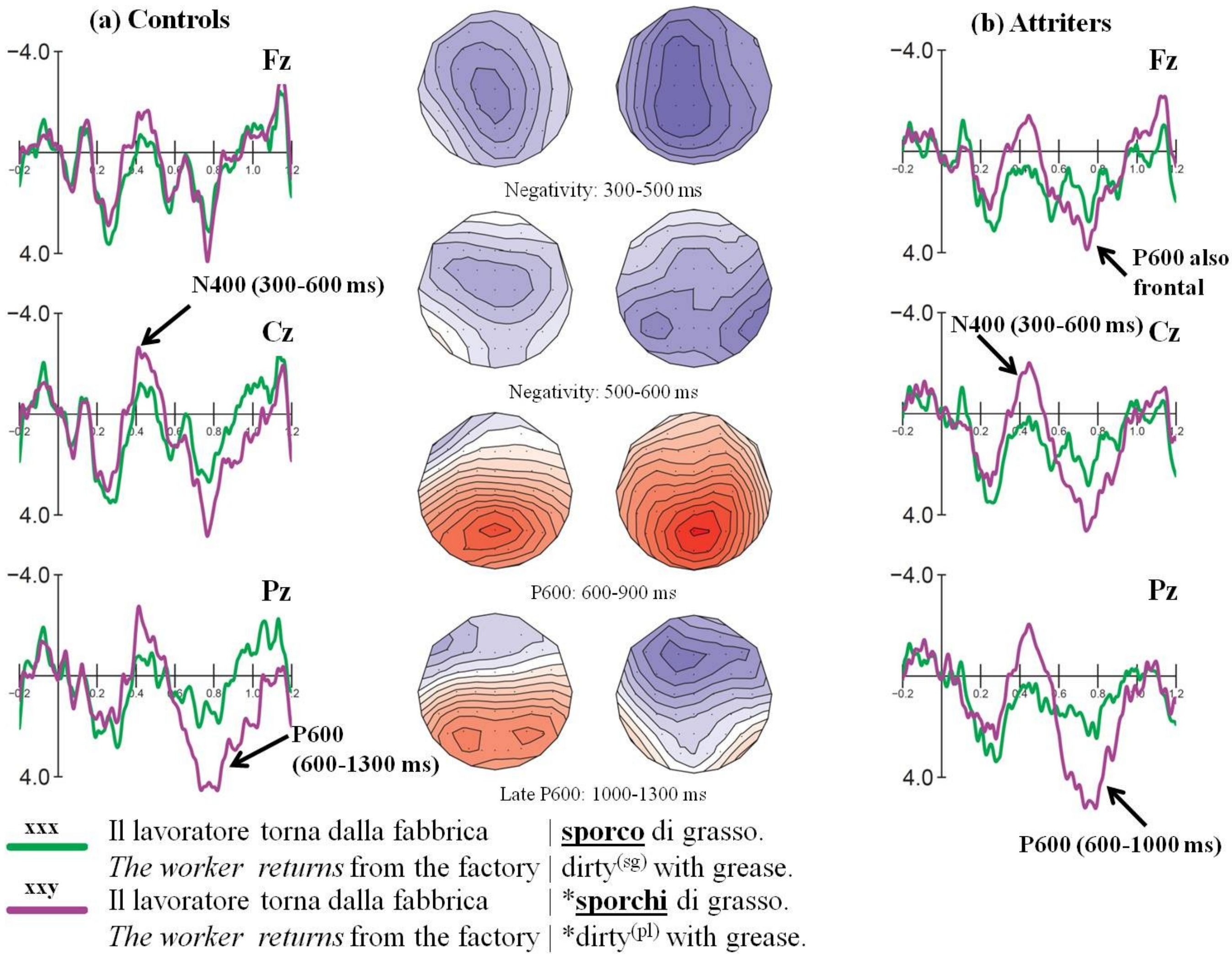


Figure 6. ERPs elicited by the modifier in response to *xxy* violations (purple) compared to *xxx* (green) shown at the midline for Controls (a) and Attriters (b). Voltage maps illustrate that the topography of the effects are similar in both groups, but the P600 in Attriters appears less focal and shorter in duration.

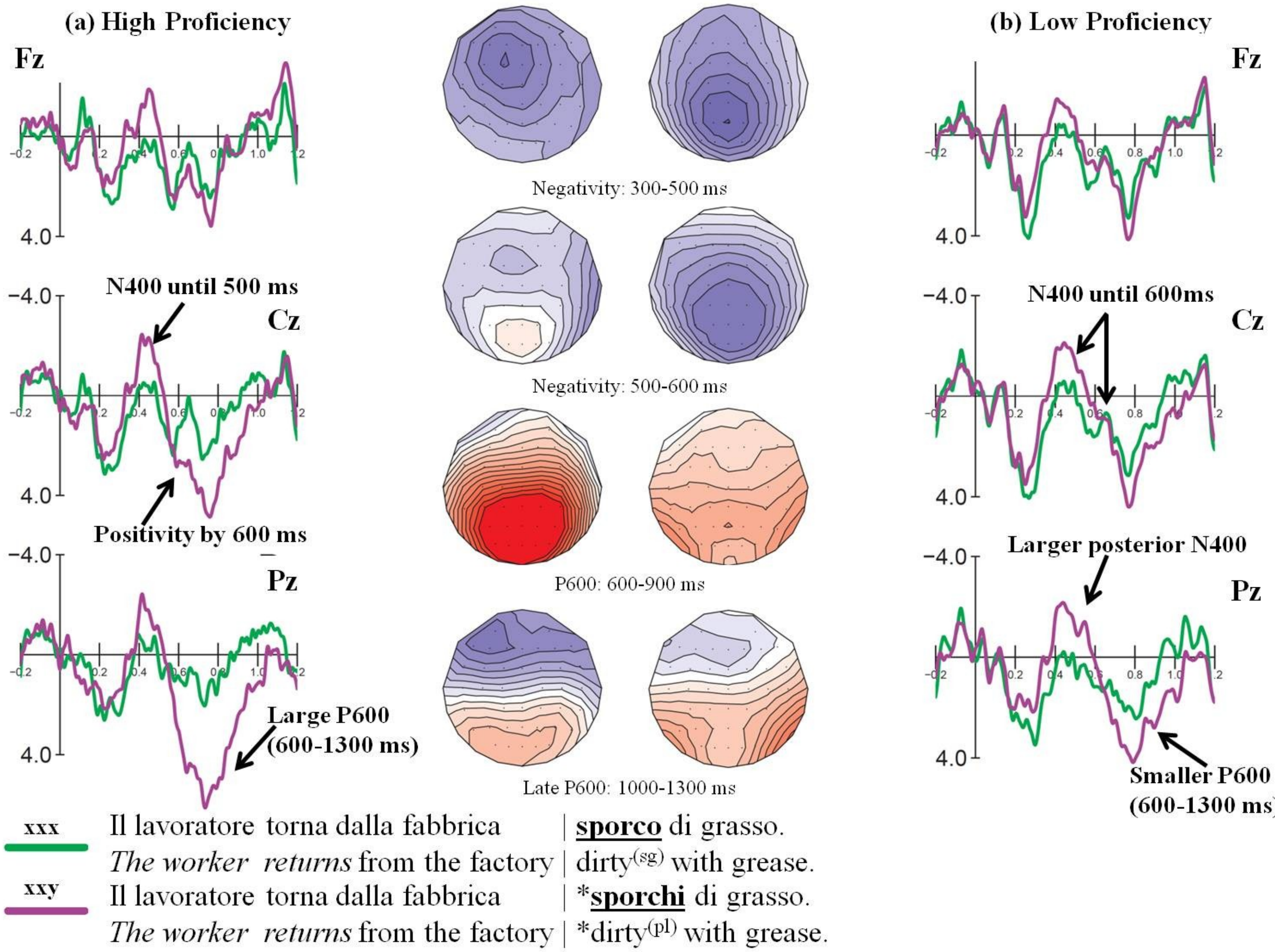


Figure 7. ERPs elicited by the modifier for *xxy* violations (purple) compared to *xxx* (green) in High Proficiency (a) versus Low Proficiency (b) groups. Low proficiency individuals show a less frontal and longer lasting negativity, as well as a weaker P600 effect relative to High Proficiency speakers.

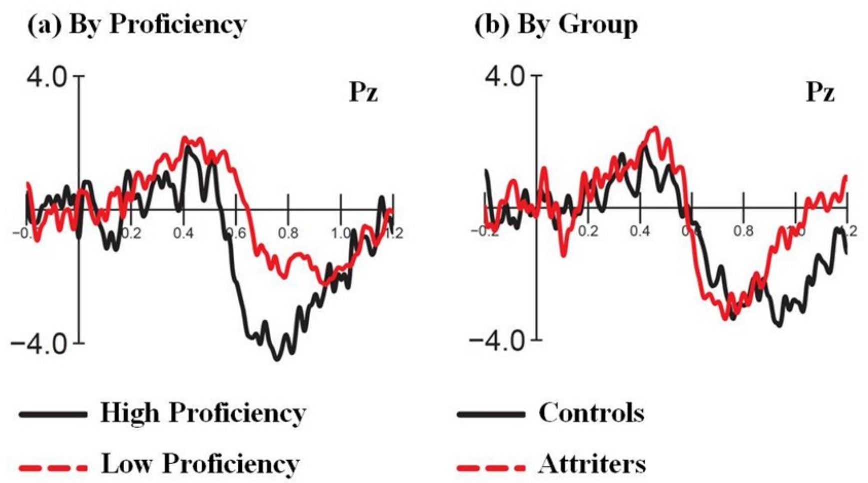


Figure 8. P600 difference waves (*xxy* - *xxx*) at Pz emphasizing proficiency differences (LP in red < HP in black) in the early time window (a) but group differences (Attriters in red < Controls in black) in the late time window (b).

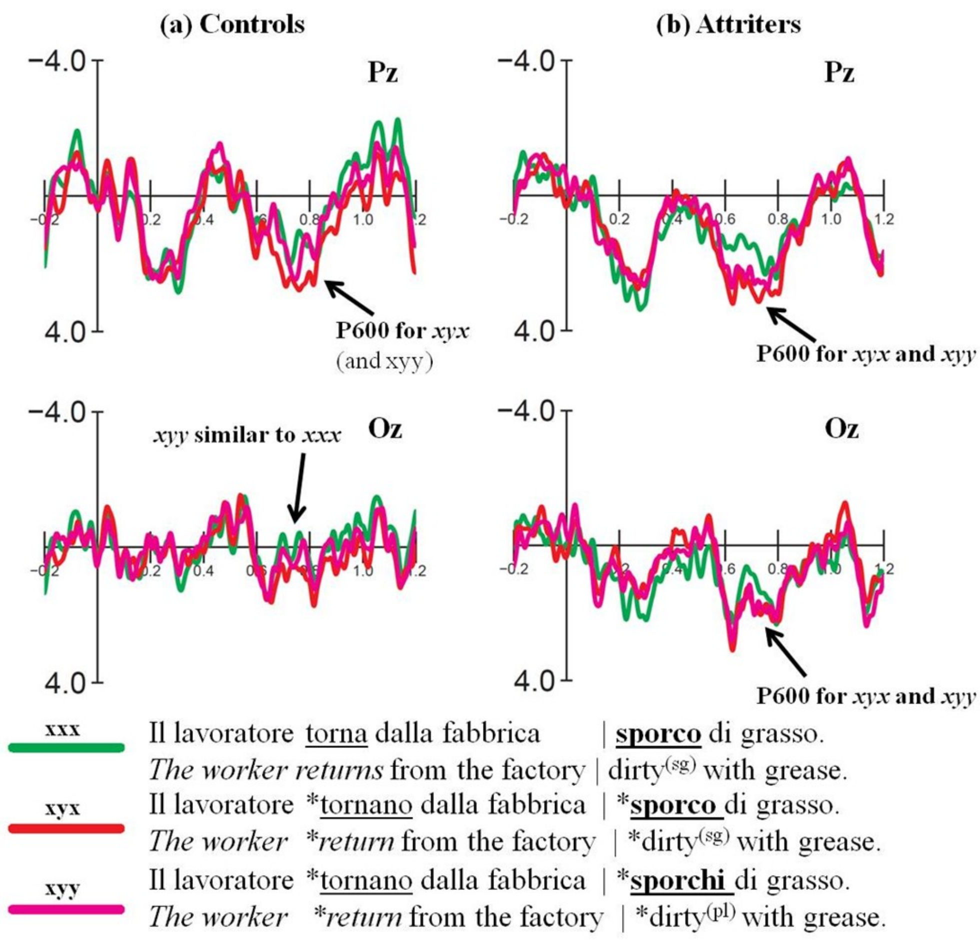
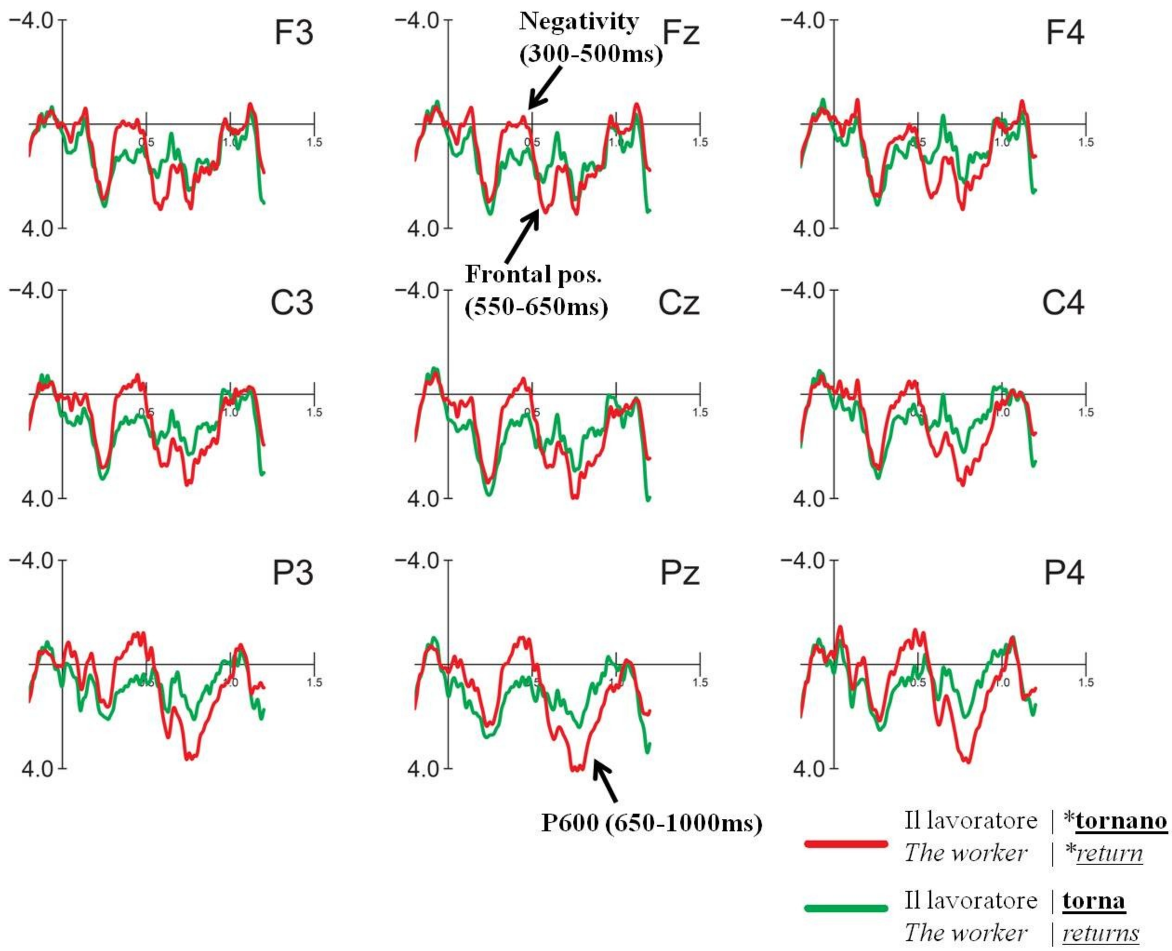
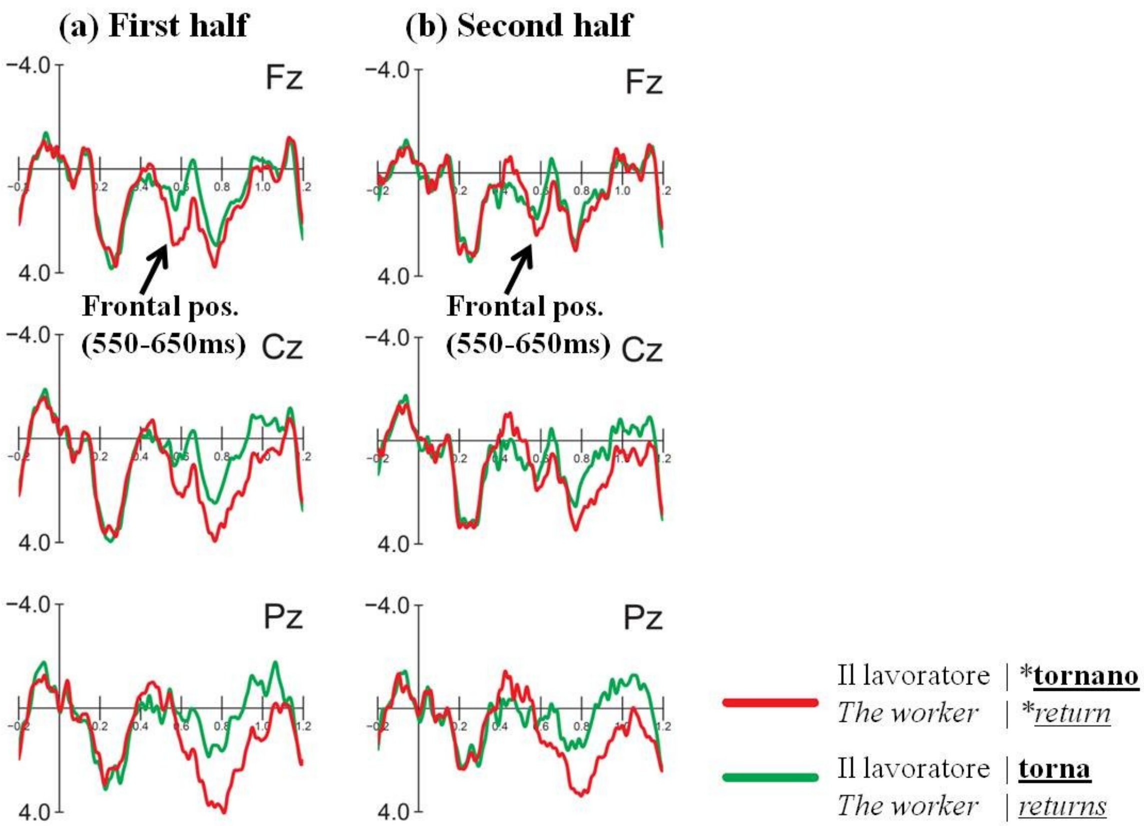


Figure 9. ERPs elicited by the modifier in response to *xyx* (red) and *xyy* (pink) violations relative to *xxx* at the most representative electrode (Pz). In Controls, *xyy* seems to elicit a smaller P600 than *xyx* violations, consistent with the 'Repair hypothesis'. In Attriters, both violations seem to overlap and to elicit a P600 effect.

Study 1: Appendices



Appendix a. An example of data patterns with the original baseline correction (-200 to 0 ms). The N400, frontal positivity and P600 effects we reported on the verb for Attriters (with a baseline of -200 to 200 ms) were also reliable with this original baseline.



Appendix b. An illustration of the larger frontal positivity / P3a amplitudes on the verb during the first half of the experiment compared to the second half (for all subjects).

Study 1: 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.

Ammerlaan, T. (1996). *You Get a Bit Wobbly...: Exploring Bilingual Lexical Retrieval Processes in the Context of First Language Attrition*. PhD dissertation, Nijmegen University.

Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccarelli, L., Palomba, D., & Stegagno, L. (2002). Cortical brain responses to semantic incongruity and syntactic violation in Italian language: an event-related potential study. *Neuroscience Letters*, *322*(1), 5–8.

Balconi, M., & Pozzoli, U. (2005). Comprehending Semantic and Grammatical Violations in Italian. N400 and P600 Comparison with Visual and Auditory Stimuli. *Journal of Psycholinguistic Research*, *34*(1), 71–98.

Barber, H.A., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137-153.

Barber, H., Salillas, E., & Carreiras, M. (2004). Gender or genders agreement? In M. Carreiras & C. Clifton (Eds.), *On-line study of sentence comprehension; eye-tracking, ERP and beyond*. Brighton, UK: Psychology Press

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A crosslinguistic study. *Cognition, 11*(3), 245-299.

Bergmann, Berends, Brouwer, Meulman, Seton, Sprenger, Stowe & Schmid (2013). Processing of gender in L2 acquisition and L1 attrition of German: Evidence from event-related potentials. Talk presented at the Workshop on Neurobilingualism, Groningen, The Netherlands, August 25-27, 2013.

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

Bornkessel-Schlesewsky, I., Kretschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language, 117*, 133–152.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*(13), 2492–2511.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, *20*(1), 98–105.

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,* 161–174

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(01), 3–42.

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

Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58.

Datta, H. (2010). *Brain bases for first language lexical attrition in Bengali-English speakers*. City University of New York. Retrieved from http://gradworks.umi.com/34/09/3409174.html

De Bot, K. (1996). Language loss. In H. Goebl, P. Nelde, Z. Stary, & W. Wölk (Eds.), *Contact linguistics: An international handbook of contemporary research*, Vol. 1 (pp. 579 – 585). Berlin: Walter de Gruyter.

De Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(1), 33-40.

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L. & Vespignani, F. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and Language, 85*(2), 280-296.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45*(2), 200-224.

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

Friederici, A.D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31*(1), 45-63.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001).Syntactic parsing preferences and their on-line revisions: A spatio temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*, 305–323.

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*(1), 529–534.

Frisch, S., Kotz, S.A., von Cramon, D.Y., & Friederici, A.D. (2003). Why the P600 is not just a P300: the role of the basal ganglia. *Clinical Neurophysiology, 114*, 336–340.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Cognitive Neuroscience, Journal of*, *15*(6), 883–899.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*(11), 1531-1549.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439-483.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266.

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

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931.

Jasper, H. H. (1958). Report to the committee on methods of clinical examination in electroencephalography. Appendix: The ten-twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371-375.

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.

Kaan, E. (2002). Investigating the effects of distance and number interference in agreement processing: An ERP study. *Journal of Psycholinguistic Research, 31*(2), 165-193.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159-201.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98-110.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013a). When the second language takes over: ERP evidence of L1-attrition in morphosyntactic processing. Talk presented at the International Conference on Multilingualism, Montreal, Canada, October 24-25, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013b). My Italian is not what it used to be: Investigating the neural correlates of L1 attrition and late L2 acquisition. Poster presented at the Workshop on Neurobilingualism, Groningen, The Netherlands, August 25-27, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014a). The case of the non-native-like first language: ERP evidence of L1-attrition in lexical and morphosyntactic processing. Symposium presentation, The 17th World Congress on Psychophysiology, Hiroshima, Japan, September 23-27, 2014.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014b). Neurophysiological correlates of L1 attrition and L2 acquisition: A continuum based on proficiency. Poster presented at the Society for the Neurobiology of Language, Amsterdam, Netherlands, August 27-29, 2014.

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

Köpke, B., & Schmid, M. S. (2004). Language Attrition: The Next Phase. In M. S. Schmid, B. Köpke, M. Keijzer & L. Weilemar, L. (Eds.), *First Language Attrition: Interdisciplinary perspectives on methodological issue*s (pp. 1-43). Amsterdam: John Benjamins.

Köpke, B. (1999). *L'attrition de la première langue chez le bilingue tardif: Implications pour l'étude psycholinguistique du bilinguisme*. Unpublished Doctoral Dissertation, Toulouse: Université de Toulouse-Le Mirail.

Kras, T. (2008). *L2 acquisition of the lexicon-syntax interface and narrow syntax by child and adult Croatian learners of Italian*. Unpublished doctoral dissertation, University of Cambridge.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition, 11*(5), 539-550.

Leinonen, A., Brattico, P., Jarvenpaa, M., & Krause, C.M. (2008). Event-related potential responses to violations of inflectional and derivational rules of Finnish. *Brain Research, 1218*, 181-193.

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

MacWhinney, B., & Bates, E. (1989). *The Crosslinguistic Study of Sentence Processing.* New York: Cambridge University Press.

Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011). When persons disagree: an ERP study of Unagreement in Spanish. *Psychophysiology, 48*(10), 1361–1371.

Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011). A person is not a number: Discourse involvement in subject-verb agreement computation. *Brain Research, 1410,* 64-76.

Mancini, S., Vespignani, F., Molinaro, N., Laudanna, A., & Rizzi, L. (2009). Number Agreement Processing with Different Persons: An ERP Study. Poster presented at the 15th AMLAP Conference. Barcelona, Spain, September 7-9, 2009.

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

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930.

Molinaro, N., Kim, A., Vespignani, F., & Job, R. (2008). Anaphoric agreement violation: An ERP analysis of its interpretation. *Cognition*, *106*(2), 963–974.

Molinaro, N., Vespignani, F., Zamparelli, R., & Job, R. (2011). Why brother and sister are not just siblings: Repair processes in agreement computation. *Journal of Memory and Language*, *64*(3), 211–232.

Montrul, S. (2008). *Incomplete acquisition in bilinguals: Re-examining the age factor.* Amsterdam: John Benjamins.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, *22*(2), 205–220.

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure: An ERP study in adult learners. *BMC Neuroscience, 10*: 89.

Näätänen, R., & Galliard, A.W.K. (1983). The orienting reflex and the N2 deflection of the ERP. In A.W.K. Galliard and W. Ritter (Eds.), *Tutorials in Event Related Potential Research: Endogenous Components*. Elsevier/North-Holland, Amsterdam. 1983: 119-141.

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*, 155–170.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research, 1164*, 81-94.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience, 24*(5), 1205-1223.

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*, 119–131.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-1228.

Opitz, C. (2011). *First language attrition and second language acquisition in a second language environment.* PhD dissertation. Centre for Language and Communication Studies. Trinity College Dublin. Dublin.

Orsini, A., Pezzuti, L. (2013). *WAIS-IV. Manuale.* Giunti OS: Firenze.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494-522.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 Debate. *Language and Cognitive Processes, 14*(1), 1-14.

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language-specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience, 8*(6), 507-526.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In Carreiras M and Clifton C (Eds), *The On-line Study of Sentence Comprehension: Eyetracking, ERP, and Beyond*. London: Psychology Press.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning*, *56*(s1), 199–230.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*(6), 739-773.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*(6), 509–521.

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

Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience*, *23*(10), 2752–2765.

Paradis, M. (2003). The bilingual Loch Ness Monster raises its non-asymmetric head again- or, why bother with such cumbersome notions as validity and reliability? Comments on Evans et al (2002). *Brain and Language*, 87, 441-448.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Köpke, M. S. Schmid, M. Keijzer & S. Dosterst (Eds.), *Language Attrition: Theoretical Perspectives* (pp. 9-37). Amsterdam: John Benjamins.

Penfield, W. (1965). Conditioning the uncommitted cortex for language learning. *Brain*, 88, 787-798.

Penfield, W., & Roberts, L. (1959). *Speech and Brain Mechanisms.* New York: Athenaeum.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology,* 118, 2128–2148.

Prat, C. S. (2011). The brain basis of individual differences in language comprehension abilities. *Language & Linguistics Compass, 5,* 635–649.

Roehm, D., Bornkessel, I., Haider, H., & Schlesewsky, M. (2005). When case meets agreement: Event related potential effects for morphology-based conflict resolution in human language comprehension. *NeuroReport, 16*(8), 875-878.

Rossi, S., Gugler, M., Friederici, A., & 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*(12), 2030–2048.

Royle, P., Drury, J. E., & Steinhauer, K. (2013). ERPs and task effects in the auditory processing of gender agreement and semantics in French. *The Mental Lexicon, 8*(2), 216-244.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness. *Journal of Cognitive Neuroscience 23*, 514–523.

Schmid, M. S. (2010). Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingualism: Language and Cognition*, *13*(01), 1.

Schmid, M. S. (2011). *Language attrition*. Cambridge: Cambridge University Press.

Schmid, M. S., & Fägersten, K. B. (2010). Disfluency Markers in L1 Attrition: Disfluency Markers in L1 Attrition. *Language Learning*, *60*(4), 753–791.

Schmid, M. S., & Köpke, B. (2011). L'attrition de la première langue en tant que phénomène psycholinguistique. *Language, Interaction and Acquisition*: *Special Issue on L1 attrition*, *2*(2), 197-220.

Silva-Pereyra, J. F., & Carreiras M. (2007). An ERP study of agreement features in Spanish. *Brain Research, 1185*(14), 201-211.

Squires, N., Squires, K., & Hillyard, S. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology,* 38(4), 387–401.

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., Drury, J. E., Portner, P., Walenski, M., & Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia, 48*(6), 1525-1542.

Steinhauer, K., Mecklinger, A., Friederici, A.D., & Meyer, M. (1997). Probability and strategy: An event-related potential study of processing syntactic anomalies. *Zeitschrift fuer Experimentelle Psychologie, 2*, 305-331.

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.

Tanner, D., (in press). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex*, 1-7.

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*(02), 277–293.

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(02), 367–382.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56,* 289-301.

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.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*(1), 37–69.

van Herten, M., Kolk, H. H., Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*, 241-255.

Vespignani, F., Molinaro, N., & Job, R. (in prep). ERP study of parsing strategies during repair. In Tommasi L, and Di Matteo R (Eds), *Cross-linguistic Perspectives in Language Processing*. Amsterdam: John Benjamins Publishing Company.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, *37*(1), 159–170.

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.

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain Responses before and after Intensive Second Language Learning: Proficiency Based Changes and First Language Background Effects in Adult Learners. *PLoS ONE*, *7*(12), 52318.

4. Bridge 1

From morphosyntax to lexical-semantics

Study 1 explored L1-morphosyntactic processing and showed that L1-attriters differed from non-attriting native-speakers in their repair/re-analysis of number-agreement violations in L1-Italian sentences. Our findings of group differences in the amplitude, scalp distribution and duration of LAN/N400 and P600 components suggested overall that attriters were less likely than non-attriting controls to explore structural solutions to an agreement problem during online sentence processing. Examining ERP correlates of agreement processing at two separate target positions within each sentence revealed that attriters seemed to treat subject-verb agreement errors as more salient grammatical violations (due to influence of English word-order) and to engage in shallower repair mechanisms over the course of the sentence than Italian monolinguals. Proficiency scores on L1 tasks were also shown to be a key factor in modulating N400 and early P600 responses, even among two groups of adult native-speakers. However, attrition-effects were not limited to L1-proficiency variations, Attriters as a group did not elicit a significant P600 lasting beyond 1000 ms. This late P600 effect – thought to reflect elaborated processes of sentence-repair – was instead modulated by amount of L1-exposure in attriters. The idea of less efficient processing was supported by the finding of slower reaction times in attriters than non-attriting controls. However, no other behavioral differences were detected on the acceptability judgment performed at the end of each sentence. Overall, Study 1 provided the first ERP evidence of attrition effects in L1 morphosyntax, and suggests that behavioral studies claiming that morphosyntax is resistant to attrition may have tested morphosyntactic properties that were too undemanding to pose a challenge for attriters and to reveal group/proficiency differences.

Study 2 aimed to extend these results to the domain of lexical-semantic processing, and to determine whether anecdotal reports of difficulties with lexical access in L1-attriters can be validated in online comprehension at the neurocognitive level. Contrary to Study 1, we did not exclusively focus on *L1*-processing in Study 2, but on comparing L1 *and* L2 processing by additionally including the group of late L2-Italian learners. Our main goal was to explore whether ERP profiles of lexical-semantic processing were largely determined by Italian proficiency-level, irrespective of whether Italian was individuals' L1 or L2. Such a finding would echo Study 1 in highlighting the importance of proficiency level in determining the brain's response patterns to language, and would also point towards a proficiency-based continuum between L1 acquisition, L1 attrition and L2 acquisition.

As in Study 1, we examined modulations of ERP components reflecting distinct processing stages (N400 and P600) and discuss the functional significance of these ERP effects relative to the literature on language processing, as well as to specific questions that have been raised in studies of L2 processing. Finally, as in the previous study, we selected experimental stimuli that we expected to be more difficult to process in L2 learners and/or L1 attriters, namely sentences containing a target noun that could easily be confused for its orthographic neighbor (*cappello* (hat) vs. *cappella* (chapel)).

Thus, both studies examine the neurocorrelates of L1 attrition and the factors modulating attrition effects, as well as the impact of proficiency level on the nativelikeness of ERP responses. However, Study 2 additionally makes the link between L1 processing (and L1 attrition) and L2 processing by testing bilingual speakers with the same language pair (English, Italian) but opposite age-of-acquisition profiles (i.e., L1/L2 status reversed).

5. MANUSCRIPT 2

Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late second-language acquisition

Kristina Kasparian[1,2] and Karsten Steinhauer[1,2]

[1] School of Communication Sciences & Disorders, McGill University, Montreal, Canada
[2] Centre for Research on Brain, Language and Music (CRBLM), Montreal, Canada

This manuscript has been submitted to Neuropsychologia for publication and is currently under review

ABSTRACT

First-language (L1) attrition is a unique socio-linguistic circumstance where second-language (L2) learning coincides with changes in exposure and use of the native-L1. "Attriters" often report experiencing a decline in automaticity or proficiency in their L1 after a prolonged period in the L2 environment, while their L2 proficiency continues to strengthen. Investigating the neurocognitive correlates of attrition alongside those of late L2 acquisition addresses the question of whether the brain mechanisms underlying both L1 and L2 processing are strongly determined by proficiency, irrespective of whether the language was acquired from birth or in adulthood. Using event-related-potentials (ERPs), we examined lexical-semantic processing in Italian L1-attriters, compared to adult Italian L2-learners and to Italian monolingual native-speakers. We contrasted the processing of classical lexical-semantic violations (*Mismatch* condition) with sentences that were equally semantically implausible but arguably trickier, as the target-noun was "swapped" with an orthographic neighbor that differed only in its final vowel and gender-marking morpheme (e.g., *cappello* (hat) vs. *cappella* (chapel)). Our aim was to determine whether sentences with such "confusable nouns" (*Swap* condition) would be processed as a semantically-correct sentences by late L2-learners and L1-attriters, especially for those individuals with lower Italian proficiency levels. We found that lower-proficiency Italian speakers did not elicit significant N400 effects for *Swap* violations relative to correct sentences, regardless of whether Italian was the L1 or the L2. Crucially, N400 response profiles followed a continuum of "nativelikeness" predicted by Italian proficiency scores – high-proficiency attriters and high-proficiency Italian-learners were indistinguishable from native-controls, whereas attriters and L2 learners in the lower-proficiency range showed significantly reduced N400 effects for "*Swap*" errors. Importantly, attriters and late L2 learners did not differ in their N400 responses when they belonged to the same proficiency subgroup. Attriters also showed an enhanced P600 response to both kinds of lexical-semantic anomalies, which we discuss as reflecting increased conflict-monitoring and conscious "second thought" processes specifically in attriters. Our findings provide some of the first ERP evidence of attrition effects, and are compatible with accounts of ongoing neuroplasticity for language in adulthood. Proficiency, rather than age-of-acquisition, seems to be the key factor in modulating neurocognitive responses, not only within L2 learners but also in L1-attriters.

Keywords: *neuroplasticity, lexical-semantic processing, first-language attrition, second-language processing, proficiency, age-of-acquisition, continuum*

1. INTRODUCTION

First-language (L1) *"attriters"* are a unique group of bilinguals. In contrast to most late second-language (L2) learners who maintain predominant use of their native-language and who therefore consider their L1 to be their "dominant" or default language, L1-attriters are individuals for whom advancing L2 proficiency comes at a direct cost to their L1. Due to socio-linguistic situations where the individual's contact with the L1-community becomes limited or severed (e.g., immigration to an environment where the L1 is not a majority-language), attriters gradually experience a shift from the L1 to the predominantly-used L2. The phenomenon of *L1-attrition* has been described as a non-pathological change in the L1 where individuals deviate from native-speakers in their use of the language – be it in terms of their fluency, vocabulary or use of certain grammatical structures (see Köpke, 2002; Köpke & Schmid, 2004; Schmid, 2011 for reviews). Importantly, these changes occur after the L1 had been previously acquired at the level of an age-appropriate native-speaker. In this vein, attrition is a direct consequence of bilingualism, given that attriters experience a change in automaticity and in native-like proficiency in their L1, while their proficiency in the late-learned L2 continues to strengthen.

Characteristics of attriters' L1 have been extensively explored in behavioral research, revealing aspects of non-native-like linguistic performance in a number of domains but particularly in vocabulary (e.g., de Bot, 1996; Köpke, 1999; Montrul, 2008; Opitz, 2011). Even when attriters anecdotally report having noticed changes in their L1 or having such changes pointed out to them by other native-speakers (e.g., friends or family), the most common complaints are difficulties in the vocabulary domain. In support of these self-reports, studies using *production* tasks such as picture-naming, verbal semantic fluency and free-speech have provided evidence that attriters are slower to access lexical items and may have smaller or less-diverse L1 vocabularies (De Bot, 1996; Köpke, 1999; Linck et al., 2009; Schmid & Jarvis, 2014; Schmid & Keijzer, 2009; Waas, 1996; Yagmur, 1997; Yilmaz & Schmid, 2012). Others have found that L1-attriters experience competition or transfer of lexical items from the L2 to the L1 – processes described as lexical borrowing or semantic-intrusions when words are cross-linguistically similar (Pavlenko, 2000). Some theories have accounted for these difficulties by positing breakdowns at the level of *stored* lexical representations in memory (see Ecke, 2004 for theories of forgetting), while others have claimed that attriters are rusty in *accessing* lexical items during production and that the difficulties have more to do with retrieval than with representation (e.g., Green, 1986; Paradis, 1989; 1997). However, it remains an open question whether lexical difficulties in attriters' L1 also occur during *comprehension*, and whether these difficulties are largely due to crosslinguistic competition or whether they may also occur as *intralinguistic* competition, i.e., *within* the L1.

The current study examines these questions using event-related potentials (ERPs). Although a multitude of studies have investigated the phenomenon of L1-attrition at the *behavioral* level, manifestations of attrition during real-time language processing using more sensitive temporal measures (such as eye-tracking or ERPs) are still relatively unexplored (but see Datta, Obler & Shafer, unpublished dissertation; Schmid, 2013). With their high temporal-resolution, ERPs are particularly useful in exploring the time-course of language processing as it unfolds, and are often shown to be more sensitive to processing differences than offline behavioral tasks that only tap into the end-stage of linguistic processing (e.g., Kasparian, Bourguignon, Drury & Steinhauer, 2010; McLaughlin, Osterhout & Kim, 2004; Steinhauer, White & Drury, 2009; Thierry & Wu, 2007). ERPs are thus likely to reveal more about L1-attrition than behavioral measures alone, and may even reveal a dissociation between attrition effects observed in behavioral performance versus those observed at the neurocognitive level during language processing.

1.1. Late L2 acquisition: Neuroplasticity and proficiency

Lexical access during online processing has been widely explored in ERP studies of late L2 acquisition. For decades, researchers have examined whether late L2 learners deviate from early-learners and/or native-monolingual speakers in how they process single words as well as sentences where a word is either semantically congruent or incongruent with the preceding context (e.g., Hahne, 2001; McLaughlin et al., 2004; Mueller, 2005; Ojima et al., 2005; Weber-Fox & Neville, 1996, to name a few). At the heart of the research exploring whether fundamental differences in language processing exist between late L2-learners and early-learners or native-monolinguals is the notion of a neurobiological "critical-period" for language learning (Lenneberg, 1967; Penfield & Roberts, 1959), after which the brain is believed to lose the plasticity necessary to acquire and process language using the same neurocognitive substrates as those used for the L1, thus resulting in non-native language processing mechanisms (Lenneberg, 1967; Penfield & Roberts, 1959; see also Bley-Vroman, 1989; Ullman, 2001). The corollary of this view is that the L1 has a privileged status, as it was acquired from birth and "hard-wired" in the brain within this maturationally-constrained window, and is therefore expected to remain stable beyond this "critical period" once brain networks for language are no longer plastic (Marchman, 1993; Penfield, 1965).

In line with these claims, a number of neuroimaging and neurophysiological studies on L2 acquisition have found that *age-of-acquisition* (AoA) is a critical determinant of the degree of native-like-ness of L2 processing (e.g., Hahne & Friederici, 2001; Kim et al., 1997; Weber-Fox & Neville, 1996). Some have argued that native-like lexical-semantic processing is more readily attainable in late L2 learners compared to morphosyntactic processing (e.g., as discussed in Hahne & Friederici, 2001; Newport, Bavelier & Neville, 2001; Wartenburger et al., 2003; Weber-Fox & Neville, 1996), and this differential impact of AoA has been explained in terms of a longer window of neuroplasticity for certain domains (e.g., lexical-semantics) compared to others (e.g., phonology, morphosyntax).

Increasingly, however, a counterargument to critical-period claims is gaining favor – that the attained *proficiency* level plays a crucial role in modulating the brain's responses to the L2 and that low proficiency levels rather than a late AoA are what result in deviations from native-like processing strategies. While there is no doubt that an L2 is harder to acquire than an L1, the fact that the late-acquired L2 is typically the language that speakers are less exposed to, less experienced with and less proficient in, it has been a challenge in the literature to disentangle the effects of these various factors from the effect of AoA. Advocates of this view have shown evidence that late learners with high levels of L2 proficiency do exist and do show neurocognitive response patterns that are indistinguishable from those of native-speakers, even if the L2 was acquired in adulthood (Bowden et al., 2014; Friederici, Steinhauer & Pfeifer, 2002; Morgan-Short et al., 2012; Osterhout et al., 2006; Rossi et al., 2006; Steinhauer, White & Drury, 2009). Research has shown that lexical-semantic processing patterns are also modulated by L2 proficiency-level, with beginning learners deviating most in their processing of words and sentences, compared to highly-proficient learners (e.g., McLaughlin et al., 2004; Newman et al., 2012; Osterhout et al., 2006).

To date, it is still debated whether there are indeed maturational limits on neuroplasticity that impede native-like language processing in adulthood, and the question of which factors (e.g., AoA, proficiency, exposure, cross-linguistic differences, etc.) are most crucial in determining the brain's responses to language is unresolved.

1.2. Theoretical contributions of L1-attrition to the debate on brain plasticity for language

First-language attrition is a phenomenon that offers a fresh perspective on a half-century-old problem. Not only are these speakers highly-proficient and predominantly-exposed to the *L2* they became immersed in during adulthood – thus offering a better test of whether L2-processing is native-like at high levels of L2 proficiency/exposure despite late AoA – but attriters additionally allow us to study whether one's *L1* continues to be native-like in circumstances where exposure and eventually proficiency has shifted from the L1 to the L2. If, in adulthood, attriters experience changes in the real-time language-processing mechanisms underlying their L1, such a finding would advocate for ongoing neuroplasticity for language in adulthood, and would suggest that one's L1 is not as stable and hard-wired as is generally assumed.

Investigating the neurocorrelates of L1 attrition alongside those of late L2 acquisition also allows us to determine whether both L1 *and* L2 processing in bilingual speakers are strongly determined by proficiency level, irrespective of whether the language in question is the L1 or L2 (i.e., regardless of AoA). A key question is whether attrition and acquisition may lie on the same *continuum*, with proficiency level modulating the brain's response patterns to language in both groups of bilinguals. Evaluating the idea of such a continuum based on proficiency-level was the chief goal of our current study.

It may follow from this notion of a continuum that those linguistic features that are difficult for L2 learners to process at a native-like level until they reach high-levels of proficiency may also be those features which are affected by proficiency/exposure changes in L1 attrition. If neurobiological constraints on plasticity are not supported and we have no reason to expect fundamental differences between L1 and L2 processing other than those differences driven by proficiency or frequency of exposure, then it is conceivable that developmental stages of L1 attrition and L2 acquisition may parallel each other, but in reverse (for a similar argument of L2 attrition, see Pitkänen, Tanner & Osterhout, unpublished dissertation). This continuum may particularly be the case in the domain of lexical-semantic processing, which has been argued to be less constrained by neuroplasticity than morphosyntax or phonology. Although longitudinal studies testing both lexical-semantic and morphosyntactic processing would be best suited to explore these interesting questions, the present study is one of the first attempts to systematically explore proficiency effects on ERP responses in both L1-attriters and L2 learners, with the goal of determining whether such a proficiency-based continuum of processing-patterns may be observed, regardless of AoA.

1.3. The present study

Using ERPs, we examined online lexical-semantic processing of Italian sentences in two groups of adult bilingual speakers of the same language pair but opposite AoA profiles: (1) Italian-English L1 attriters, highly-proficient in English and reporting a shift in dominance and changes in proficiency in their native-Italian; and (2) English-Italian late L2 learners, highly-advanced in their Italian level but still dominant in their native-English. Attriters were therefore tested in their L1 (in their dominant-L2 environment), while the late Italian learners were tested in their L2 (in their dominant-L1 environment). The bilingual groups were additionally compared to a group of Italian native-monolinguals still residing in Italy.

1.3.1. Confusable words

Our study tested lexical-semantic access and integration of "confusable words" *within Italian*, rather than assessing the role of crosslinguistic transfer/competition with confusable words between the individuals' two languages, such as "false-friend" homographs (but see Kasparian, Postiglione & Steinhauer, under review). The critical words in our experiment involved minimal pairs that differed in their final vowel (e.g., *cappello* vs. *cappella*) but also in their lexical-semantic meaning (*hat* vs. *chapel* respectively). These words, therefore, were orthographic and phonological neighbors of the words that were intended to fit the sentence context. We examined ERP correlates of lexical-semantic processing in response to sentences where the intended target word was swapped with its minimal pair ("*swap*"condition), as well as to sentences where an orthographically unrelated word was erroneously inserted into the sentence context ("*mismatch*" condition), compared to semantically-correct sentences (see **Table 1**).

Table 1 about here

While the "*mismatch*" sentences were similar to classical violations tested in ERP paradigms of lexical-semantic processing (e.g. "*He spread the warm bread with socks*"; Kutas & Hillyard, 1980; 1984), the "*swap*" sentences – though equally semantically-implausible – arguably consisted of a less salient type of lexical-semantic violation, given the confusability of the target words. Similar to interlingual "false friends", pairs of words *within Italian* that differ in their final vowel and semantic meaning are notoriously difficult for Italian learners, and are the subject of many cautionary notes in Italian textbooks and online language blogs. These lexical items are perhaps additionally difficult to master due to their interface with gender properties – a word-final vowel change in Italian typically corresponds to a change in gender, such that one word in the confusable pair is often feminine in gender while the other is masculine. Studies with L2 learners conducted in both visual and auditory domains have shown that word-recognition is slower and/or more difficult when words can be confounded with orthographically/phonologically-similar words within the L2 (Carreiras, Perea & Grainger, 1997; Rüschemeyer, 2005; Rüschemeyer, Nojack & Limbach, 2008).

One such study by Rüschemeyer and colleagues (2008) showed that German L2-learners were more likely than native-monolinguals to judge auditorily-presented semantically-anomalous sentences (e.g., *Der Tisch wurde am Nachmittag geangelt = The table was caught (fished) in the afternoon*) as semantically-acceptable if replacing the incongruent noun by its phonological neighbor (e.g., *Fisch = fish*) would have rendered the sentence plausible, compared to anomalous sentences with no confusable noun. In an ERP experiment using a cross-modal lexical-priming paradigm, the authors revealed that the L2-learners (but not native-German monolinguals) showed N400 amplitude differences on a visually-presented prime (e.g. *Lachs = salmon*) between neighbor vs. non-neighbor conditions. The authors concluded that L2 learners "are more likely to be tricked by the presence of a phonological neighbor than are native-speakers" (p. 134) and that they experience interference from semantically-unrelated orthographic/phonological neighbors during L2 word-recognition. These results seemed to be in line with other claims that L2 learners are less efficient in inhibiting competing activation of intralingual competitors (FitzPatrick & Indefrey, 2009; Sebastián-Galles, Echeverria & Bosch, 2005).

An important point, however, is that Rüschemeyer and colleagues only tested behavioral/ERP responses to the two violation conditions without including semantically-correct sentences in their analyses. Without correct control sentences, one cannot determine whether L2 learners' response pattern was a graded one similar to graded N400 effects reported in monolinguals. Reduced N400s have been reported for conditions where the presented target-word overlaps with the correct target-word in semantics (e.g., *The pizza was too hot to eat < *drink < **cry*; Kutas & Hillyard, 1984; see also Federmeier & Kutas, 1999; Federmeier et al., 2007; Wlotko & Federmeier, 2015) or in orthographic/phonological form (Laszlo & Federmeier, 2009). The important question is whether the pattern observed in L2 learners reflects a different underlying mechanism than the pattern observed in native-controls, where the N400 for related target-words is reduced relative to unrelated target-words but still large relative to correct control words. The crucial analysis would therefore consist of the comparison between each violation condition relative to the correct control condition, in order to argue that L2 learners (unlike monolinguals) did not distinguish between sentences containing an incongruent orthographic/phonological neighbor and semantically-correct sentences.

Moreover, the orthographic neighbors tested in studies such as those by Rüschemeyer and colleagues (2008) and Laszlo and Federmeier (2009) were words that differed in their initial phoneme (e.g. *Tisch/Fisch* or *wish/dish* respectively), which may potentially make them less confusable than Italian nouns that are identical until their final vowel and gender marker (e.g., see Cohort model, Marslen-Wilson, 1987). Arguably, the point of disambiguation being a word-*final* phoneme and an inflectional morpheme may make the distinction between these items less salient in the input and may add an additional dimension of confusability – not only would learners have to tell the nearly-identical forms apart, but they would have to learn that one concept is associated with the feminine noun, and the other with the masculine noun. These inherent difficulties may result in less robust form-concept mappings in the lexicon, at least at lower levels of proficiency. Thus, and perhaps more so than for words that differ in their word-*initial* phoneme as in the previous studies, it is possible that effects of orthographic/phonological neighbors are not (only) due to inefficiency at inhibiting similar but semantically-inappropriate forms that are co-activated in parallel, but the problem may potentially stem from blurry or under-specified lexical representations where the two distinct concepts are not robustly associated with one form/gender. Investigating lexical items that overlap except for the word-final phoneme / gender morpheme may add an interesting dimension to previous investigations.

Lastly, the effect of L2 proficiency-level on response patterns was not systematically examined in the studies cited above, and only self-reported ratings of proficiency and language exposure were obtained as background measures.

1.3.2. P600 effects in lexical-semantic processing: integration, conflict monitoring and re-analysis

Although ERP investigations of lexical-semantic processing have mainly focused on modulations of the N400 component (Kutas & Hillyard, 1980; 1984; Kutas & Federmeier, 2011 for a review), a systematic review by Van Petten and Luka (2012) identified a number of studies that revealed post-N400 parietal positivities consistent with a P600 effect (but see review for studies reporting *frontal* positivities). It has been shown that the P600 is not merely elicited in response to sentences that are *syntactically* anomalous or dispreferred (Friederici et al., 1996; Hagoort et al., 1993; Kaan & Swaab, 2003; Osterhout & Holcomb, 1992; 1994; Phillips et al., 2005), but also when semantically-based predictions in sentence or discourse contexts are disconfirmed (Bokhari et al., in prep; DeLong et al., 2011; Federmeier et al., 2007; Otten & Van Berkum, 2008; Steinhauer et al., 2010). Van Petten and Luka describe the P600 as a reflection of a "processing cost" associated with having to review the preceding context to diagnose the problem and repair it.

A related and perhaps even more general explanation of P600 effects elicited in response to lexical-semantic violations is that the P600 reflects an underlying "conflict monitoring" process (e.g., Kolk et al., 2003; Kolk & Chwilla, 2007; Van Herten et al., 2005; 2006; Van de Meerendonk et al., 2009; Vissers, Chwilla & Kolk, 2006; Vissers et al., 2008). A monitoring process involves double-checking the input for possible processing errors once an anomaly has been detected, as the participant asks him/herself, "Did I read that correctly?" (Van de Meerendonk et al., 2009; p. 69). Crucially, all forms of input are argued to be checked, including orthographic and phonological features. According to Van de Meerendonk and colleagues, the presence of the P600 depends on the severity of the conflict during lexical integration; a mild conflict leads to integration difficulties that are resolved without having to re-analyze the input, thus eliciting an N400 effect in the absence of a subsequent P600, whereas strong conflicts trigger a re-analysis process, thereby eliciting a P600.

Although Van de Meerendonk and colleagues define a strong conflict as arising from "deeply implausible" sentences, it is not clear whether, in their view, a conflict on a *non-semantic* level such as orthography or phonology (in addition to conflict on a semantic-level) might modulate P600 amplitudes among semantically-implausible sentences. In line with this possibility, Laszlo and Federmeier (2009) found an enhanced late-positivity (LPC: late positive complex) in response to semantically-incongruent sentences that ended with the orthographic neighbor of the predicted completion (e.g., *The genie was ready to grant his third and final dish*), and argued that the larger positivity was due to the explicit recognition of the orthographic similarity with the expected word (*wish*). In addition to examining potential P600 differences between semantically-anomalous sentences where the target-word is either an orthographic/phonological neighbor or an entirely unrelated word, it was of interest to determine whether L1-attriters and/or late L2-learners show processing differences from native-monolingual speakers at the *re-analysis* stage, depending on the degree of the perceived conflict.

1.3.3. Research questions and predictions

Our study examined N400 and P600 effects in response to *Mismatch* and *Swap* anomalies, relative to correct control sentences (refer back to **Table 1**). In terms of condition effects, we expected both lexical-semantic violations to elicit large N400 effects compared to correct sentences in native-Italian monolingual controls. Although the N400 in the *Swap* condition could potentially be reduced in comparison to *Mismatch* violations (i.e., yielding a graded pattern due to the orthographic/phonological similarity between the presented word and the intended (correct) word in *Swap* but not *Mismatch* violations), we nonetheless expected a robust N400 effect relative to the correct condition, demonstrating that the Controls would have detected the semantic incongruity between the target-word and its preceding context, which was instead congruent with the non-presented member of the minimal pair (e.g., *cappella* (*chapel*) instead of *cappello* (*hat*)). It was conceivable that both violation conditions would additionally elicit a P600 effect. Although both *Mismatch* and *Swap* sentences are equally implausible, one could argue that *Swap* violations are less severe (i.e., resulting in a weaker conflict) because simply replacing the target-word with its nearly-identical orthographic neighbor would rescue the sentence. In this view, we would expect *Mismatch* violations to elicit larger P600 responses than *Swap* sentences, as *Mismatch* errors would require more elaborate revision processes in order to arrive at a plausible sentence interpretation. Conversely, if the replacement of a target-noun with an orthographic competitor in *Swap* violations causes enhanced conflict and results in readers having to double-check the ambiguous input for the source of the integration problem (e.g., orthographic/phonological features, gender, sentence-context), then it is conceivable that *Swap* violations would elicit larger P600 effects than *Mismatch* sentences due to the confusability of the nouns. Examining condition differences in the P600 time-window would therefore clarify whether the amplitude of the effect is sensitive only to the degree of semantic implausibility, or to different aspects of conflict (such as orthographic/phonological ambiguity) as well.

In terms of group differences, our main motivation behind our experimental design was to determine whether late L2 learners would be less able to immediately detect *Swap* errors compared to native Italian speakers, due to the increased likelihood of confusing these nearly-identical word forms. Late L2 learners obviously differ from native-speakers in their language-learning experiences and lexical acquisition – not solely in terms of age-of-acquisition but also in amount/type of exposure and context of learning. These "tricky" lexical items are likely to have been encountered not only in naturalistic input but especially in terms of a more explicit way where learners' conscious attention was drawn to these items to aid them in telling them apart. It was therefore conceivable that L2 learners would have more difficulty in rapidly accessing these items and efficiently processing *Swap* sentences as implausible in as automatic a process as in native-speakers. We expected processing patterns to depend on Italian proficiency-level, where individuals with higher Italian proficiency scores would show processing patterns more similar to those observed in native-controls. More specifically, we expected late L2 learners of lower Italian proficiency to show a reduced N400 effect in response to *Swap* violations compared to *Mismatch* violations, with – crucially – little to no difference between *Swap* and *Correct* sentences. In other words, low-proficiency L2 learners were expected to show a different pattern from monolingual native-controls for the comparison between *Swap* vs. *Correct* sentences.

A key question is whether we might expect processing patterns in L1-attriters to resemble those observed in native-monolinguals, or those of late L2 learners. If L1-Italian attriters remain like native-Italian speakers in their processing of confusable lexical items, then we would expect both lexical-semantic violations to elicit large N400 effects (with *Swap* responses potentially being reduced relative to *Mismatch* conditions but still significantly more negative than to *Correct* sentences). If the N400 is reduced in response to *Swap* errors, it is counterintuitive for this to be due to *enhanced* orthographic-overlap effects that underlie the potentially graded N400 pattern in monolingual-controls (*Swap* < *Mismatch*). Instead, we may find that Attriters with lower Italian proficiency scores may also be more inclined to process *Swap* sentences as correct, at least in the N400 time-window.

Related to this point, we also considered whether ERP profiles might fall along a continuum, with proficiency modulating N400 and P600 amplitudes regardless of L1/L2 status. Such a finding would strongly advocate in favor of ongoing neuroplasticity for language in adulthood, and proficiency-based brain responses, even in one's L1. On the other hand, it was conceivable that proficient L2-learners might still differ from L1-attriters in their processing patterns, due to experiential factors that are not characteristic of native-speakers, such as their late age-of-acquisition but also their different type of experience and exposure to lexical items, particularly confusable words. For example, L2 learners are more likely than native-speakers to have had their attention explicitly drawn to confusable words within the language, potentially leading to differences in processing that reflect this differential exposure / acquisition experience.

Finally, we also examined whether proficiency and/or group differences were observed at different *stages* of lexical access/integration, namely on the early detection of the anomaly vs. on later processes reflecting revision and conflict-resolution mechanisms. One possibility was that the groups (or proficiency subgroups) would differ most from native-controls in the *N400* time-window, failing to automatically detect the anomaly or to efficiently inhibit competition from the irrelevant orthographic neighbor, but would converge on their P600 responses, having detected the violation in a more conscious "double-take" or “second-thought” realization of the conflict further downstream (see Kasparian, Postiglione & Steinhauer, under review, for a similar finding in L1-attriters' and L2-learners' processing of Italian-English homographs/cognates). An alternate possibility was that the groups would deviate most from native-controls in the re-analysis stage reflected by the late P600 rather than in the N400 time-window, thus demonstrating processing differences at the level of repairing the input (see Kasparian, Vespignani & Steinhauer, under review).

To date, ours is the first ERP study to examine lexical-semantic processing of confusable and non-confusable words in sentence contexts in L1-attriters and L2-learners, compared to native-monolingual speakers, in an attempt to shed light on the debate surrounding the existence of maturational constraints on brain plasticity and the role of proficiency in shaping the brain's responses to language.

2. METHODS

2.1. Participants

Three groups of participants were tested: (1) Italian-English attriters; (2) English-Italian late L2 learners; and (3) Italian native-monolinguals (see *General Methods*, p. 47-50 for demographic information).

2.2. Behavioral measures

Participants completed background questionnaires and several behavioral measures (see *General Methods*, p. 50-54). Group means are provided in **Table 2**. Attriters scored numerically lower on all four proficiency measures, but did not differ significantly from Controls (*p*s > 0.1). Although our recruitment process targeted L2 learners in the highly-advanced range of Italian proficiency, our group of L2 learners was not matched to Controls or Attriters on their proficiency level. This allowed us to study whether ERP correlates of lexical-semantic processing approach and converge upon native-like profiles with increasing proficiency. Recall that our main aim was to determine whether ERP responses fall along a continuum modulated by Italian proficiency level (but see *Discussion* for questions our data cannot immediately address and for required follow-up studies).

Table 2 about here

Based on their scores on the 3 individual proficiency tests (excluding self-report measures, as monolingual controls rated themselves at ceiling), subgroups of "*high*" and "*low*" proficiency were derived by median split. *High* and *low* proficiency subgroups are described in **Table 3**. *High* vs. *low* proficiency groups (including within native-monolinguals and within L1-Attriters) differed significantly on most proficiency tasks.

Table 3 about here

2.3. Stimuli

Sentence examples of the three experimental conditions are provided in **Table 1**. The target words in this experiment were each constituents of a minimal pair in Italian. We searched for pairs of nouns that differed only in their final vowel (e.g., *cappello* vs. *cappella*) and, consequently, in their semantic meaning (e.g., *hat* vs. *chapel*). We avoided nouns where the vowel change resulted in a difference in pronunciation or length, as well as nouns that overlapped in semantic meaning such that it would be difficult to create a sentence context that highly constrained the meaning of only one word in the minimal pair (e.g. *cero* (candle) / *cera* (wax)). We also avoided vowel-initial words, so that the determiner preceding the target noun would be "*il*" ($\text{the}_{\text{-masc}}$) or "*la*" ($\text{the}_{\text{-fem}}$) across all items. As we did not use indefinite determiners ("*un*", "*uno*", "*una*"), we avoided nouns that could fit only in an indefinite sentence context. These rather stringent selection criteria allowed us to select 48 pairs of nouns, for a total of 96 target words.

All target words were inanimate nouns, either bi- or trisyllabic, ranging from 4 to 8 letters in length, and from low to high lemma frequency (CoLFIS database; Bertinetto et al., 2005). Given the limited number of minimal pairs that survived the strict selection process, and due to the fact that the aim of our study was to scrutinize lexical access in advanced and native-speakers of Italian, we opted against limiting our items to medium- or high-frequency nouns. In order to check for likely frequency effects and group interactions, we coded each target noun as "high" or "low" frequency, based on a median split. Given that grammatical gender in Italian is largely predictable by the word-final vowel ("o" = masculine; "a" or "e" = feminine, barring some exceptions), our target words systematically consisted of 48 masculine nouns and 48 feminine nouns. Masculine and feminine words were identical in length, as we only changed their final vowel (*M* length = 5 letters). There were no significant differences between masculine and feminine nouns on word frequency (Masc: 297.4; Fem: 296.7; $p >$ 0.1).

A set of 96 different sentences were constructed and realized in each of the three conditions. First, target nouns were inserted into sentence contexts that primed the correct meaning of the noun (e.g., literally translated: *To cover his head, the fisherman wears the* <u>*hat*</u> *of wool*). Sentences in the "*Swap*" condition were created by switching nouns within each minimal pair (such that *cappello* (hat) occurred in the *cappella* (chapel) context, and vice versa). Sentences in the "*Mismatch*" condition were created by first pair-wise matching each minimal pair with another minimal pair, resulting in sets of quadruples (e.g., *cappello*/*cappella* paired with *cartello*/*cartella*). Thus, a "*Mismatch*" involved replacing "*cappello*" with "*cartella*", and "*cappella*" with "*cartello*". Note that, for the *Mismatch* condition to be analogous to the *Swap* condition – where swapping the noun within a minimal pair inherently resulted in changing the gender and the preceding determiner – we decided to replace a masculine noun by a feminine one in the *Mismatch* condition as well. Quadruples (i.e., the two sets of minimal pairs) were matched as closely as possible on word length and frequency, but re-assigned in cases where *Swap* and *Mismatch* sentences happened to be semantically plausible (even in a remote, figurative sense). The advantage of our design is that it is completely counterbalanced – every target noun contributed to each main condition, thus ruling out that any differences between conditions were due to target noun characteristics such as frequency or length.

All sentences were eleven words long and were identical in grammatical structure. The target noun was in the 9th position in the sentence, preceded by its determiner, and followed by a prepositional phrase. In all cases, the preceding determiner agreed with the target noun, as our aim was to investigate lexical-semantic processing, not grammatical gender mismatches between determiner-noun. All verbs were conjugated in the present tense and were transitive, such that, in all items, the target noun was the direct object of the verb. Sentences were created by the main author and checked by three Italian native-speakers who did not take part in the study.

Although reporting effects of cloze probability is beyond the scope of the current paper, we obtained and coded cloze probability values for our experimental sentences for future examination, as sentence-constraint is viewed as the most important factor influencing participants' expectations/predictions of upcoming content-words, thus modulating N400 amplitudes (see Federmeier & Kutas, 2011 for a review). We assessed cloze probability with a sentence-completion task administered to 12 Italian native-speakers who did not participate in the study. These individuals were shown each of the 96 sentence contexts (truncated after the determiner) and were asked to complete the sentence with the first word that came to mind. Cloze probability was operationally defined as the percentage of times a word was provided, and the stimuli were coded for high or low cloze probability, as determined by a median split (*Median* = 37%; *Mean* = 43 %).

Three experimental lists were created such that, across lists, each sentence contributed equally to each condition, while no sentence context nor target noun was repeated within an experimental list. Each participant was presented with each noun within a minimal pair (e.g. *cappello*/*cappella*) only once, and in different conditions (i.e., different preceding contexts). Importantly, nouns belonging to the same quadruple (2 minimal pairs, e.g. *cappello*/*cappella*/*cartello*/*cartella*) were kept as separated as possible during the pseudo-randomization process, and never occurred within less than 8 intervening trials. Each participant also saw 228 filler sentences, which were part of the larger Italian study (testing morphosyntactic number agreement processing and relative clause word-orders) and which will be reported in forthcoming papers (see Kasparian, Vespignani & Steinhauer, 2013a, b, c; 2014a, b). Out of the total of 324 pseudorandomized stimuli (96 experimental and 228 fillers) per participant, 146 sentences (approx. 45%) were acceptable (grammatically and semantically), while 178 were expected to receive a rating of 3 or lower on a five-point rating scale (approx. 55%).

2.4. Procedure

The experimental procedure unfolded as described in the *General Methods* (p. 56-57).

2.5. EEG recording and analysis

For details on EEG recording, see *General Methods* (p. 57). Trials containing artifacts due to blinks, eye-movements and excessive muscle activity were rejected prior to averaging, using a moving-window standard deviation of 30 microvolts. One participant in the attrition group was excluded due to baseline problems. On average, participants contributed 30 artifact-free trials per condition (range: 66-100%), with no differences across conditions (*p*s > 0.1).

ERPs were time-locked to the onset of the target noun with a baseline correction of -500 to 0 ms. Our original baseline of -200-0 ms was adjusted in order to compensate for early differences triggered by the preceding verb in some attriters and L2 learners. Data with the original baseline are available as supplemental materials. Two consecutive and non-overlapping time-windows of interest were examined, based on the ERP components we expected as well as visual inspection of the grand average data for each group: (1) 300-550 (N400) and (2) 650-1000 (P600).

Repeated-measures ANOVAs were performed separately for 4 midline electrodes (Fz, Cz, Pz, Oz) and 12 lateral electrodes (F3/4, C3/4, P3/4 and F7/8, T3/4, T5/6). Global ANOVAs for the midline sites included within-subject factors *Condition* (correct, swap, mismatch) and *Ant-Post* (anterior, central, parietal, occipital). Lateral ANOVAs additionally included factors *Hemisphere* (left, right) and *Laterality* (lateral, medial). For all ANOVAs, *Group* (Controls, Attriters, L2 learners) and *Proficiency* (High, Low) were between-subjects factors (but see Section 3.4). All global ANOVAs which yielded significant interaction including the factor *Condition* were followed up with pair-wise ANOVAs to clarify the nature of the interaction. Where appropriate, Greenhouse-Geisser correction was applied to analyses with more than two levels (e.g., *Ant-Post*). In these cases, the corrected p values but original degrees of freedom are reported. As a default, reported analyses are restricted to the midline only, except in cases where the lateral ANOVAs revealed additional effects.

3. RESULTS

3.1. Acceptability judgments

Acceptability ratings (1-5) for each sentence condition are shown in **Fig 1a**. A repeated-measures ANOVA with within-subjects factor *Condition* (Correct, Swap, Mismatch) and between-subjects factor *Group* (Controls, Attriters, L2 learners) revealed a significant main effect of *Condition* ($F(2,142) = 307.914$, $p < 0.0001$ after Greenhouse-Geisser correction), a significant main effect of *Group* ($F(2,71) = 9.238$, $p < 0.0001$), as well as a significant *Condition* x *Group* interaction ($F(4,142) = 28.303$, $p < 0.0001$). A one-way ANOVA by *Condition* revealed that *Group* differences existed for *Correct* ($F(2,73) = 7.267$, $p < 0.0001$) as well as for *Swap* ($F(2,73) = 13.793$, $p < 0.0001$) and *Mismatch* sentences ($F(2,73) = 16.029$, $p < 0.0001$). Post-hoc Bonferroni-corrected multiple comparisons between groups confirmed that the significant group interaction was due to the Italian L2 learners differing from both native-Controls and Attriters on each of the three conditions (ps < 0.01). Controls and Attriters, however, did not differ significantly on any of the conditions (ps > 0.1).

Figure 1a about here

Both violation conditions were rated lower than the correct control condition (Correct vs. Swap: Condition: $F(1,71) = 225.198$, $p < 0.0001$; Correct vs. Mismatch: $F(1,71) = 451.055$, $p < 0.0001$). Differences in ratings between *Mismatch* vs. *Swap* sentences ($F(1,71) = 135.086$, $p < 0.0001$) were not significantly affected by *Group* ($p > 0.1$). It remains to be seen whether online measures such as ERPs would reveal group differences in the real-time processing of *Swap* sentences relative to *Mismatch* violations, contrary to offline responses provided at the end of the sentence.

Group differences were driven by the L2 learners who seemed more willing to accept semantically-anomalous sentences as correct, either because they did not perceive the target-word errors, or because they were in a syntactic "processing mode", giving higher acceptability ratings overall when errors involved vocabulary items rather than grammar (i.e., in our other sub-experiments). Although the present study does not enable us to disentangle the two possibilities nor to systematically verify the processing mode that participants were in, comparing L2 learners' ratings across the different Italian sub-experiments that were part of the larger study confirmed that syntactic anomalies did receive lower ratings than lexical-semantic anomalies. While L2 learners' average rating in response to lexical-semantic violations was 3.5 on 5, violations of number agreement received a significantly lower average rating of 2.88 ($p < 0.0001$) and infrequent/dispreferred (but grammatical) relative-clause word orders received a lower rating of 3.23 ($p = 0.09$). Thus, it could be conceivable that the L2 learners in our study were more strict with grammatical errors overall than with vocabulary errors.

Correlational analyses indicated that Italian proficiency level was positively correlated with participants' ratings of correct control sentences (Overall-proficiency: $r = 0.391$, $p < 0.0001$: C-test: $r = 0.482$, $p < 0.0001$; Error test: $r = 0.557$, $p < 0.0001$; Semantic fluency: $r = 0.44$, $p < 0.0001$). Proficiency was negatively-correlated with ratings of both *Swap* (C-test: $r = -0.357$, $p < 0.001$; Error test: $r = -0.408$, $p < 0.0001$; Semantic fluency: $r = - 0.363$, $p < 0.001$) and *Mismatch* sentences (C-test: $r = -0.457$, $p < 0.001$; Error test: $r = -0.509$, $p < 0.0001$; Semantic fluency: $r = -0.450$, $p < 0.0001$), such that individuals with higher-proficiency scores gave lower acceptability ratings for violation conditions.

3.2. Reaction times

Reaction times between the onset of the prompt and participants' button-press are depicted in **Fig 1b**. A repeated-measures ANOVA with within-subjects factor *Condition* (Correct, Swap, Mismatch) and between-subjects factor *Group* (Controls, Attriters, L2 learners) revealed a significant main effect of *Condition* ($F(2,142) = 13.888$, $p < 0.0001$ after Greenhouse-Geisser correction), as well as a significant main effect of *Group* ($F(2,71) = 7.264$, $p < 0.001$). Follow-up Bonferroni-corrected comparisons indicated that L2 learners took significantly longer to respond than native-Controls ($p < 0.005$) but not compared to Attriters ($p > 0.1$). In turn, native-Controls and Attriters only differed marginally in their reaction times ($p = 0.09$). However, the interaction between *Condition* x *Group* was only marginal ($F(4,142) = 2.606$, $p = 0.06$), suggesting that the bilingual subgroups had longer RTs overall.

Figure 1b about here

Post-hoc Bonferroni-corrected comparisons were performed to clarify the nature of the *Condition* main effect on RTs. *Mismatch* sentences were responded to significantly slower than those to *Correct* sentences ($F(1,71) = 14.534$, $p < 0.0001$) and to *Swap* sentences ($F(1,71) = 32.045$, $p < 0.0001$). Reaction times to *Swap* sentences were not significantly different from *Correct* sentences ($F(1,71) = 0.181$, $p > 0.1$).

Correlational analyses indicated that the only proficiency measure to significantly correlate with reaction times for all three conditions was semantic fluency, with lower scores being associated with longer reaction times for *Correct* ($r = - 0.285$, $p < 0.0001$), *Swap* ($r = - 0.273$, $p < 0.0001$), and *Mismatch* conditions ($r = - 0.295$, $p < 0.0001$). The Error-detection test was only marginally correlated with reaction times for *Correct* ($r = - 0.186$, $p = 0.06$) and *Swap* ($r = - 0.168$, $p = 0.08$) sentences, but was significantly correlated with reaction times to *Mismatch* violations ($r = - 0.217$, $p < 0.05$), with higher scores associated with shorter reaction times.

3.3. ERPs elicited on the target noun

Grand average ERP waveforms elicited in native-speaker Controls are presented in **Fig. 2** Controls showed a broadly-distributed N400 in response to both *Swap* and *Mismatch* violations, followed by a posterior P600 effect that seemed more pronounced in the *Swap* than the *Mismatch* condition.

Figure 2 about here

High-proficiency Attriters (**Fig. 3a**) showed a similar, broadly-distributed N400 in response to *Mismatch* and *Swap* violations, but the P600 effect that followed was larger and less focal than the posterior P600 observed in Controls. Low-proficiency Attriters (**Fig. 3b**) showed a broad N400 in response to *Mismatch* sentences but a visibly reduced N400 for the *Swap* condition, compared to High-Proficiency Attriters and Controls. A P600 effect was elicited in response to both kinds of lexical-semantic violations, but this P600 seemed smaller and more broadly-distributed (i.e., less posterior) in Low- than in High-Proficiency Attriters.

Figures 3a and 3b about here

High-proficiency L2 learners (**Fig. 4a**) showed an N400 effect in response to both semantically anomalous sentences. On the other hand, low-proficiency L2 learners (**Fig 4b**) showed a somewhat posterior N400 only in the *Mismatch* condition, whereas the *Swap* condition overlapped with the *Correct* condition in the N400 time-window. Overall, neither subgroup of L2 learners elicited a P600 effect in either violation condition. Instead, the High Proficiency L2 learners elicited a negativity in the *Mismatch* condition, and the Low Proficiency L2 learners showed an apparent positivity in the *Swap* condition, but the latter turned out to be non-significant and was primarily driven by two participants.

Figures 4a and 4b about here

Figure 5 illustrates the continuum-like pattern of ERP responses to *Swap* (**Fig. 5a**) and *Mismatch* (**Fig. 5b**) sentences by proficiency subgroup. In sum, visual inspection of ERP patterns suggests that proficiency level modulated the amplitude and scalp distribution of the N400 (300-500 ms), and more so for *Swap* than for *Mismatch* violations. The P600 effect (650-1000 ms) appeared to be affected by *Condition* (Swap > Mismatch), by *Proficiency* (High > Low) but also by *Group* (Attriters > Controls > L2 learners), in terms of its amplitude and scalp distribution.

Figures 5a and 5b about here

3.3.1. N400 between 300-550 ms

The global ANOVA in the 300-550 ms time-window for midline electrodes revealed a significant main effect of *Condition* ($F(2,134) = 25.41$, $p < 0.0001$) and a significant *Condition* x *Ant-Post* interaction ($F(6,402) = 3.98$, $p < 0.001$). Interactions with *Group* did not reach significance ($p > 0.1$), indicating that the N400 was shared between Controls, Attriters and L2 learners. However, a significant interaction between *Condition* x *Proficiency* was found ($F(2,134) = 3.81$, $p < 0.05$). No significant interactions involved *both* between-subject factors *Proficiency* x *Group*, suggesting that the impact of both factors was independent.

Follow-up analyses by *Condition* pairs were then performed to elucidate the nature of the interactions. The comparison between *Mismatch* and *Correct* conditions yielded a significant main effect of *Condition* ($F(1,67) = 58.02$, $p < 0.0001$) as well as a significant *Condition* x *Ant-Post* interaction ($F(3,201) = 5.00$, $p < 0.05$). The interaction between *Condition* x *Proficiency* also reached significance ($F(1,67) = 6.49$, $p < 0.05$), as did the three-way interaction between *Condition* x *Ant-Post* x *Proficiency* ($F(3,201) = 3.21$, $p < 0.05$). When followed-up by proficiency subgroup, results indicated that high proficiency individuals showed a main effect of *Condition* ($F(1,36) = 49.71$, $p < 0.0001$) qualified by a significant *Condition x Ant-Post* interaction ($F(3,108) = 5.46$, $p < 0.05$), reflecting the maximal amplitude of the N400 among midline electrodes at Cz and Pz in high proficiency individuals. Conversely, low proficiency individuals only showed a significant main effect of *Condition* ($F(1,35) = 17.05$, $p < 0.05$), suggesting that the N400 effect for *Mismatch* sentences was more evenly distributed in low proficiency compared to high proficiency speakers.

The comparison between *Swap* and *Correct* conditions also revealed a main effect of *Condition* ($F(1,67) = 30.41$, $p < 0.0001$) and a significant interaction between *Condition* x *Proficiency* ($F(1,67) = 7.60$, $p < 0.01$), which when followed-up by proficiency subgroup indicated that only high proficiency individuals showed a significant N400 effect in the *Swap* condition ($F(1,36) = 32.86$, $p < 0.0001$), while for low proficiency individuals, *Swap* sentences were not significantly different from *Correct* control sentences in the N400 time-window ($p > 0.1$).

The comparison between *Mismatch* and *Swap* violation conditions yielded a marginal main effect of *Condition* ($F(1,67) = 2.92$, $p = 0.09$) and a significant interaction between *Condition* x *Ant-Post* ($F(3,201) = 6.19$, $p < 0.0005$), reflecting a graded pattern in N400 responses at Pz ($F(1,67) = 13.15$, $p < 0.05$) and Oz ($F(1,67) = 4.69$, $p < 0.05$) but not Fz or Cz (ps > 0.1). Proficiency level, however, did not significantly modulate the difference between the two violation conditions in the N400 time-window (ps > 0.1).

Finally, we compared the difference waves between *Mismatch-Swap* and *Swap-Correct* as two levels of *Condition* in an additional ANOVA and expected to find a significant *Condition* x *Proficiency* interaction, indicating that, for High-proficiency individuals, the major N400 difference lies in the *Swap-Correct* condition whereas *Mismatch-Swap* does not add much more (as both violations elicited large N400 effects in High-proficiency individuals). Conversely, in Low-proficiency subgroups, we would expect an N400 difference only for the *Mismatch-Swap* condition and not for *Swap-Correct*. Our prediction was confirmed, although the effect was marginal ($F(1,71) = 3.47$, $p = 0.06$).

3.3.2. P600 between 650-1000 ms

On the midline in the P600 window, a significant main effect of *Condition* ($F(2,134) = 4.92$, $p < 0.01$) was qualified by a significant interaction between *Condition* x *Ant-Post* ($F(6,402) = 4.67$, $p < 0.005$), reflecting the largely posterior prominence of the positivity. Contrary to the preceding N400, the presence and distribution of the P600 was significantly modulated both by *Proficiency* (<u>*Condition* x *Ant-Post* x *Proficiency*</u>: $F(6,402) = 2.48$, $p < 0.05$) and by *Group* (<u>*Condition* x *Group*</u>: $F(2,134) =$

3.14, $p < 0.05$; *Condition* x *Ant-Post* x *Group*: $F(12,402) = 2.35$, $p < 0.05$). No interactions involving both *Proficiency* x *Group* were close to significance (*p*s > 0.1), indicating that the effects of the two factors were independent.

The comparison between *Mismatch* and *Correct* sentences yielded a significant *Condition* x *Ant-Post* interaction ($F(3,201) = 5.18$, $p < 0.005$). The amplitude and the scalp distribution of the P600 effect in response to *Mismatch* sentences was influenced by *Group* (*Condition* x *Group*: $F(2,67) = 5.10$, $p < 0.01$; *Condition* x *Ant-Post* x *Group*: $F(6,201) = 3.21$, $p < 0.005$). Follow-up analyses within each *Group* demonstrated a main effect of *Condition* in Attriters ($F(1,22) = 5.30$, $p < 0.05$), reflecting the broad positivity elicited by *Mismatch* sentences, and a main effect of *Condition* in L2 learners ($F(1,19) = 5.11$, $p < 0.05$), reflecting the unexpected negativity observed in this time-window. Controls showed only a significant *Condition* x *Ant-Post* interaction ($F(3,87) = 10.18$, $p < 0.0005$), confirming the more posterior focus of the P600 in Controls (Fz: $p > 0.1$; Cz: $p > 0.1$; Pz: $p < 0.05$; Oz: $p < 0.005$) compared to Attriters (Fz: $p > 0.1$; Cz: $p < 0.05$; Pz: $p < 0.05$; Oz: $p < 0.05$). In addition to *Group* differences, the P600 effect for *Mismatch* vs. *Correct* sentences was also significantly modulated by *Proficiency* (*Condition* x *Ant-Post* x *Proficiency*: $F(3,201) = 3.82$, $p < 0.05$), with high-proficiency individuals eliciting a robust and more posterior P600 (*Condition* x *Ant-Post*: $F(3,108) = 6.96$, $p < 0.005$; Pz and Oz: $p < 0.05$), compared to low proficiency individuals (*Condition* x *Ant-Post*: $F(3,105) = 2.94$, $p = 0.06$).

Next, the comparison between *Swap* and *Correct* sentences revealed a *Condition* main effect ($F(1,67) = 9.10$, $p < 0.005$) qualified by a *Condition* x *Ant-Post* interaction ($F(3,201) = 8.97$, $p < 0.0001$), which in turn was significantly modulated both by *Group* (*Condition* x *Ant-Post* x *Group*: $F(6,201) = 2.97$, $p < 0.05$) and by *Proficiency* (*Condition* x *Ant-Post* x *Proficiency*: $F(3,201) = 3.11$, $p < 0.03$). Follow-up analyses by *Group* demonstrated that *Swap* violations elicited a P600 effect in Controls (*Condition*: $F(1,29) = 3.28$, $p = 0.08$; *Condition x Ant-Post*: $F(3,87) = 9.98$, $p < 0.0001$) and Attriters (*Condition*: $F(1,22) = 9.99$, $p < 0.005$; *Condition x Ant-Post*: $F(3,66) = 6.80$, $p < 0.01$), but that the P600 was less focal/posterior in Attriters (Fz: $p = 0.1$; Cz: $p < 0.01$; Pz: $p < 0.005$; Oz: $p < 0.005$) than in Controls (Fz: $p > 0.1$; Cz: $p = 0.07$; Pz: $p < 0.005$; Oz: $p < 0.005$). As a group, L2 learners did not elicit a P600 for *Swap* violations in this time-window ($p > 0.1$). Follow-up analyses by *Proficiency* indicated a significant, posterior P600 in high proficiency individuals (*Condition* x *Ant-Post*: $F(3,108) = 7.97$, $p < 0.001$), and a broadly-distributed P600 in low proficiency individuals (*Condition*: $F(1,35) = 8.69$, $p < 0.001$).

Finally, comparing *Mismatch* and *Swap* violations revealed a significant main effect of *Condition* ($F(1,67) = 5.09$, $p < 0.05$) that was shared across groups, therefore confirming the trend observed during visual inspection that *Swap* violations generally elicited a larger positivity than *Mismatch* violations in the time-window between 650-1000 ms. However, this difference was not modulated by *Proficiency* level ($p > 0.1$).

3.3.3. P600 differences between Controls and Attriters

Given that Controls and Attriters were the only two subgroups to show a clear P600 effect in response to lexical-semantic anomalies in our study, we conducted a lateral ANOVA with only those two groups (keeping *Proficiency* as a second between-subjects factor), in order to assess whether both groups of native-Italian speakers differed significantly in the amplitude and/or scalp distribution of the P600. A *Condition* x *Group* interaction ($F(4,196) = 13.64$, $p < 0.0001$) revealed that Controls and Attriters differed significantly in their P600 responses, both for *Mismatch* vs. *Correct* (*Condition* x *Group:* $F(1,49) = 5.31$, $p < 0.05$) and *Swap* vs. *Correct* comparisons (*Condition* x *Group:* $F(1,49) = 4.38$, $p < 0.05$). A significant five-way interaction between *Condition* x *Laterality* x *Ant-Post* x *Proficiency* x *Group* reflected the pattern that Attriters elicited more broadly-distributed P600 effects than Controls, and that amplitudes were larger in high vs. low proficiency individuals. Thus, the significant *Group* differences found in our ANOVA with all 3 groups were not simply due to L2 learners' lack of a P600 effect, but also to Attriters eliciting larger and less focal/posterior P600 effects (for both violations) compared to non-attriting Italian native-speakers.

3.4. **Further investigations of proficiency subgroups**

As an additional investigation of proficiency effects, we ran two supplementary analyses on the same time-windows of interest with "*Group*" (3 levels) as the only between-subjects factor: (1) all Controls compared to High Proficiency Attriters and High Proficiency L2 learners; (2) all Controls compared to Low Proficiency Attriters and Low Proficiency L2 learners. Our aim was to confirm the findings from the global ANOVAs above (Section 3.3) that native-like-ness of N400 responses depended on proficiency level regardless of group membership (i.e., L1/L2 status), and to rule out that proficiency effects observed in our initial ANOVAs were due to the L2 learners alone.

We expected that the two High groups would be indistinguishable from Controls in the N400 time-window (and would be indistinguishable from one another), whereas the two Low groups would differ significantly from Controls, at least in the Swap condition, though not differing significantly from one another. On the P600, we expected *Group* differences for *Controls vs. Highs* as well as *Controls vs. Lows* comparisons, given that the L2 learners stood out from the other two groups and did not elicit P600 effects, regardless of their proficiency level. The results confirmed our predictions.

3.4.1. Controls vs. High Proficiency Attriters vs. High Proficiency L2 learners

In the N400 time-window, a significant *Condition* main effect ($F(2,100) = 22.50$, $p < 0.0001$) and *Condition* x *Ant-Post* interaction ($F(6,300) = 3.39$, $p < 0.05$) reached significance in the absence of any interactions with *Group* (ps > 0.5). The correct condition significantly differed from both *Mismatch* ($F(1,50) = 41.36$, $p < 0.0001$; *Condition* x *Ant-Post*: $F(3,150) = 4.50$, $p < 0.01$) and *Swap* ($F(1,50) = 38.67$, $p < 0.0001$; *Condition* x *Ant-Post*: $F(3,150) = 3.01$, $p < 0.05$) violations, which in turn were only marginally different from each other (*Condition* x *Ant-Post:* $F(3,150) = 2.68$, $p = 0.07$). Crucially, planned comparisons revealed that the two High proficiency bilingual groups did not differ from each other in any of the N400 effects they elicited (all ps > 0.2).

In the P600 window, however, *Group* did exert an effect ($F(4,10) = 2.70$, $p < 0.05$), confirming that, for both violations, the P600 effect was present only in the two L1-Italian groups (*Condition* x *Ant-Post* effects in Controls: $F(6,174) = 7.68$, $p < 0.001$; Attriters: $F(6,72) = 2.26$, $p < 0.05$), but not in High-Proficiency L2 learners (ps > 0.1).

3.4.2. Controls vs. Low Proficiency Attriters vs. Low Proficiency L2 learners

The global ANOVA revealed a marginal interaction between *Condition* x *Group* $F(4,94) = 2.01$, $p = 0.09$) which, when followed-up by *Condition* pairs, followed the direction of our hypotheses, namely that the groups differed most on the N400 effect in *Swap* conditions rather than *Mismatch* conditions. For *Mismatch* vs. *Correct* sentences, a significant *Condition* x *Group* interaction ($F(2,47) = 2.98$, $p < 0.05$) reflected the large N400 response elicited in Controls ($F(1,29) = 39.45$, $p < 0.0001$) relative to the weaker N400 effect in Low Proficiency Attriters ($F(1,9) = 6.34$, $p < 0.05$) and Low Proficiency L2 learners ($F(1,9) = 4.13$, $p < 0.05$). For Swap-Correct sentences, the significant *Condition* x *Group* interaction ($F(2,47) = 4.26$, $p < 0.05$) confirmed that Controls were the only group to show a significant main effect of *Condition* ($F(1,29) = 23.16$, $p < 0.0001$), while neither one of the

Low Proficiency bilingual groups showed a significant N400 for the *Swap* condition (*p*s > 0.1). Importantly, planned comparisons confirmed that the two Low Proficiency bilingual groups did not differ from each other on any N400 effects (ps > 0.2).

As predicted, *Group* also exerted an effect in the P600 window (*Condition* x *Ant-Post* x *Group* ($F(12,282) = 3.09$, $p < 0.005$), given Low Proficiency L2 learners did not elicit P600 responses in either violation condition relative to correct sentences (*p*s > 0.1).

In sum, High Proficiency Attriters and High Proficiency L2 learners were more "native-like" (similar to native-Controls) than Low Proficiency subgroups in the N400 time-window, but the two bilingual groups did not differ from one another in their N400 responses when they were assigned to the same proficiency subcategory (High or Low). The only *Group* differences that persisted between Attriters and L2 learners were in the P600 time-window, where the L2 Italian learners did not elicit P600 effects in response to either violation, regardless of proficiency level.

3.5. Proficiency as a continuous variable

To evaluate the impact of proficiency level on ERP responses with proficiency as a continuous rather than categorical variable, we examined the relationship between individual proficiency measures in correlations with ERP amplitudes in the N400 and P600 time-windows.

3.5.1. Proficiency measures and the N400

Correlational analyses indicated that the amplitude of the N400 effect for *Swap-Correct* was negatively correlated with scores on the *C-test* ($r = -0.306$, $p < 0.0001$), the *Error-test* ($r = -0.387$, $p < 0.0001$) as well as *Semantic fluency* ($r = -0.272$, $p < 0.01$), such that individuals with higher proficiency scores elicited a larger N400 (i.e., less positive amplitude). The N400 effect for *Mismatch-Correct* was also negatively correlated with scores on the C-test ($r = -0.332$, $p < 0.01$), *Error-detection test* ($r = -0.293$, $p < 0.01$) and *Semantic fluency* ($r = -0.302$, $p < 0.01$).

3.5.2. Proficiency measures and the P600

The posterior P600 elicited by *Swap* violations was positively correlated with participants' performance on the *Error-test* ($r = 0.259$, $p < 0.05$) and on the *Semantic fluency* task ($r = 0.285$, $p < 0.05$), while the *C-test* was not a significant predictor of the P600 following *Swap* errors ($p > 0.1$). In contrast, all three individual tests were positively correlated with P600 amplitudes elicited in response to *Mismatch* errors (*C-test*: $r = 0.321$, $p < 0.01$; *Error-test*: $r = 0.390$, $p < 0.01$; *Semantic fluency*: $r = 0.281$, $p < 0.01$), such that individuals with greater proficiency scores elicited a larger P600 effect following *Mismatch* violations.

3.6. The effect of background factors in L1 attrition and L2 processing

There were no significant correlations between *age at testing* or *education level* with any observed ERP patterns, in either time-window (ps > 0.1).

In Attriters, length of residence (LoR) was significantly correlated with the N400 amplitude in *Mismatch* violations relative to *Correct* sentences (Fz: $r = 0.356$, $p < 0.05$; Cz: $r = 0.369$, $p < 0.05$), indicating that individuals who spent a greater number of years outside of Italy elicited weaker N400 effects in response to *Mismatch* violations, and therefore were *less* native-like in Italian. Hours of L1-Italian exposure were also negatively correlated with N400 amplitudes in response to *Mismatch* violations (Pz: $r = -0.397$, $p < 0.05$), such that Attriters with more exposure to Italian elicited a larger N400 effect than Attriters with less exposure to Italian.

Interestingly, in L2 learners, correlations between age-of-acquisition of Italian (AoA) and ERP responses did not reach significance, for any conditions in either time-window (ps > 0.1).

4. DISCUSSION

The present study examined the neurocognitive correlates of real-time lexical-semantic processing in adult L1-attriters and late L2-learners of Italian compared to non-attriting Italian monolingual-speakers. We investigated whether L1-attriters, despite being native-speakers of Italian, showed evidence of non-native-like processing patterns during comprehension, particularly for pairs of words that are notoriously "confusable" in Italian. A second critical aim was to determine whether L1-attriters and late L2-learners exhibited similar processing patterns despite their opposite AoA profiles, and whether ERP responses to lexical-semantic anomalies fell along a continuum modulated by proficiency-level, regardless of whether Italian was the L1 or L2. Our ERP findings will be summarized and interpreted in turn for each of the ERP components of interest.

4.1. N400 (300-550 ms)

Our findings revealed that proficiency level but not group membership (i.e., L1/L2 status) modulated the amplitude and scalp distribution of the N400 (350-500 ms) in response to lexical-semantic violations relative to correct control sentences. When comparing *Mismatch* vs. *Correct* sentences, high-proficiency individuals (from all groups) were shown to elicit large N400 effects which were maximal at Cz and Pz, whereas low-proficiency individuals elicited significantly smaller and more broadly-distributed N400 effects. As predicted, proficiency effects were even more robust for *Swap* sentences, with only high-proficiency individuals showing an N400 effect for *Swap* violations, whereas these sentences were statistically indistinguishable from *Correct* sentences in low-proficiency subgroups. Proficiency differences therefore modulated the graded response pattern between *Mismatch*, *Swap* and *Correct* sentences – in high-proficiency subgroups, the major N400 difference was between *Swap*-*Correct* conditions, whereas the two violation conditions did not significantly differ from one another. This pattern differed in low-proficiency subgroups, however, who showed an N400 difference for the *Mismatch*-*Swap* comparison but not for *Swap*-*Correct*.

Proficiency effects in the N400 time-window in the absence of group effects were further supported by analyses separately comparing the high-proficiency bilingual subgroups and low-proficiency bilingual subgroups to native-monolingual Controls. For both violation conditions, native-Controls and high-proficiency bilinguals (i.e., L1-attriters and late L2-learners) were statistically indistinguishable in their N400 responses. Conversely, group differences existed between native-Controls and both low-proficiency subgroups of L1-attriters and L2-learners, with low-proficiency subgroups failing to show a significant N400 effect in response to *Swap* violations. Crucially, however, L1-attriters and L2-learners were indistinguishable in their responses when they belonged to the same Italian-proficiency cluster.

For decades, the N400 has been reliably elicited in response to lexical-semantic anomalies, and has been argued to index processes of lexical-access and integration, with its amplitude being sensitive to a number of factors in addition to semantic-plausibility, such as familiarity, word-frequency, sentence-constraint, orthographic/phonological similarity, etc. (see Federmeier & Kutas, 2011 for a review). In our study, the two lexical-semantic violations differed in the word-form ambiguity of the target noun – in one case (*Swap* condition), the target-word and the word intended for the sentence-context belonged to a minimal pair, such that they were orthographic neighbors. Otherwise, both violation conditions were equally semantically-implausible, and the experimental design was optimally balanced such that the same target words contributed to each condition, with no asymmetrical loss of trials across conditions. Thus, N400 amplitude differences between *Mismatch* vs. *Swap* violations were due to our experimental manipulation (i.e., confusability due to word-form overlap).

In native-Italian controls, both violation conditions elicited an N400 effect of almost identical amplitude. Previous studies have provided evidence that sentences where the target noun has been replaced by an orthographic or phonological neighbor elicit reduced N400 effects compared to non-neighbor sentence conditions, even though both types of violation sentences are semantically-implausible to the same degree (Lazslo & Federmeier, 2009). This graded N400 response has been argued to reflect co-activation during lexical access, where the presented target word and similar word-forms are activated in parallel. However, our results were not entirely in line with such findings, as the N400 difference between *Mismatch* and *Swap* violations was only marginal in native-Controls and high-proficiency subgroups of L1-attriters and L2-learners. A likely reason for this discrepancy is that Lazslo and Federmeier had examined highly-constraining sentence-contexts whereas our study included a range of high and low cloze-probability items. In highly-constraining contexts, one might expect readers to make stronger predictions for the target-word, and it is therefore conceivable that readers would more strongly activate the target-word before its actual presentation, thus also priming its orthographic/phonological neighbor. However, if orthographic neighbors are indeed simultaneously activated along with the expected target-words, readers must – at some point in the lexical retrieval process – select the appropriate word to fit the sentence context and inhibit competing neighbors efficiently in order to arrive at the correct interpretation of the sentence (e.g., Zwitserlood, 1989). Thus, they must be able to tell the words apart, inhibit the irrelevant semantic-meaning and pursue the online interpretation in such a way so as not to erroneously process "*Swap*" sentences as "*Correct*". Native-monolingual controls and high-proficiency attriters and L2 learners showed evidence of this process, as *Swap* violations elicited large N400 effects relative to correct sentences.

On the other hand, lower-proficiency individuals were shown to be less likely than high-proficiency groups and native-controls to show N400 effects for *Swap* violations, which were indistinguishable from correct sentences in processing responses. There are two conceivable possibilities for the overlap between *Swap* and *Correct* sentences in low-proficiency individuals: (1) either readers confounded the minimal pair words at the form-concept level and did not immediately detect the error in the N400 time-window because they mistook "*cappella*" for "*cappello*"; or (2) readers showed inefficient inhibition of non-target semantic meanings, such that the simultaneous activation of the other word in the minimal pair led to a reduction in the N400 violation effect. In the first case, the problem may stem from the level of lexical representations, where the two distinct concepts are not strongly linked to the appropriate form and gender. In the latter case, a simultaneous activation of both "*cappello*" and "*cappella*" would be indicative of a lesser degree of specificity in lexical access / selection by maintaining co-activation of related forms despite their semantic incongruity with the preceding context. Previous studies have argued that L2 learners are less efficient in inhibiting co-activation of intralingual competitors such as orthographically/phonologically-similar word-forms (FitzPatrick & Indefrey, 2009; Rüschemeyer et al., 2008; Sebastián-Galles, Echeverria & Bosch, 2005). While there was little to no evidence of such co-activation in the monolingual Controls and high-proficiency bilinguals, it is conceivable that lower-proficiency participants may be more influenced by competing word-forms during sentence processing. Conversely, it may be that the mechanism underlying these processing differences in the lower-proficiency subgroups is not simply a matter of degree but is qualitatively different from what is observed in higher-proficiency groups, namely that the specific lexical-semantic forms are not immediately available to these readers.

Our current study does not permit us to distinguish between these two possibilities, but a future investigation of effects of sentence-constraint, frequency and word-familiarity/exposure may help clarify the nature of the reduced online sensitivity to *Swap* sentences in these individuals. For example, if the problem is a general insensitivity in readily detecting the error due to having confounded the words in the minimal pair, then it is conceivable that error-detection should be easier when the presented target-nouns are high in frequency, as high-frequency nouns are arguably more accessible both to L1-attriters (e.g., Paradis, 1989; 1997) and L2-learners (e.g., Ullman, 2001). Given that high-frequency words are more easily recognized, lower-proficiency individuals may approach the N400 response patterns observed in higher-proficiency subgroups and in native-Controls when target-words appearing in semantically-incongruent contexts are highly-frequent rather than infrequent.

Alternatively, if the reduced N400 is the result of an inhibition / selectivity problem, then N400 effects may be more greatly affected by factors that have been shown to influence co-activation / inhibition effects both within and between languages, such as sentence-constraint and the *relative* frequency between the expected target-word and its competitor. Inhibition of intralinguistic competitors should be more efficient when sentence-contexts are highly-constraining (i.e., larger N400 effects for "*cappella*" when sentence strongly primes "*cappello*") and if the presented competitor is *lower* in frequency than the noun that was primed by the sentence-context but not presented (Elston-Guttler, Paulmann & Kotz, 2005; Libben & Titone, 2009; Kerkhofs et al., 2006; Van Hell & De Groot, 2008 for a similar argument about cross-linguistic interference).

These future directions aside, the key finding in our study was that, for both lexical-semantic violations but *especially* for the trickier and less salient *Swap* violation, proficiency level significantly predicted N400 amplitudes irrespective of group, suggesting that whether lexical items were acquired early in life in a monolingual-L1 context or as a late-learned L2 did not influence processing patterns of confusable words. We found that lower-proficiency L1-attriters were more similar to lower-proficiency L2 learners than to native-controls in the N400 time-window. Our findings were in line with the idea of a continuum in ERP profiles, with lower Italian proficiency scores being associated with smaller N400 amplitudes (especially for the *Swap* condition), regardless of whether Italian was the L1 or L2.

4.2. P600 (650-1000 ms)

Our finding of P600 effects following N400 responses to lexical-semantic violations was in line with a number of previous studies, suggesting that neurocognitively demanding revision of the input triggers a P600 effect even in lexical-semantic violations as opposed to strictly morphosyntactic ones (e.g., Diaz & Swaab, 2007; Kutas, Lindamood & Hillyard, 1984; Laszlo & Federmeier, 2009; van den Brink, Brown & Hagoort, 2001; Van de Meerendonk et al., 2009; see Van Petten & Luka, 2012 for a review). The presence, amplitude and scalp distribution of the P600 was influenced by the type of violation, proficiency-level and by group membership. Insights on the functional significance of the P600 and implications for L1-attrition and L2-processing will be discussed following a brief summary of the observed patterns.

A P600 effect for *Mismatch* sentences relative to *Correct* sentences was found in both native-speaker groups, but the P600 response in Attriters was larger in amplitude and less focal/posterior than the P600 in non-attriting Controls. Late L2 learners, on the other hand, showed a negativity between 800-900 ms for the *Mismatch* condition instead of a positivity. Contrary to the two native-speaker groups who elicited a significant P600 in response to *Mismatch* violations relative to correct sentences, our group of L2 learners did not show any indication of a P600 effect. In addition to group differences, both the amplitude and the scalp distribution of the P600 was additionally modulated by proficiency-level, such that individuals with higher Italian proficiency scores showed more robust and more posterior P600 responses for *Mismatch* violations compared to lower-proficiency individuals.

The P600 effect elicited in response to *Swap* violations was similarly influenced by *Group* and *Proficiency*, but its amplitude was significantly larger than that of the P600 effect for *Mismatch* sentences. Again, only the two L1-groups showed P600 effects for *Swap* vs. *Correct* sentences, whereas L2 learners elicited no significant effects in the P600 time-window. When L2 learners were removed from the analysis, Attriters were found to differ significantly from non-attriting native-speakers by eliciting larger and less focal (more broadly-distributed) P600 responses to *Swap* sentences, thus echoing the pattern we had observed for *Mismatch* violations. These differences in P600 amplitudes and distribution between Attriters and Controls were not modulated by proficiency-level, suggesting that the larger and broader P600 responses were associated with characteristics of the attrition group overall. We will return to this point further below.

What does the P600 reflect in our semantically-anomalous conditions? Why was the effect larger for Swap violations, largest and less posterior for L1-Attriters, more robust in high-proficiency than in low-proficiency Italian speakers, and absent even in highly-proficient late L2 learners? First, the condition effect on P600 amplitudes (*Swap* > *Mismatch*) was consistent with those reported in the study by Laszlo and Federmeier (2009) that contrasted sentences where the sentence-final target-noun was either semantically-congruent, a semantically-incongruent orthographic neighbor of the expected ending, or incongruent and dissimilar in form to the expected ending. As in our experiment, participants performed an acceptability judgment task, though judgments were binary decisions (yes/no). Although the target-words in their study were sentence-final and sentence-contexts were highly-constraining (*M* cloze-probability = 89%), a late-positive-complex (LPC) over posterior sites of the scalp was reported between 450 and 750 ms, and was larger in amplitude for neighbor violations than unrelated violations. The authors discussed the effect as reflecting the explicit realization that the target-word was orthographically close to the expected word.

In line with this view, one could argue that if the larger P600 merely reflects the "realization" of the orthographic similarity between the intended word and the presented word, late L2 learners do not elicit a large P600 effect in response to *Swap* sentences because they did not recognize that the incongruent target-noun was only one letter/phoneme away from the noun that was intended to fit the sentence-context. However, this would seem at odds with our finding that at least high-proficiency L2 learners were able to detect the erroneous insertion of an orthographic neighbor in the *Swap* condition, eliciting a similar N400 effect as high-proficiency native-speakers. Furthermore, it is not intuitive why Attriters would be "better" at recognizing the orthographic similarity of the nouns compared to non-attriting controls. Thus, it would seem that a process other than merely *noticing* the orthographic similarity between congruent and incongruent targets underlies the P600 response to these stimuli.

Alternatively, the posterior P600 effects elicited in response to both lexical-semantic violations have been discussed as as reflecting a general conflict monitoring process. When a strong conflict is detected, the reader must re-analyze the input in order to determine where processing errors may have occurred (e.g., Van de Meerendonk et al., 2009; Vissers, Chwilla & Kolk, 2006; Vissers et al., 2008). The more severe the conflict, the larger the P600 response. In our study, *Mismatch* and *Swap* sentences were equally semantically-implausible, but only the *Swap* condition involved an additional conflict at the word-form level, as the target-noun differed in its (gender-marking) final vowel, and was expected to cause intralingual competition during lexical-access. The enhanced conflict caused by semantic anomaly + orthographic ambiguity may have triggered more elaborated re-analysis processes to double-check the preceding input relative to *Mismatch* violations where the only conflict was on a semantic level. Although proponents of the conflict monitoring view of the P600 do not explicitly discuss conflict on a level other than semantic implausibility, they do mention that all aspects of input – including orthographic – are double-checked during re-analysis while the reader attempts to pinpoint the source of the problem.

The proficiency and group differences we found to modulate P600 amplitudes would also seem to fit this theoretical framework. A number of studies have shown that highly-proficient speakers engage in re-analysis / repair processes as indexed by the P600, whereas low-proficiency learners may not elicit P600 responses during early stages of learning (e.g., Osterhout et al., 2006; White, Genesee & Steinhauer, 2012). However, the large P600 responses elicited in *both* proficiency subgroups of L1-Attriters, together with the finding that the significant differences in amplitude and scalp distribution between Attiters and non-attriting native-controls (when analyzed without L2 learners) were *not* modulated by proficiency level indicate that there is more to our P600 differences than normal variability in proficiency-level among native L1-speakers. The larger P600 effects overall for L1-Attriters may be indicative of increased conflict-monitoring and a more explicit second-thought or double-take further downstream from the initial detection of the anomaly, particularly for *Swap* violations where they may re-analyze the sentence and determine "It should have been X".

This idea of a "second-thought" resulting in a more elaborated re-analysis process from native-Controls is supported by the finding that Attriters arrived at appropriate acceptability ratings but after longer reaction-times than non-attriting Controls. Given that Attriters are a special group of bilinguals who are aware of their special circumstances and of the change they have been experiencing in their L1 since immigration, it is intuitive that increased attention and a more explicit metalinguistic analysis of the sentences during the experiment would characterize this group. This motivation to perform well in their "attriting" native-language may be stronger than L2-learners' motivation to perform well in a second-language. In other words, showing you can maintain your native-language may be a stronger motivating factor than showing you can master a second-language. This idea is currently speculative but may be worth exploring in a follow-up study, given that our background questionnaires contained a number questions that were motivational and socio-affective in nature.

The P600 effect was absent in late L2 learners, even in those with higher proficiency scores. Instead, only in response to *Mismatch* violations, L2 learners showed a negativity in the P600-window. It is unclear whether this negativity was an N400-effect on the word following the target-noun (a function word) or a sustained negativity driven by the target-word. The N400 on the target-word in the *Mismatch* condition did return to baseline prior to the N1/P2 complex elicited by the subsequent function word. It is unlikely that the following function word would elicit an N400 in the *Mismatch* condition relative to the other two conditions, given that sentence-contexts were optimally balanced in our experimental design and, in general, the function words did not differ in length or frequency. It is conceivable that individual differences in N400 latency may have contributed to this group pattern, as it has been shown that late L2 learners (particularly low-proficiency learners) show delayed N400 effects in lexical-semantic process (Elston-Guttler & Friederici, 2007; Hahne, 2001; Moreno & Kutas, 2005; Weber-Fox, Davis & Cuadrado, 2003). However, the main finding was that L2 learners did not elicit a P600 in response to either violation condition, and therefore did not seem to engage in the same re-analysis process as the two groups of native-speakers. This lack of a re-analysis may be related to our finding that late L2 learners gave higher acceptability ratings to violation conditions compared to the native-speaker groups, despite their slow reaction-times which echoed the slow responses of Attriters.

This brings us to the question of whether the conflict-monitoring and re-analysis processes that are indexed by the P600 in response to our experimental stimuli are processes than only native-speakers engage in, or whether the proficiency level of our L2 learners was still not high enough for these processing patterns to emerge. Testing a group of English-Italian late-learners living in Italy and therefore using Italian regularly and implicitly (rather than only in classroom contexts) may reveal that more native-like use of the language leads to the development of re-analysis processes even in late-learners. It is conceivable that a fundamental difference in how such lexical items are acquired (i.e., properties relating to exposure, frequency and implicit vs. explicit learning) may have contributed to the differences we observed between L1- and L2-speakers of Italian. At least, examining trials where lexical items were of high-frequency may indicate whether P600 patterns in late-learners converged upon those of native-speakers at least for words of higher familiarity/exposure.

In our current study, late L2 learners of high Italian proficiency are indistinguishable from native-speaker Controls and high-proficiency Attriters in the N400 time-window (indexing lexical-access and integration), but diverge from those same groups in processing mechanisms further downstream from the anomaly. This unexpected but interesting finding emphasizes the usefulness of ERPs in tapping into distinct stages of processing and that simply focusing on the N400 effect in studies of lexical-semantic processing may not tell the whole story.

4.3. Neuroplasticity, L1-attrition and L2-acquisition

Our chief aim in the present study was to turn to a unique population of late-bilinguals – *L1-attriters* – to contribute in a novel way to the debate about maturational limits on neuroplasticity that are believed to dictate how native- and second-languages are processed in the brain in terms of the native-like-ness of the neurocognitive processes involved as well as their temporal dynamics.

Our results revealed that Attriters, though processing the L1 they were exposed to as exclusively monolingual speakers until adulthood, showed processing differences from non-attriting native-Controls and similarities with late *L2* learners of Italian during real-time sentence comprehension. Given that we tested lexical-semantic access and integration using an all-Italian paradigm where the words were confusable only *within* Italian, response patterns cannot be explained in terms of cross-linguistic influence or co-activation of Italian and English. Thus, we argue that the similarities in N400 responses we observed between L1-Attriters and late L2-learners were due to a proficiency continuum rather than to a general effect attributable to being bilingual. Advocates of a strong version of the claim that brain plasticity for language is limited to an early maturational window *prior* to adulthood would have difficulty reconciling the finding of a proficiency-based continuum (regardless of L1/L2 status) with their views. Instead, it has been argued that, once "fully acquired" and hard-wired, one's L1 should be stable.

Behavioral accounts of attrition have provided support for this view with their findings of attrition effects in pre-pubescent but not adult migrants (Ammerlaan, 1996; Bylund, 2009; Pelc, 2001; also see Köpke, 2004; Köpke & Schmid, 2004 for reviews). The argument is the corollary of the *age-of-acquisition* claim for L2-learning: the older you are when you are exposed to the new language, the less likely you are to experience changes in the L1 and the less likely you are to rely on native-like brain mechanisms when processing the L2, because the brain has lost plasticity. Our finding that lower-L1 proficiency Attriters were more similar to lower-proficiency L2 learners of Italian than to native-Controls in their N400 responses to confusable Italian words (and that this pattern was *not* modulated by age-at-immigration) can be constituted as evidence against hard maturational limits on neuroplasticity for language. Furthermore, Attriters as a *group* differed significantly from non-attriting native-Controls in the amplitude and scalp distribution of the P600 responses they elicited to both kinds of semantically-anomalous sentences. These results, taken together with those from other recent findings of attrition-effects from our laboratory (Kasparian, Vespignani & Steinhauer, under review; Kasparian, Postiglione & Steinhauer, under review; Kasparian, Vespignani & Steinhauer, 2013; 2014), are in favor of ongoing neuroplasticity for language in adulthood, and even for one's L1.

With respect to *L2* processing in late-learners, we showed that proficiency was the critical determinant of N400 responses in Italian-L2 learners, and that our L2-learners' ERP responses to these experimental stimuli were not modulated by AoA. Our results are therefore in line with studies showing that proficiency-level predicts the neurocognitive mechanisms underlying language processing, and that high-proficiency L2 learners can show processing mechanisms that are indistinguishable from native-speakers, even if the L2 was acquired at a late AoA (Bowden et al., 2014; Friederici, Steinhauer & Pfeifer, 2002; Osterhout et al., 2006; Rossi et al., 2006; Steinhauer, White & Drury, 2009; White, Genesee & Steinhauer, 2012). The notion of a proficiency-based *continuum* of ERP responses fits well with those revealed in longitudinal studies of L2 acquisition (e.g., McLaughlin et al., 2009; Osterhout et al., 2006; White, Steinhauer & Genesee, 2012), as well as one unpublished study comparing stages of L2 acquisition and L2 attrition (Pitkänen, Tanner & Osterhout, unpublished dissertation). The finding that this continuum also spans across *L1*-groups (such as Attriters) is a novel contribution. It remains to be seen whether a proficiency-based continuum in N400 responses across L1/L2 groups (i.e., regardless of AoA) was observed because lexical-semantic processing is argued to benefit more from ongoing neuroplasticity than morphosyntactic processing (as has been argued by Newport, Bavelier & Neville, 2001; Wartenburger et al., 2003; Weber-Fox & Neville, 1996), or whether similar results may be observed for complex morphosyntax (Kasparian, Vespignani & Steinhauer, under review). The latter case would pose a greater challenge for advocates of maturational constraints on neuroplasticity for language processing.

The question we are currently unable to answer is whether the differences we observed in the P600 effects elicited in response to lexical-semantic violations relative to correct sentences were attributable to (1) differences in L1/L2 status, (2) still much-too-low L2 proficiency and exposure-levels, and/or (3) more general characteristics such as attention or motivation. We attributed the P600 effects we found to increased conscious effort during conflict-monitoring and re-analysis processes. Follow-up studies are required to further examine P600 effects in response to semantically-anomalous sentences, to determine what kinds of lexical-semantic factors modulate the degree of perceived conflict in such sentences, and whether input-checking and re-analysis are stages of processing that may also become native-like at high proficiency-levels and/or when a language has been acquired in implicit (immersion) contexts.

5. CONCLUSIONS

Using ERPs, we examined the neurocognitive processes underlying real-time lexical-semantic processing in Italian L1-Attriters who reported experiencing changes in their native-L1 after a prolonged period of limited L1 exposure/use, compared to late L2-learners of Italian and monolingual non-attriting Italian native-speakers. We showed that, in lower-proficiency Italian speakers, sentences with "confusable nouns" (ambiguous until their final vowel and gender-marking) were not processed as lexical-semantic violations and did not elicit a significant N400 effect, contrary to "classical" lexical-semantic (*Mismatch*) violations. Crucially, N400 responses followed a continuum of "native-like-ness" predicted by Italian proficiency scores, regardless of L1/L2 status. Both lexical-semantic violations also elicited a posterior P600 effect, but this effect larger for *Swap* rather than *Mismatch* violations, larger and more broadly-distributed in Attriters than in Controls, and absent in late-learners. We discussed the P600 in the framework of conflict-monitoring and re-analysis processes, and argued that larger P600 responses reflected a more conscious, explicit, elaborated "second-thought" process in Attriters as a group.

Our findings provided the first ERP evidence of attrition effects within the L1 (rather than L1-L2 competition / transfer) during online comprehension in L1-natives who lived in an exclusively monolingual L1-context until adulthood. We advocate for ongoing neuroplasticity for language, even in adulthood, and emphasize that proficiency is a key factor in modulating native-like neurocognitive responses, regardless of whether the language being processed was acquired as the L2 or the L1. However, attrition effects are not entirely accounted for by proficiency modulations, and group differences from non-attriting controls in conflict-monitoring and conscious "second-thought" processes may be due to factors surrounding this demographic group's special socio-linguistic circumstances.

ACKNOWLEDGEMENTS

K.K was supported by a Vanier Canada Graduate Scholarship from the Canadian Institutes of Health Research, a Richard Tomlinson Doctoral Fellowship awarded by the Faculty of Medicine of McGill University, and a Michael Smith Foreign Study Supplement (MSFSS-CGS). This research was supported by grants awarded to K.S. by the Canada Research Chair program and the Canada Foundation for Innovation (CRC/CFI; project # 201876), the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPGP 312835-05 and RGPIN 402678-11), the Social Sciences and Humanities Research Council of Canada (SSHRC; # 435-2013-2052), and the Fonds de Recherche Société et Culture, Québec (FQRSC; # 2010-SE-103727). The authors sincerely thank Francesco Vespignani for his collaboration and support during K.K's research stay in Italy, as well as John Drury, Eleonora Rossi and Francesca Postiglione for valuable feedback on our experimental stimuli. We are also indebted to research assistants Linna Jin, Kristina Maiorino, Filippo Vicari and Paolo Zandomeneghi for their invaluable help with data collection.

Study 2: Tables and figures

Table 1. *Four examples of experimental sentences are provided each condition. English translations are presented in italics. The target noun is underlined. The asterisk marks a lexical-semantic violation.*

CONDITION	
Correct	Per coprire la testa, il pescatore porta il **cappello** di lana. *To cover his head, the fisherman wears the hat of wool.* Per assistere alla messa, la coppia frequenta la **cappella** ogni giorno. *To assist the mass, the couple attends the chapel every day.* Per indicare gli orari, il negoziante lascia il **cartello** in vitrina. *To indicate the opening-hours, the shopkeeper leaves the sign in the window.* Per nascondere i documenti, la spia mette la **cartella** in cassaforte. *To hide the documents, the spy puts the briefcase in the safe.*
Swap	Per coprire la testa, il pescatore porta la ***cappella** di lana. *To cover his head, the fisherman wears the *chapel of wool.* Per assistere alla messa, la coppia frequenta il ***cappello** ogni giorno. *To assist the mass, the couple attends the *hat every day.* Per indicare gli orari, il negoziante lascia la ***cartella** in vitrina. *To indicate the opening-hours, the shopkeeper leaves the *briefcase in the window.* Per nascondere i documenti, la spia mette il ***cartello** in cassaforte. *To hide the documents, the spy puts the *sign in the safe.*
Mismatch	Per coprire la testa, il pescatore porta la ***cartella** di lana. *To cover his head, the fisherman wears the briefcase of wool.* Per assistere alla messa, la coppia frequenta il **cartello** ogni giorno. *To assist the mass, the couple attends the *document every day.* Per indicare gli orari, il negoziante lascia la ***cappella** in vitrina. *To indicate the opening-hours, the shopkeeper leaves the *chapel in the window.* Per nascondere i documenti, la spia mette il ***cappello** in cassaforte. *To hide the documents, the spy puts the *hat in the safe.*

Table 2. *Group means (standard deviation) for proficiency and control tasks*

BEHAVIORAL MEASURES	CONTROLS (n = 30)	ATTRITERS (n = 24)	L2 LEARNERS (n = 20)
Self-report of proficiency (7 point scale)	7 (0)	6.87 (0.2)	4.94 (0.8)
Listening comprehension	7 (0)	7 (0)	5.45 (1.1)
Reading comprehension	7 (0)	7 (0)	5.35 (0.9)
Pronunciation	7 (0)	6.96 (0.2)	5.05 (1.1)
Fluency	7 (0)	6.79 (0.6)	4.7 (0.9)
Vocabulary	7 (0)	6.63 (0.7)	4.6 (0.9)
Grammar	7 (0)	6.83 (0.4)	4.5 (0.9)
C-test (%)	96.3 (4.4)	95.2 (4.6)	68.4 (19.4)
Error-detection test (%)	90.0 (5.1)	89.5 (5.9)	52.0 (19.4)
Verbal semantic fluency (average of 2 categories)	23.4 (5.5)	21.5 (3.9)	8.6 (3.4)
Reading fluency (# correct in 3 minutes)	71.6 (13.0)	75.3 (15.0)	46.0 (13.1)
Working memory			
Correct	11.2 (2.7)	11.9 (2.6)	9 (1.9)
Span	5.4 (1.1)	5.7 (1.1)	4.7 (0.7)

Table 3. *Proficiency subgroup means (standard deviation) for proficiency and control tasks.*
** p < 0.05. ** p < 0.01. *** p < 0.005. ns = not significant (p > 0.1)*

BEHAVIORAL MEASURES	HP CONTROLS		LP CONTROLS	HP ATTRITERS		LP ATTRITERS	HP L2		LP L2
Self-report of proficiency (7pt.scale)	7 (0)	--	7 (0)	6.97 (0.1)	*	6.74 (0.3)	5.1 (0.8)	ns	4.8 (1.1)
Listening comprehension	7 (0)	--	7 (0)	7 (0)	--	7 (0)	5.5 (1.3)	ns	5.5 (0.9)
Reading comprehension	7 (0)	--	7 (0)	7 (0)	--	7 (0)	5.7 (0.7)	ns	5.0 (0.9)
Pronunciation	7 (0)	--	7 (0)	7 (0)	ns	6.9	5.2 (0.8)	ns	4.9 (1.3)
Fluency	7 (0)	--	7 (0)	7 (0)	ns	6.5	4.7 (0.8)	--	4.7 (0.9)
Vocabulary	7 (0)	--	7 (0)	6.9	*	6.3	4.7 (0.7)	ns	4.5 (1.1)
Grammar	7 (0)	--	7 (0)	6.9	ns	6.7	4.7 (0.7)	ns	4.3 (1.2)
C-test (%)	98.8 (1.1)	***	93.6 (2.8)	97.6 (2.7)	***	92.3 (4.7)	83.7 (7.6)	***	53.2 (14.9)
Error-detection test (%)	94.6 (3.4)	***	86.2 (4.3)	93.9 (3.6)	***	84.3 (3.4)	62.4 (14.9)	**	42.1 (18.5)
Verbal semantic fluency (average)	23.5 (6.9)	ns	21.6 (5.6)	23.8 (3.0)	***	18.9 (3.3)	9.5 (3.3)	ns	7.7 (3.3)
Reading fluency (# correct)	73.4 (9.1)	ns	70.0 (15.8)	82.4 (12.4)	**	66.8 (13.9)	46.3 (8.5)	ns	45.7 (17.0)
Working memory									
Correct	11.5 (2.4)	ns	10.9 (2.9)	13.2 (2.7)	**	10.5 (1.5)	9.10 (1.1)	ns	8.9 (2.6)
Span	5.4 (1.1)	ns	5.3 (1.3)	6.2 (1.2)	*	5.1 (0.7)	4.7 (0.5)	ns	4.7 (0.8)

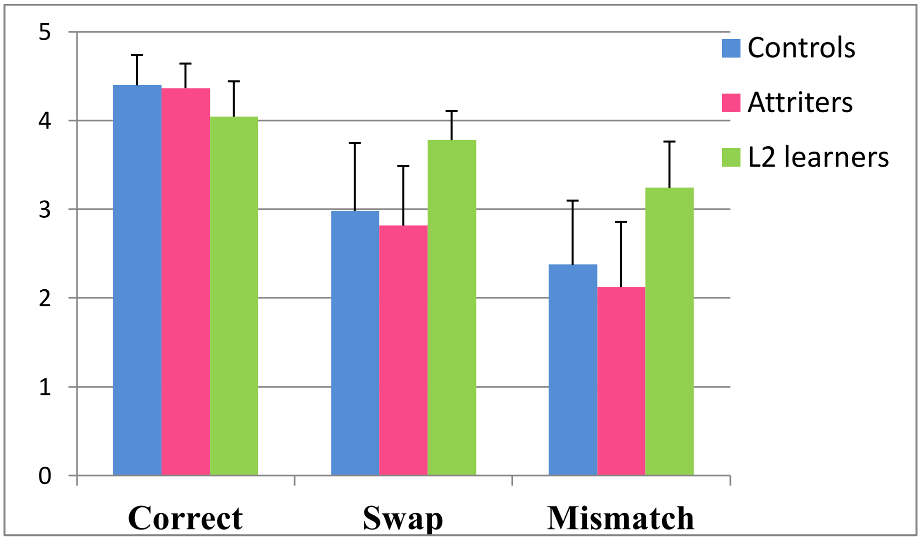


Figure 1a. Group acceptability ratings on a scale from 1 (completely unacceptable) to 5 (perfect) by condition. Both violation conditions were rated lower than the correct condition (*p*s < 0.0001). L2 learners were more willing to accept violation sentences as correct, and differed from both Controls and Attriters (*p*s < 0.01). Attriters did not differ overall from Controls. Error bars represent standard deviation.

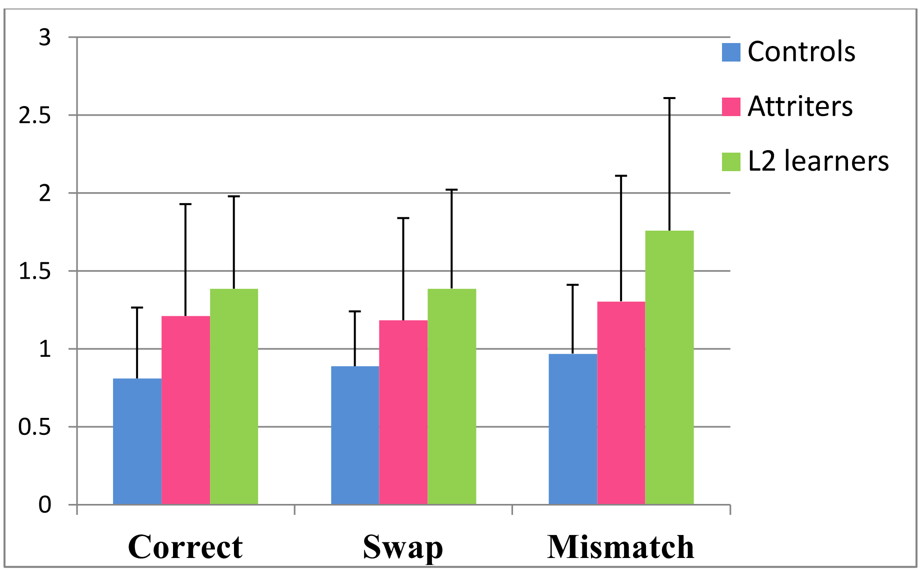


Figure 1b. Group reaction times (in seconds) by condition. Reaction times to *Mismatch* sentences were significantly slower than to *Correct* and *Swap* Sentences (*p*s < 0.0001). Group differences did not reach significance. Error bars represent standard deviation.

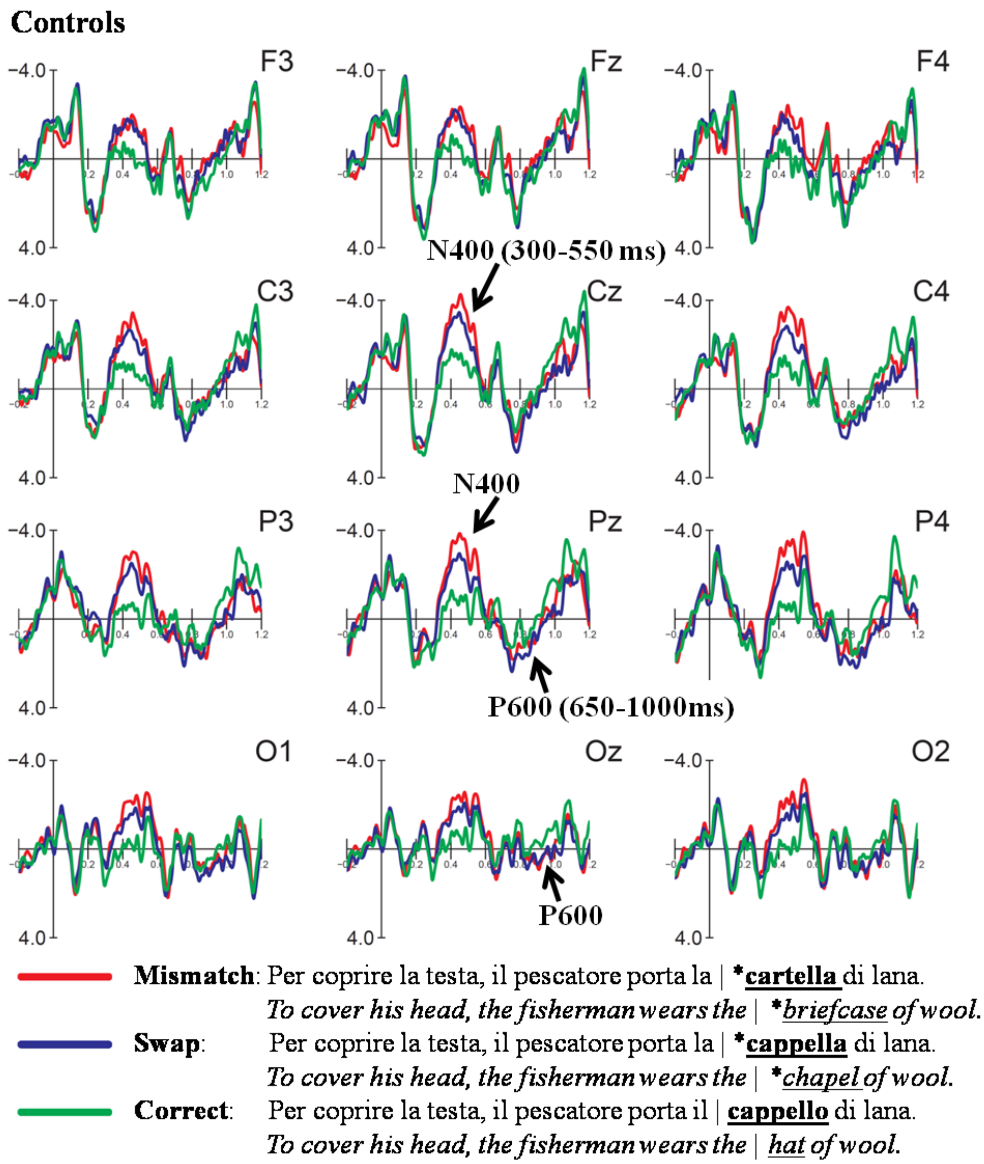


Figure 2. ERPs elicited by the target noun in response to *Mismatch* violations (red) and *Swap* violations (blue) compared to *Correct* sentences (green) in native-monolingual Controls. Time ranges (in milliseconds) depicted on the x-axis are relative to the onset of the target noun (0 ms). Negative values are plotted up. Controls elicited a biphasic N400+P600 pattern in response to both violation conditions.

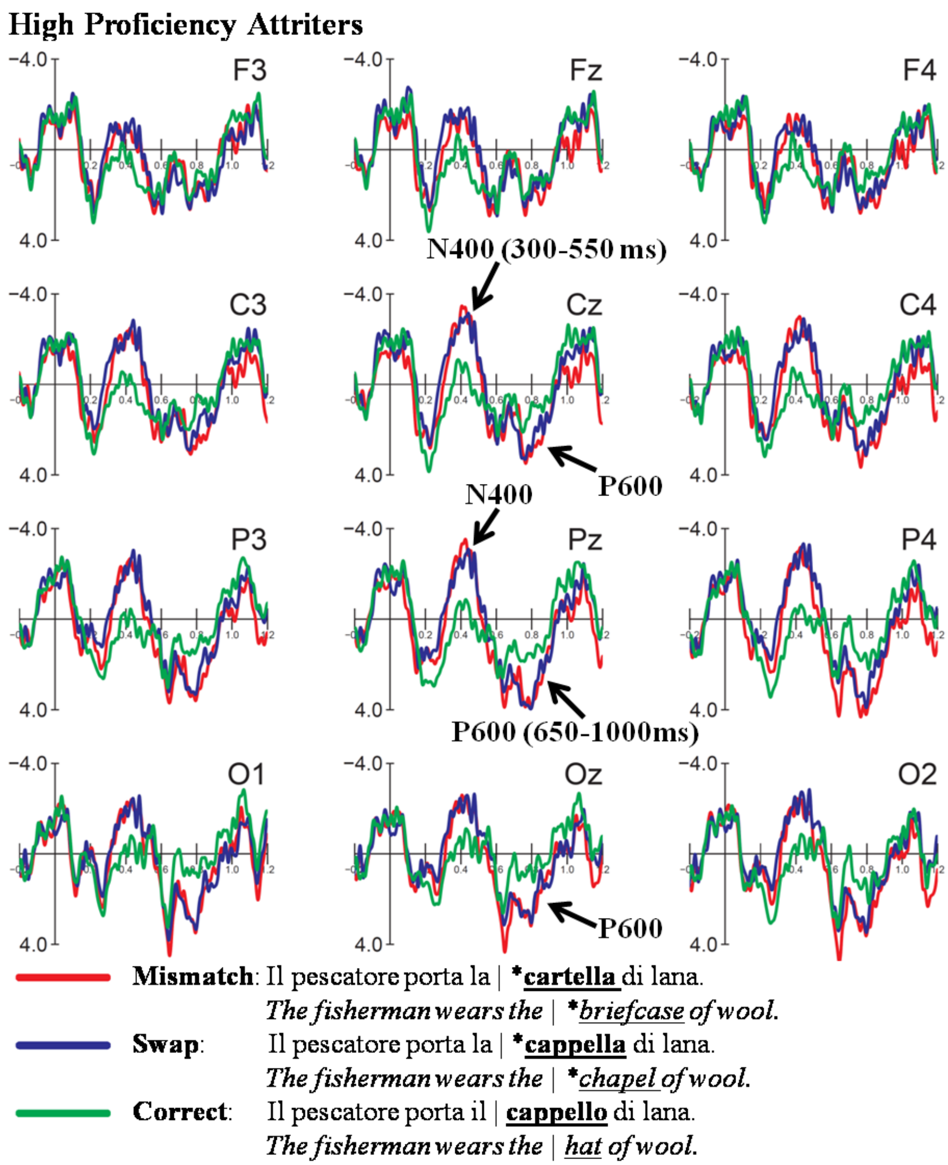


Figure 3a. ERPs elicited by the target noun in response to *Mismatch* violations (red) and *Swap* violations (blue) compared to *Correct* sentences (green) in Attriters with high Italian proficiency scores. Time ranges (in milliseconds) depicted on the x-axis are relative to the onset of the target noun (0 ms). Negative values are plotted up. High-proficiency Attriters elicited a broad N400 in response to *Mismatch* and *Swap* violations, followed by a P600 that was larger and less focal than the posterior P600 observed in Controls.

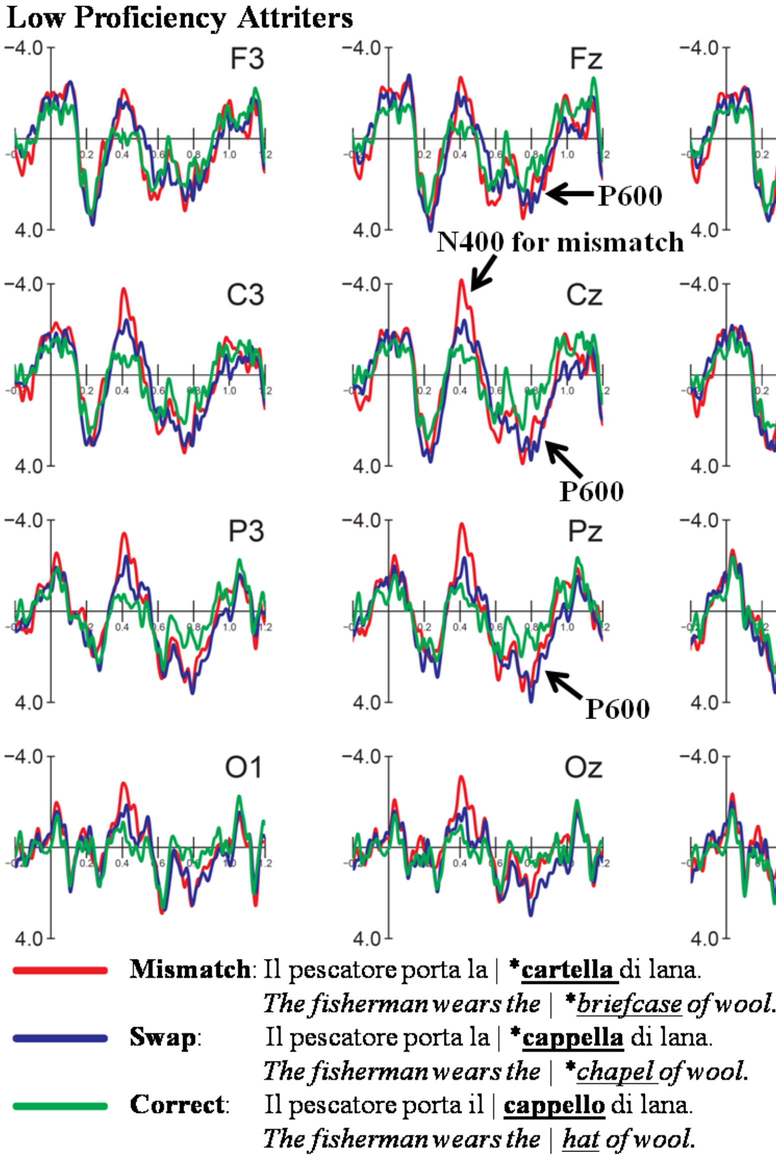


Figure 3b. ERPs elicited by the target noun in response to *Mismatch* violations (red) and *Swap* violations (blue) compared to *Correct* sentences (green) in Attriters with low Italian proficiency scores. Low-proficiency Attriters showed a broad N400 for *Mismatch* violations but a reduced N400 for *Swap* violations. The P600 effect seemed smaller and more frontally-distributed than in high-proficiency Attriters.

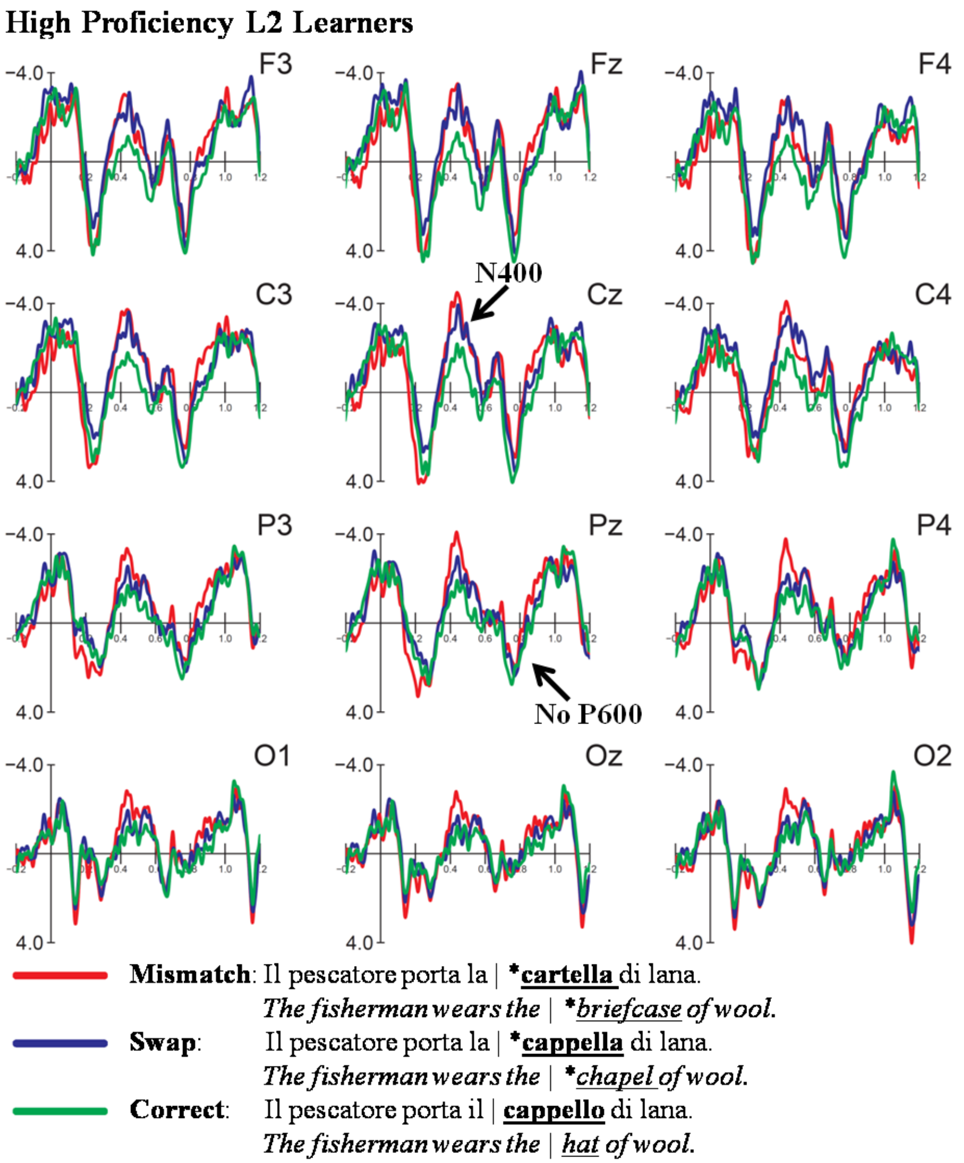


Figure 4a. ERPs elicited by the target noun in L2 learners with high Italian proficiency scores. *Mismatch* (red) and *Swap* (blue) conditions elicited an N400 effect relative to *Correct* (green) sentences, in the absence of a subsequent P600 response.

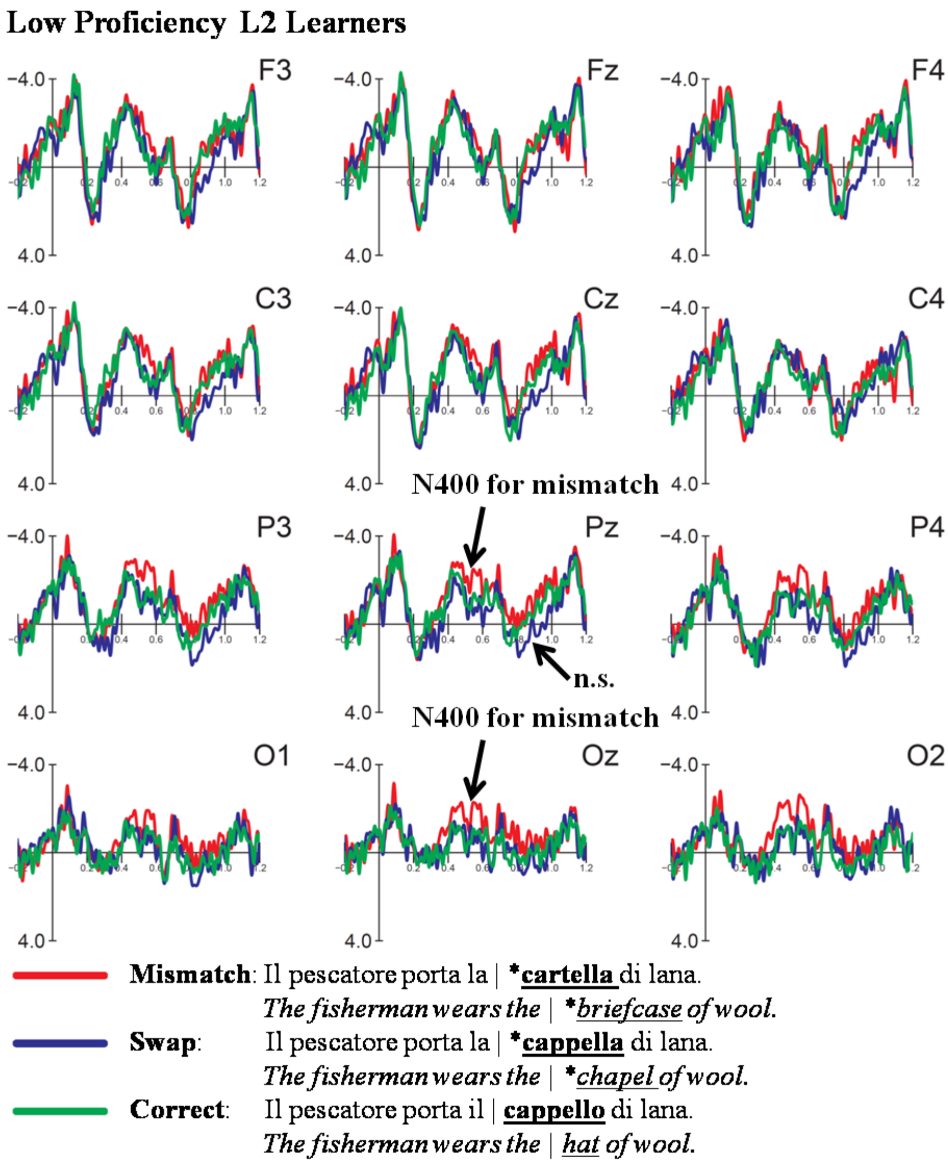


Figure 4b. ERPs elicited by the target noun in L2 learners with low Italian proficiency scores. Only the Mismatch condition elicited a posterior N400 effect, whereas the Swap condition overlapped with the Correct control condition.

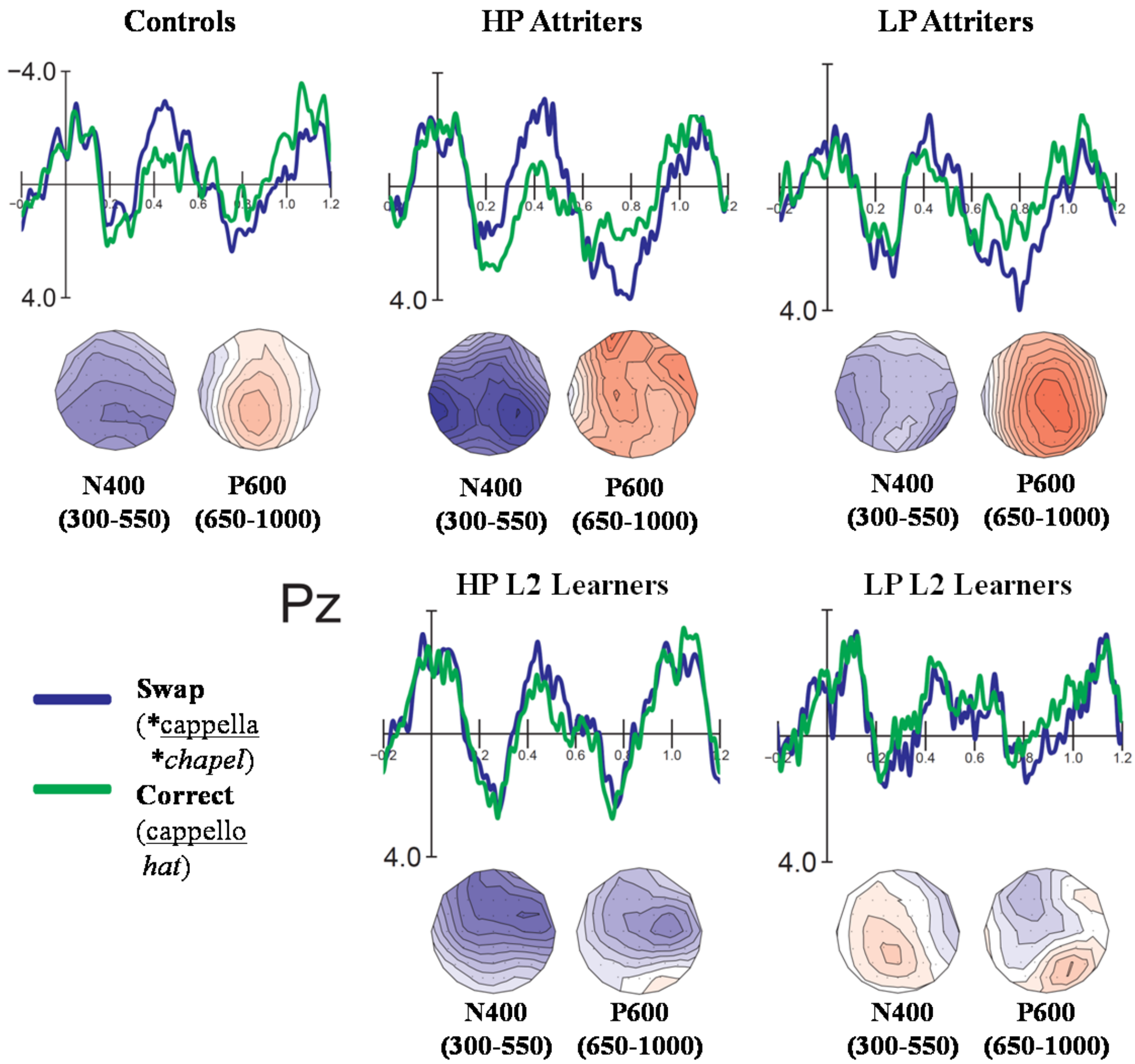


Figure 5a. ERP responses (at Pz) to *Swap* (blue) anomalies relative to *Correct* (green) sentences are compared between native-Controls and high/low proficiency subgroups of Attriters and L2 learners, to illustrate the continuum-like reduction of the N400 with decreasing proficiency. Voltage maps illustrate the scalp distribution of the N400 (300-550 ms) and the P600 (650-1000 ms). Proficiency appears to modulate the amplitude of the N400, whereas both proficiency and group membership appear to affect the P600.

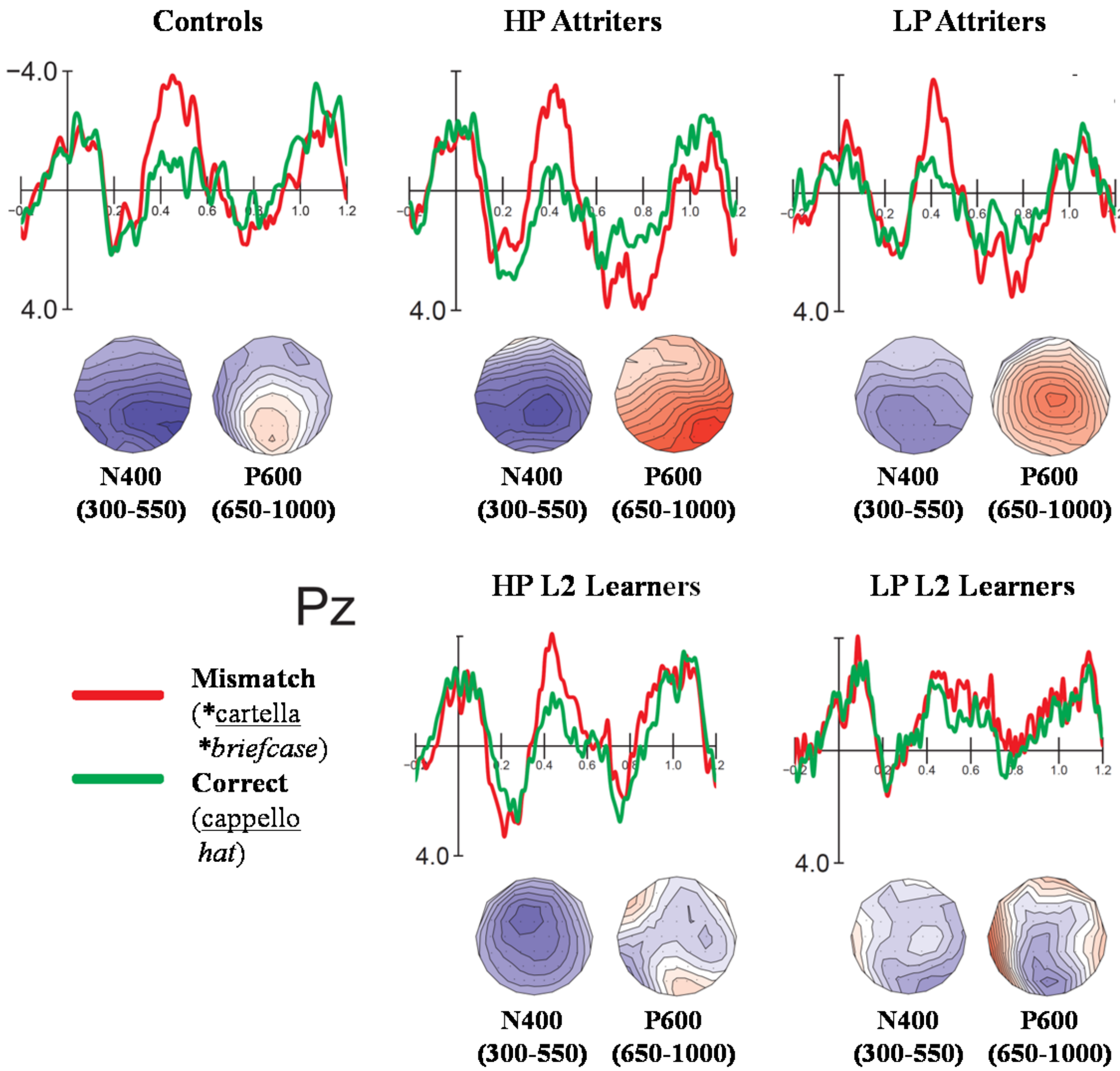


Figure 5b. ERP responses (at Pz) to *Mismatch* (red) anomalies relative to *Correct* (green) sentences are compared between native-Controls and high/low proficiency subgroups of Attriters and L2 learners. Voltage maps illustrate the scalp distribution of the N400 (300-550 ms) and the P600 (650-1000 ms). The N400 in the *Mismatch* condition appears to be less modulated by Proficiency level than *Swap* violations (Fig 5a). Both proficiency and group membership appear to affect the P600.

Study 2: References

Ammerlaan, T. (1996). *You Get a Bit Wobbly...: Exploring Bilingual Lexical Retrieval Processes in the Context of First Language Attrition*. PhD dissertation, Nijmegen University.

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C. and Thornton, A. M. (2005). Corpus e Lessico di frequenza dell'Italiano Scritto (CoLFIS).

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

Bokhari, F. S., & Steinhauer, K. (2015). Investigating the neurocognitive mechanisms underlying truth-conditional and logical semantic aspects of sentence processing: An event-related brain potential study. Unpublished Master's thesis, McGill University.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*(13), 2492–2511.

Bylund, E. (2009). Maturational constraints and first language attrition. *Language Learning*, *59*(3), 687–715.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 857-871.

Datta, H. (2010). *Brain bases for first language lexical attrition in Bengali-English speakers*. City University of New York. Retrieved from http://gradworks.umi.com/34/09/3409174.html

De Bot, K. (1996). Language loss. In H. Goebl, P. Nelde, Z. Stary, & W. Wölk (Eds.), *Contact linguistics: An international handbook of contemporary research*, Vol. 1 (pp. 579 – 585). Berlin: Walter de Gruyter.

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations: Dual ERPs to low probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207.

Diaz, M. T., & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research, 1146,* 85–100.

Ecke, P. (2004). Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism*, *8*(3), 321–354.

Elston-Güttler, K. E., & Friederici, A., D. (2007). Ambiguous words in sentences: Brain indices for native and non-native disambiguation. *Neuroscience Letters, 414*(1), 85-89.

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, *17*, 1593–1610.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84.

FitzPatrick, I., & Indefrey, P. (2009). Lexical competition in nonnative speech comprehension. *Journal of Cognitive Neuroscience, 22*(6), 1165–1178.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). The temporal structure of syntactic parsing: Early and late event-related brain potential effects elicited by syntactic anomalies. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 1219–1248.

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*(1), 529–534.

Green, D. W. (1986). Control, activation and resource: a framework and a model for the control of speech in bilinguals. *Brain and Language, 27*, 210–223.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439-483.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266.

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

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98-110.

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. *Poster presentation at the Donostia Workshop on Neurobilingualism, Basque Center on Cognition, Brain and Language*, Donostia-San Sebastián, Spain, Sept 30 – Oct 2, 2010.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013a). When the second language takes over: ERP evidence of L1-attrition in morphosyntactic processing. Talk presented at the International Conference on Multilingualism, Montreal, Canada, October 24-25, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013b). My Italian is not what it used to be: Investigating the neural correlates of L1 attrition and late L2 acquisition. Poster presented at the Workshop on Neurobilingualism, Groningen, The Netherlands, August 25-27, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014a). The case of the non-native-like first language: ERP evidence of L1-attrition in lexical and morphosyntactic processing. Symposium presentation, The 17th World Congress on Psychophysiology, Hiroshima, Japan, September 23-27, 2014.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014b). Neurophysiological correlates of L1 attrition and L2 acquisition: A continuum based on proficiency. Poster presented at the Society for the Neurobiology of Language, Amsterdam, Netherlands, August 27-29, 2014.

Kasparian, K., Vespignani, F., & Steinhauer, K. (under review). First-language attrition induces changes in online morphosyntactic processing and re-analysis: An ERP study of number agreement in complex Italian sentences.

Kasparian, K., Postiglione, F., & Steinhauer, K. (under review). In My Other Language: An ERP Study of Italian-English Cognates and Homographs in L1-Attrition and L2-Acquisition.

Kerkhofs, R., Dijkstra, T., Chwilla, D. J., & de Bruijn, E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research, 1068,* 170–183

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

Kolk, H. H. J., & Chwilla, D. J. (2007). Late positivities in unusual situations. *Brain and Language, 100*, 257–261.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1-36.

Köpke, B. (1999). *L'attrition de la première langue chez le bilingue tardif : Implications pour l'étude psycholinguistique du bilinguisme*. Unpublished Doctoral Dissertation, Toulouse: Université de Toulouse-Le Mirail.

Köpke, B. (2002). Activation thresholds and non-pathological L1 attrition". In F. Fabbro (Ed.), *Advances in the Neurolinguistics of Bilingualism. Essays in Honor Of Michel Paradis* (pp.119-142). Undine: Forum.

Köpke, B., & Schmid, M. S. (2004). Language Attrition: The Next Phase. In M. S. Schmid, B. Köpke, M. Keijzer & L. Weilemar, L. (Eds.), *First Language Attrition: Interdisciplinary perspectives on methodological issue*s (pp. 1-43). Amsterdam: John Benjamins.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science, 207, 203–205.

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

Kutas, M., Lindamood, T., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes*. Hillsdale, NJ: Erlbaum.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338.

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

Libben, M. R, & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 381–390.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immerse in a second language. *Psychological Science*, *20*, 1507–1515.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71–102.

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

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

Montrul, S. (2008). *Incomplete acquisition in bilinguals: Re-examining the age factor.* Amsterdam: John Benjamins.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, *22*(2), 205–220.

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.

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure: An ERP study in adult learners. *BMC Neuroscience, 10*: 89.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience, 24*(5), 1205-1223.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Doupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: MIT Press.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-1228.

Opitz, C. (2011). *First language attrition and second language acquisition in a second language environment.* PhD dissertation. Centre for Language and Communication Studies. Trinity College Dublin. Dublin.

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

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech, *Language and Cognitive Processes, 8*(4), 413-437.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning*, *56*(1), 199–230.

Otten, M., & van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes 45*, 464–496.

Paradis, M. (1989). Bilingual and polyglot aphasia. In F. Boller, and J. Grafman (Eds.), *Handbook of Neuropsychology* (pp. 117–140). Amsterdam: Elsevier.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. De Groot and J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 331-354).Erlbaum, Mahwah, NJ

Pavlenko, A. (2000). L2 influence on L1 late bilingualism. *Issues in Applied Linguistics, 11*(2), 175–205.

Pelc, L. (2001). *L1 lexical, morphological and morphosyntactic attrition in Greek-English bilinguals*. Unpublished doctoral dissertation, City University of New York, United States.

Penfield, W. (1965). Conditioning the uncommitted cortex for language learning. *Brain*, 88, 787-798.

Penfield, W., & Roberts, L. (1959). *Speech and Brain Mechanisms.* New York: Athenaeum.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*(3), 407-428.

Pitkänen, I., Tanner, D., McLaughlin, J., & Osterhout, L. (in prep). Use it or lose it: Second language attrition in the brain looks like acquisition in reverse.

Rossi, S., Gugler, M., Friederici, A., & 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*(12), 2030–2048.

Rüschemeyer, S. A. (2005). The processing of lexical semantic and syntactic information in spoken sentences: Neuroimaging and behavioral studies of native and non-native speakers. MPI Series in Human Cognitive and Brain Sciences, 60.

Rüschemeyer, S., Nojack, A., & Limbach, M. (2008). A mouse with a roof? Effects of phonological neighbors on processing of words in sentences in a non-native language. *Brain and Language, 104*, 132–144.

Schmid, M. S. (2011). *Language attrition*. Cambridge: Cambridge University Press.

Schmid, M. S. (2013). First language attrition as a window to constraints on bilingual development. Keynote lecture at the International Symposium on Bilingualism (ISB9), Singapore, June 10, 2013.

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, 1–20.

Schmid, M. S., & Keijzer, M. (2009). First language attrition and reversion among older migrants. *International Journal of the Sociology of Language, 200*, 83–101.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240 –255.

Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., & Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia, 48*(6), 1525-1542.

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.

Thierry, G., & Wu, Y. J. (2007). Brain potential reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*, 12530–12535.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*(1), 37–69.

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., and Vissers, C. T. W. M. (2009). Monitoring in language perception. *Lang. Linguist. Compass* 3, 1211–1224.

van den Brink, D., Brown, C., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 Effects. *Journal of Cognitive Neuroscience, 13*(7), 967-985.

Van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*(3), 431–451.

van Herten, M., Chwilla, D. J., & Kolk, H. H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18*, 1181–1197.

van Herten, M., Kolk, H. H., Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*, 241-255.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 76–190.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. *Brain Research, 1106*, 150–163.

Vissers, C. T. W. M., Kolk, H. H. J., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: Evidence from ERPs in a picture–sentence matching task. *Neuropsychologia, 46*(4), 967-982.

Waas, M. (1996). *Language attrition downunder: German speakers in Australia*. Frankfurt: Peter Lang.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early settings of grammatical processing in the bilingual brain. *Neuron*, *37*, 159–170.

Weber-Fox, C., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language, 85*(2), 231-244.

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.

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain Responses before and after Intensive Second Language Learning: Proficiency Based Changes and First Language Background Effects in Adult Learners. *PLoS ONE*, *7*(12), 52318.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, in press.

Yağmur, K. (1997). *First language attrition among Turkish speakers in Sydney*. Tilburg: Tilburg University Press.

Yilmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon, 7*(3), 249-274.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*, 25-64.

6. Bridge 2

Lexical access and integration in L1-attriters and L2-learners: Within-language *and* between-language effects?

Study 2 tested lexical access and integration processes during online sentence comprehension in Italian-L1 attriters and late L2-learners of Italian, compared to non-attriting native-Italian monolinguals. We chose to examine the processing of minimal pairs because of their confusability in form (ambiguous until the final vowel, different gender and semantic meaning), which we expected would make these violations less salient or "trickier" to detect, particularly for L2 learners and/or L1 attriters with lower Italian proficiency levels. We found that lower Italian proficiency scores were associated with reduced N400 effects in response to both violation conditions relative to correct sentences. As expected, the proficiency effect was even more robust for *Swap* sentences than for *Mismatch* violations and, interestingly, proficiency modulated N400 responses in the absence of significant group effects, i.e., regardless of whether Italian was the participants' L1 or L2. Thus, the role of proficiency level in predicting ERP responses to language was further emphasized by our findings in Study 2, and the inclusion of the L2 learners allowed us to evaluate the notion of a proficiency-based continuum between L1 and L2 processing.

Similar to Study 1, we showed that attrition effects could not entirely be ascribed to effects of lower-proficiency, as the group of Attriters stood apart from Italian monolinguals and L2 learners with their enhanced P600 effects. We attributed the larger P600 amplitudes to a more attention-driven "second-thought" or attriters' double-checking the preceding input in order to diagnose the error, particularly in response to *Swap* anomalies. Given that both Study 1 and 2 reported differences in attriters' P600 responses (albeit in different time-windows of the P600 and thus reflecting different underlying processes of repair vs. conflict monitoring/diagnosis respectively), we were interested in whether P600 differences between groups might also be revealed in Study 3.

The focus of Study 3 was also on lexical-semantic processing but in English rather than Italian. Attriters were therefore tested in their late-acquired but predominantly-used L2, whereas English-Italian learners were tested in their L1. The two bilingual groups were also compared to native English monolingual controls. As in Study 2, we tested lexical access and integration of confusable words embedded in sentence contexts. However, while Study 2 had explored the potential of difficulties in lexical access *within* Italian, Study 3 tested whether English lexical-processing was subject to *cross-linguistic* influence from Italian, and thus the confusable words in Study 3 involved *homographs* (identical in form across languages but different in semantic meaning) and *cognates* (identical or nearly-identical in form and in semantic meaning).

The main goals of Study 3 were to determine whether, in cases of cross-linguistic ambiguity, word-meanings from the non-target (i.e., Italian) language would be activated in parallel during online English processing, and whether these effects might differ for the two bilingual groups, given the opposite direction of crosslinguistic transfer (L1 to L2 in Italian-attriters, and L2 to L1 in English-Italian learners). A second aim was to determine whether co-activation of attriters' L1 during L2 processing would be modulated by factors associated with attrition, such as high L2 proficiency, lower L1 proficiency, length of residence, amount of exposure to L1 and L2, etc. Like in Study 3, we examined modulations in amplitude, scalp distribution and timing of N400 and P600 effects as a measure of how lexical-semantic anomalies of different degrees of conflict were detected and resolved during online sentence processing.

7. MANUSCRIPT 3

In My Other Language: An ERP Study of Italian-English Cognates and Homographs in L1-Attrition and L2-Acquisition

Kristina Kasparian[1,2], Francesca Postiglione[3] and Karsten Steinhauer[1,2]

[1] School of Communication Sciences & Disorders, McGill University, Montreal, Canada
[2] Centre for Research on Brain, Language and Music (CRBLM), Montreal, Canada
[3] Dipartimento di Psicologia e Scienze Cognitive, Università degli studi di Trento, Italy

This manuscript has been submitted to PLOS-ONE for publication and is currently under review

ABSTRACT

Using event-related-potentials (ERPs), we examined lexical-semantic processing of sentences containing Italian-English homographs and cognates in (1) English monolingual speakers; (2) Italian-English "*attriters*" (highly-proficient and predominantly-exposed to late-acquired L2 while reporting changes in L1); and (3) English-Italian learners (highly-proficient in late-acquired L2 but still L1-dominant). Sentence-contexts either constrained the English meaning of the homograph, the Italian meaning of the homograph, the dual-language meaning of the cognate, or were semantically implausible in either language. In line with language non-selective models of bilingual lexical access, we observed co-activation of Italian during an English experiment, not only in the N400 time-window but also on the P600. At early processing-stages, sentences that highly-constrained the Italian homograph meaning were perceived as less of a lexical-semantic violation (reduced N400), particularly for Italian-English bilinguals, thus indicating interference of the Italian meaning in an English task. A cognate facilitation effect was found for both groups of bilinguals, though in different processing-stages; while Italian-English Attriters showed reduced N400 *and* P600 effects for correct cognate vs. correct homograph conditions, English-Italian learners only showed this facilitation effect on the P600, and tended instead to show cognate-interference on the N400. In both bilingual groups, homograph-targets elicited larger posterior P600s relative to cognate-targets – a novel finding in sentence-contexts which we discuss as reflecting a conflict-monitoring process during a late "language-selection" stage of lexical access. These effects were not mirrored by offline acceptability ratings. Crucially, co-activation was modulated by proficiency (both in English, as well as English *relative* to Italian), and by language-exposure and length-of-residence in Attriters, but not by age-of-acquisition. By turning to L1-attrition, we examine how a shift in dominance and proficiency from L1-to-L2 affects lexical-access and integration during real-time comprehension. Our findings support claims of ongoing neuroplasticity for language-learning in adulthood, and of proficiency and exposure determining the brain's responses to language.

1. INTRODUCTION

In an increasingly multilingual world, research into mysteries and possibilities of multilingualism has flourished over the last few decades. How second languages (L2) are learned, how they may interact with one's first language (L1), and how multiple languages are organized and processed relative to one another in the brain are questions that continue to garner attention. Infused in most of the interdisciplinary research on L2 acquisition is the claim in favor of a neurobiological "critical-period" for language-learning, after which the brain is believed to lose the plasticity necessary to acquire and process language at a native-like level (Lenneberg, 1967; Penfield & Roberts, 1959). A number of neuroimaging studies have shown that L2 learners with a late *age-of-acquisition* (AoA) deviate from early-learners and from native-speakers in the neurocognitive mechanisms used to process the L2 (Kim et al, 1997; Hahne & Friederici, 2001; Weber-Fox & Neville, 1996). Until relatively recently, however, *proficiency level* was a neglected factor in many studies and was typically confounded with AoA, making it difficult to disentangle whether processing differences between L2 and L1 speakers were indeed due to a late AoA or rather to a low level of L2 proficiency.

Even when researchers began to consider proficiency, it was often quantified only through participants' own self-report ratings or indirectly defined by months/years of classroom instruction. Studies that have more systematically examined the impact of proficiency level have provided strong evidence that, with increasing proficiency, L2 learners show processing profiles that converge upon and are indistinguishable from those of native-speakers, even if the L2 was acquired in adulthood (Bowden et al., 2014; Friederici, Steinhauer & Pfeifer, 2002; Hopp, 2010; Osterhout et al., 2006; 2008; Perani et al., 1998; Rossi et al., 2006; Steinhauer, White & Drury, 2009; White, Genesee & Steinhauer, 2012). Cross-linguistic influence from bilinguals' L1 to their L2 has also been shown to be modulated by proficiency level, with low proficiency L2 learners more prone to transfer or interference from the L1 and, consequently, a less native-like L2 (e.g., Ojima, Nakata & Kakigi, 2005; Tokowicz & MacWhinney, 2005). Thus, while cross-linguistic similarity has been deemed an important predictor of the kinds of linguistic structures that are likely to be processed in a native-like-way in late L2 learners, proficiency has been shown to modulate these effects. Further evidence of the crucial role of proficiency in determining the brain's responses to language comes from studies of *monolingual* native-speakers in whom, obviously, AoA plays no role (Pakulak & Neville, 2010; Prat, 2011). Despite these findings, however, the debate is still controversial, as others argue that the role of AoA cannot be disregarded (e.g., Clahsen & Felser, 2006; Hahne, 2001; Ojima et al., 2005).

Another potentially influential factor is language *dominance*, which is related to the amount of exposure and use of the L2 comparatively to the L1. Although a number of studies with L2 learners have been conducted in L2 (rather than L1) environments (e.g., Chinese learners of English in Canada in White, Genesee & Steinhauer, 2012; Spanish learners of English in USA in Tanner, Inoue & Osterhout, 2013), participants' *L1* proficiency, exposure and use relative to their L2 was typically not assessed in a way to determine whether participants tested were still dominant in the L1 and continued to use their native-language on a daily basis, although they were highly-advanced in the L2. In sum, it is currently unresolved whether it is AoA that shapes the neurocognitive basis of L2 processing, or whether other experiential factors (proficiency level, typological similarities/differences, or amount of exposure/use) are stronger predictors of "native-like-ness" in L2 processing.

For a new perspective on the problem, we turn to a special group of late learners – "*first language (L1) attriters*" who are not only highly-proficient in the late-acquired L2 after having immigrated to the L2 environment, but who additionally experience a shift in dominance from the L1 to the L2, as they have little or no use of their L1 on a daily basis. Immersion into the L2 environment with limited use of the L1 gradually has been shown to result in changes to the L1 that are often perceptible to attriters themselves – a phenomenon called "*attrition*" (for reviews, see Schmid and Köpke, 2007; Schmid, 2011). The present study does not directly examine L1 attrition as it does not assess L1 processing in these individuals, but rather their L2 (but see our other studies in Kasparian, Vespignani & Steinhauer, under review; Kasparian, Vespignani & Steinhauer, 2013a;b; Kasparian, Vespignani & Steinhauer, 2014a;b). However, given their unique L2-learning circumstances, "attriters" are a critical bilingual group in order to address questions of neuroplasticity, critical-periods and the relative impact of factors such as AoA, proficiency, exposure, or other experiential factors on L2 processing, and thus we refer to them as "attriters" even in the current paper (rather than simply as late L2 learners of English).

Using event-related-potentials (ERPs), our goal in the present study was to investigate lexical-semantic processing in (1) English monolinguals, (2) Italian-English Attriters (i.e., late L2 learners of English) and (3) English-Italian bilinguals (i.e., late L2 learners of Italian), using a design that tested whether bilingual participants' Italian lexicon would be implicitly co-activated though reading exclusively English sentences, during an English experimental session, in an English environment. Crucially, we aimed to determine whether the two groups of bilinguals (with the same language pairs but opposite AoA profiles) might co-activate Italian to a similar degree, and whether this lexical co-activation would be modulated not only by their proficiency-level in English (target-language) but particularly by Italian-English *relative* proficiency. We expected that L1-to-L2 transfer or co-activation would decrease with advancing attrition, i.e., with increasing L2 proficiency and decreasing L1 proficiency. Moreover, we examined whether Attriters would show "native-like" lexical-semantic processing compared to English monolinguals and English-Italian bilinguals, especially for conditions where our design did not explicitly promote cross-linguistic co-activation of Italian.

1.1. Lexical access in bilinguals

A fundamental question in bilingualism research has centered upon whether the bilingual lexicon stores and accesses words in a *language-selective* way (e.g. Scarborough, Gerard & Cortese, 1984; Soares & Grosjean, 1984), or in a *parallel, non-selective* process that contemporaneously activates representations of both languages during early stages of word identification, irrespective of the target-language (e.g., "*Bilingual Interactive Activation (BIA+) Model*", Dijkstra & Van Heuven, 2002; "*Revised Hierarchical Model (RHM)*", Kroll & Stewart, 1994; Van Hell, Tokowicz & Green, 2010).

As evidence for non-selective lexical access, a number of studies have attempted to show that late L2 learners activate L1 translation equivalents when processing L2 words (e.g. Potter et al., 1984; Keatley et al., 1994; Basnight-Brown & Altarriba, 2007; Sanchez-Casas et al., 1992). The vast majority of research has focused on a particular category of words that are similar across bilinguals' languages. *Interlingual homographs* are words which share an identical orthographic form but are distinct in their semantic meanings (e.g., English-Italian: *estate* (= *property* vs. *summer*); English-French: *coin* (= *money* vs. *corner*); English-Dutch: *room* (= *space* vs. *cream*). In contrast, *cognates* are identical or nearly-identical in their form, but also share a semantic meaning (e.g., English-Italian: *music*/*musica*; English-French: *angle*/*angle*; English-Dutch: *baker*/*bakker*). The rationale behind investigating the processing of these types of words is to precisely capitalize on their cross-linguistic similarity; if cognates and homographs are found to be processed differently from words that bear no such resemblance across the two languages, the likely inference is that the representations of these words in both languages are activated in parallel, resulting in either facilitation (due to convergence of meanings) or conflict (due to divergence of meanings) in processing the relevant target-language.

In line with this hypothesis, the main finding across early behavioral studies was that bilinguals generally show facilitation effects (i.e., faster and more accurate processing) for cognates, but interference effects for homographs, compared to matched control words (De Groot & Nas, 1991; Dijkstra, Grainger & Van Heuven, 1999; Dijkstra, Van Jaarsveld & Ten Brinke, 1998; Dijstra, De Bruijn, Schriefers & Brinke, 2000; Duyck, Van Assche, Drieghe & Hartsuiker, 2007; Lemhofer & Dijkstra, 2004; Libben & Titone, 2009; Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen, 2010; Sanchez-Casas, Davis & Garcia-Albea, 1992; Schwartz & Kroll, 2006; Van Hell & Groot, 2008; Van Hell & Dijkstra, 2002). The direction and strength of facilitatory/inhibitory effects, however, have been shown to be affected by a number of methodological factors, such as the composition of word lists (Dijkstra, Van Jaarsveld & Ten Brinke, 1998; Dijkstra, Timmermans & Schriefers, 2000), the nature of the task and whether bilinguals were placed in a monolingual or dual-language context (Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen, 2010; Dijkstra, Van Jaarsveld & Ten Brinke, 1998), as well properties of the stimuli, such as the degree of orthographic/semantic/phonological overlap (Dijkstra, Grainger & Van Heuven, 1999; Dijkstra et al., 2010). For example, cognates may cause *inhibition* rather than facilitation if mixed with interlingual homographs in a language-identification task (Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen, 2010), while homographs may suddenly benefit from a processing *advantage* if the task requires bilinguals to decide if a word exists in *either* one of the languages (Dijkstra, Van Jaarsveld & Ten Brinke, 1998).

Although parallel activation of the non-target language was an overwhelming finding in studies with single words, such experimental designs initially neglected potential effects of context. Evidence of crosslinguistic activation in *sentence* contexts in the target-language has provided even stronger support for non-selective lexical access (Duyck, Van Assche, Drieghe & Hartsuiker, 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Assche, Drieghe, Duyck, Welvaert & Hartsuiker, 2010; Van Hell & De Groot, 2008). Moreover, these studies have shown that the degree of sentence-constraint modulates the strength of co-activation. For example, while cognates showed facilitation in low-constraint sentences, their processing advantage was attenuated in high-constraint sentences (Duyck et al., 2007; Elston-Güttler, 2000;

Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Hell & De Groot, 2008, but see Van Assche et al., 2010 for cognate facilitation effects regardless of sentence-context).

Using eye-tracking techniques or event-related brain potentials (ERPs), researchers have increasingly begun investigating lexical co-activation effects during real-time processing, i.e., as comprehension unfolds. Such measures are likely to tap into more automatic and implicit effects that may have decayed by the time an "offline" behavioral judgment at the end of stimulus presentation (see Thierry & Wu, 2007). Moreover, insights can be gained on the time-course of these effects (e.g., see Libben & Titone, 2009 for evidence of co-activation in sentence contexts at early- but not late-stages of processing).

In ERP studies, facilitation in lexical processing has been reliably associated with a reduction in the N400 component (Kutas & Hillyard, 1980; 1984), the amplitude of which is typically also reduced when words are more highly-frequent, more highly-familiar, primed by semantically-related words, or embedded in highly-constraining sentences where the target-word is predictable (see Kutas & Federmeier, 2011 for a review). ERP investigations have replicated the cognate facilitation effect in paradigms using lists of single words (Midgley, Holcomb & Grainger, 2011; Peeters, Dijkstra & Grainger, 2013), and have also shown semantic priming effects between homographs and their L1 meanings (Elston-Güttler, Gunter & Kotz, 2005) or L2 meanings (De Bruijn et al., 2001; Kerkhofs et al., 2006), whether the homographs were presented as the prime (De Bruijn et al., 2001) or as the target (Kerkhofs et al., 2006), in sentence contexts (Elston-Güttler, Gunter & Kotz, 2005) or in pairs/triplets of words (De Bruijn et al., 2001). ERP studies have also corroborated the claim that context – be it the "global" context of the experimental session or the "local" context of sentence constraint – modulates the extent of co-activation of the non-target language, and that such effects can be reduced or eliminated if the target-language context highly-constrains the target-language meaning of the word (Elston-Güttler & Gunter, 2008; Elston-Güttler, Gunter & Kotz, 2005).

However, a gap in the literature to date is that studies have not systematically investigated whether priming the *L1*-meaning of a crosslinguistically ambiguous word in an L2 experiment leads to enhanced L1 co-activation, and how context may modulate this effect. For example, priming studies showing that homographs prime or are primed by their L2-meanings do not say much about *co-activation* of the non-target language and whether context modulates access to the *L1*-concept, if they did not test priming between the homograph forms and their L1 semantic-meanings (e.g., De Bruijn et al., 2001; Kerkhofs et al., 2006). Even studies that embedded homographs in sentence-contexts and manipulated the degree of sentence constraint did not test sentences that *exclusively* supported the *L1*-meaning of the ambiguous word rather than also/only supporting its target-L2 meaning (e.g., Elston-Güttler, Gunter & Kotz, 2005; Elston-Guttler, Paulmann & Kotz, 2005).

1.2. Proficiency effects on bilingual lexical access

According to models of bilingual lexical access (e.g., *BIA+ Model*: Dijkstra & Van Heuven, 2002; *RHM:* Kroll & Stewart, 1994), another factor likely to determine the selectivity of lexical access in bilinguals is L2 proficiency-level. Such models posit that increasing proficiency leads to stronger form-meaning links and, thus, better efficiency in inhibiting non-target semantic meanings (e.g., Jared & Kroll, 2011). However, proficiency was not explicitly considered in the majority of studies, or participants' proficiency was quantified only by their self-report ratings, which are likely to be insensitive to proficiency differences in specific linguistic areas or on specific structures, in addition to potentially being unreliable. As highlighted by Van Hell and Tanner (2012), only a handful of studies have systematically examined the impact of proficiency by using appropriate experimental designs – for example, by testing (1) the same bilinguals both in their L1 or L2, (2) bilingual speakers with the same L1 but with differing levels of L2 proficiency, or (3) bilingual speakers of the same language pair combination but where groups differ on their L1 (e.g., English-French vs. French-English).

Studies examining L2 lexical processing have shown that cross-linguistic activation is more pervasive in the *weaker* language (i.e., from the L1 to the L2 when the task is performed in the L2). The weaker the language, the stronger the observed interference effects when word-forms with divergent semantic meanings compete for selection (Elston-Güttler, Paulmann & Kotz, 2005), and the stronger the facilitation effects for cognates (Costa et al., 2000; Libben & Titone, 2009; Poarch & Van Hell, 2012). A minimal level of L2 proficiency seems to be required, however, for cognate facilitation to occur at all, i.e., for learners to *recognize* L2 words as L1 cognates (Van Hell & Dijkstra, 2002). In fact, at initial stages of learning, cognates may even create an interference effect, especially if intermixed with homographs, and may be recognized more slowly than non-cognate controls (Brenders, Van Hell & Dijkstra, 2011). Furthermore, proficiency-level seems to interact with sentence-constraint, with reduced or eliminated interference effects in highly-constraining sentence contexts for high- but not low-proficiency learners (Elston-Güttler, Paulmann & Kotz, 2005), suggesting that sentence-context enables L2 learners to better control L1 activation in an L2 task, but only if their L2-proficiency is sufficiently high.

In the majority of this research, however, the weaker language invariably coincides with the L2. A few studies have turned to situations where bilinguals had been immersed in their L2-environments for an extended period of time, and have revealed that co-activation effects tend to be bidirectional in bilinguals with more balanced proficiency levels in their languages (Basnight-Brown & Altarriba, 2007; Costa et al., 2000; Duñabeitia, Perea & Carreiras, 2010; Perea et al., 2008; Zhao, Li, Liu, Fang & Shu, 2011). Although such findings suggest that the co-activation of the non-target language is not only modulated by proficiency but also by relative-dominance of the bilingual's languages, several of these studies tested early simultaneous bilinguals (e.g., Basnight-Brown & Altarriba, 2007; Duñabeitia et al., 2010; Perea et al., 2008), or individuals who were still L1-dominant though living in the L2-environment (e.g., Costa et al., 2000), in which case it is not possible to disentangle whether co-activation effects are due to L1-L2 status, AoA or language dominance, all of which are confounded. The next logical step, then, is to further extend this research to cases of late AoA where the L2 is not only highly-proficient, but also where the dominance shifts from the L1 to the L2.

1.3. Lexical access in L2-dominant attriters

First-language (L1) attrition has been defined as a non-pathological decline in a previously native-like L1, due to a particular linguistic situation where the individual's contact with the L1 community is limited or severed and the L2 becomes to predominantly used language (see Köpke, 2002; Köpke & Schmid, 2004). In our study, *attriters* are operationally defined as first-generation immigrants who lived in an exclusively monolingual L1 environment until adulthood (> 28 years) before moving to a new country where, due to full immersion into the majority-L2 environment and limited use of their minority-L1, they experienced a shift in dominance from the L1 to the L2. Similarly to typical L2-learners, proficiency in the L2 gradually increases, but this comes at a cost to their L1, due to their special socio-linguistic circumstances. Attriters typically report experiencing a decrease in automaticity, fluency and native-like proficiency in their L1, whereas their proficiency level in the late-learned L2 continues to strengthen. This interplay in relative L1-L2 dominance and proficiency makes attriters different from the bilingual populations that are typically recruited in L2 processing studies, including the literature reviewed above.

Characteristics of attriters' L1 have been extensively explored in behavioral research, revealing deviations from native linguistic performance, in a number of domains but particularly in vocabulary (e.g., de Bot, 1996; Köpke, 1999; Montrul, 2008; Opitz, 2011). There is some evidence that attriters show slower access or smaller/less-diverse L1 vocabularies in tasks such as picture-naming, verbal semantic fluency or free-speech (De Bot, 1996; Köpke, 1999; Linck et al., 2009; Schmid & Jarvis, 2014; Schmid & Keijzer, 2009; Waas, 1996; Yagmur, 1997; Yilmaz & Schmid, 2012), as well as indications of L2-to-L1 transfer/competition in the domain of lexical access in reports of increased lexical borrowing or semantic-intrusions from the L2 (Pavlenko, 2000). Even anecdotally, attriters may report having noticed changes in their L1, and these reported changes often center on L2 effects on L1 vocabulary and/or grammar (Kasparian, Vespignani & Steinhauer, 2013b). In fact, a common complaint among the attriters in our own series of studies (prior to participation and without any knowledge of the linguistic structures that would be tested) was the pervasiveness of "false-friend" errors and other kinds of interference from their L2 (English) vocabulary onto their L1 (Italian), especially when words are similar but diverge in semantic meaning.

If cognate-facilitation and homograph-interference effects in bilinguals have been found to be more robust when processing the weaker language (i.e., from L1 to L2) and have been shown to be bidirectional only when L2-proficiency is sufficiently advanced, what happens in L1-attriters who are highly-proficient in the L2 but are also potentially less native-like in their L1 compared to other L1-dominant bilinguals? Do attriters still show L1-to-L2 effects and, if so, is this modulated by proficiency level, or does their L2-dominance preclude co-activation effects in the L1-to-L2 direction, making them more similar to native-speakers of that language?

While the field is replete with behavioral studies and anecdotal reports, neurocognitive investigations of language-processing (L1 or L2) in attriters are only now beginning to contribute to the picture (Kasparian, Vespignani & Steinhauer, under review; Schmid et al., 2013).

1.4. The present study

Our study tested processing of Italian-English homographs and cognates embedded in English sentences in English native-monolinguals and two groups of bilinguals with the same language pairs but opposite AoA profiles, namely (1) Italian-English attriters, dominant and highly-proficient in their L2, and (2) English-Italian L2 learners, dominant in their L1 but advanced in L2 proficiency. Attriters were thus tested in their dominant-L2 (in an L2 experiment in the L2 environment), whereas English-Italian learners were tested in their dominant-L1. As summarized in **Table 1**, sentences were either semantically-plausible in English only (EH = English reading of homograph), plausible in Italian only (IH = Italian reading of homograph), plausible in both (CC = cognate) or plausible in neither (EC, IC, CH).

Table 1 about here

1.4.1. Research questions and hypotheses

(1) In terms of processing the L2 (Italian-English Attriters), we were interested in the following questions:

- Replicating previous findings of non-selective bilingual lexical access, do homograph conditions (EH, IH) cause interference and cognates (CC) cause facilitation in behavioral performance and/or ERP responses?
- If co-activation from L1 to L2 occurs in Attriters despite their L2-dominance and "attriting" L1, are these effects mediated by factors such as proficiency, AoA, language use and/or length of residence (LoR) in the L2 environment?
- On conditions that do not promote co-activation between the two languages (EC, IC, CH), are Italian-English Attriters indistinguishable from native-English monolinguals and English-Italian bilinguals in the amplitude, latency and distribution of their ERP responses?
- Are L1-to-L2 effects correlated with *L2* proficiency level or rather on *relative* proficiency level (L1-L2), given that Attriters are a unique subgroup of bilinguals with more variability in their L1?

(2) In terms of processing the L1 (English-Italian learners):

- Do English late-learners of Italian show co-activation effects in the L2-to-L1 direction (i.e., interference for homographs, facilitation for cognates), and does this depend on AoA and/or L2-proficiency?

(3) ERP modulations and time-course of co-activation:

- Are co-activation effects reflected by modulations of the N400 component?
- Is a P600 effect *also* elicited in response to lexical-semantic violations, reflecting more controlled processes such as conflict monitoring and revision (van de Meerendonk, Kolk et al., 2009; 2010; Vissers et al., 2006)?
- If a P600 occurs, does it occur only in some conditions (e.g., homograph interference) rather than others (e.g., cognate facilitation)?
- Do the ERP effects mirror the patterns observed in behavioral responses, or are ERPs more sensitive to implicit co-activation effects than offline acceptability judgments?

We expected English monolingual controls to show a large N400 effect, potentially followed by a P600 effect, in response to all lexical-semantic violation conditions (EC, IH, IC, CH) relative to correct sentences (EH, CC), with no differences across the four violation conditions, nor across the two correct conditions. We predicted that both bilingual groups (Attriters and Italian-learners) would show some degree of Italian co-activation during English reading, in the N400 window and/or during later stages of processing, especially in those cases where the English context did not strongly constrain co-activation, namely in the IH condition (where the translation of the sentence would be semantically-plausible in Italian), and the CC condition (where we expected facilitation due to the doubly correct nature of the cognate sentence contexts). More specifically, we anticipated a reduction in the N400 effect for IH sentences relative to other violations, and a reduction in the correct CC condition relative to correct EH sentences (which are actually incongruous with the Italian reading of the homograph). Thus, we expected co-activation effects to occur both in the direction of the L1 to the L2 (and vice versa) in those conditions where we explicitly promoted it.

Crucially, we expected that proficiency-level but not AoA would modulate ERP responses in both groups of bilinguals, regardless of whether Italian was the L1 or the L2. We also expected that relative L1-L2 proficiency would be a significant predictor of the degree of co-activation, with less co-activation (more native-like ERP patterns) occurring in Attriters with a larger asymmetry in proficiency scores favoring the L2. Finally, given that homographs and cognates have a detectable dual-language status, we could have expected a target-word effect in the two bilingual groups, whereby merely recognizing that a homograph in an English context is identical in form to an Italian word might contribute to additional orthographic and/or interference effects for homographs in general, irrespective of whether the sentence was semantically plausible in English or not.

1.4.2. Novelty and methodological considerations

The present study aimed to extend the literature on bilingual lexical processing by not only turning to a unique population of bilinguals (i.e., L1-attriters) but also including two groups of bilingual speakers of the same language pair, but different L1s – a design advantage uncommon in the literature so far (as discussed in Van Hell & Tanner, 2012). In doing so, we are able to assess the differential contribution of L1 and L2 proficiency effects, AoA, and language exposure, in a design where L1-L2 status is not confounded with language dominance/use, and where proficiency in the late-acquired L2 is not by default lower than proficiency in the native-L1. Furthermore, the present study scrutinized participants' L1 *and* L2 proficiency by means of multiple comprehension and production tasks, administered also to monolinguals. The language pairs we chose also allowed us to assess whether crosslinguistic dissimilarity between Italian and English (in contrast to Dutch and English – the most widely-studied language pair in the field to date) might impact our findings, relative to previous studies.

Another contribution of the present study is the inclusion of a sentence context *exclusively* favoring the non-target reading of the homograph (IH). Although a similar manipulation was done by De Bruijn and colleagues (2001) with triplets of words for which they had found no significant modulations of context (i.e., the language of the first word), contextual cues are arguably stronger in a sentential context. An additional advantage of our design (Table 1) is that its six conditions prevented a particular type of target word to be consistently linked to the same type of response (in contrast to experiments using lexical decision tasks, for example). Instead, our design balanced the contexts in which homographs and cognates appeared, such that homographs and cognates equally contributed to sentences that were violations in English (E<u>C</u>, I<u>C</u>, I<u>H</u>, C<u>H</u>), and behavioral responses could not therefore be predicted on the basis of the target-word alone.

A related point concerns our choice of task – in line with much of the ERP research on sentence-processing (but in contrast with most of the literature on bilingual lexical processing), participants performed an end-of-sentence acceptability judgment task rather than a lexical decision task or a dual-language task such as translation, generalized LDT, language-identification, etc. Moreover, rather than asking participants to provide binary yes/no judgments, we asked them to *rate* (on a scale from 1 (completely unacceptable) to 5 (perfect)) the acceptability of sentences they read. We expected that this task might tap into processes of both early automatic lexical access but also semantic-integration of the word into the sentence context, and that having to rate the sentences might lead participants to pursue semantic processing in more depth, comparatively to deciding whether a word was legal or not. Given our task, we thought it plausible to find co-activation effects beyond early time-windows such as the N400, and were interested in exploring the effects of homographs and cognates on later stages of processing associated with forming an acceptable interpretation of the sentence. For example, we might expect an enhanced P600 effect for interference conditions in bilinguals, compared to instances of cognate facilitation where there is no conflict to resolve at a later stage of processing. The P600 effect has not been examined in studies of bilingual lexical access; only one study by Peeters, Dijkstra & Grainger (2013) reported a significant difference between cognate words and control words in the P600 time-window for control-words, where cognates elicited a significantly smaller positive-going wave than control words.

In sum, in addition to investigating questions of neuroplasticity, the native-like-ness of L2 processing, the non-selective nature of the bilingual lexicon and the special profile of L1-attriters, it was also of interest to explore how the novel elements of our experimental design (participant groups, language pairs, sentence stimuli, proficiency measures, task effects and stages of processing examined) may affect patterns of lexical co-activation in bilinguals, thus further advancing the extensive literature on the topic so far.

2. METHODS

2.1. Participants

Three groups of participants were tested in English: (1) Italian-English attriters; (2) English-Italian late L2 learners; and (3) English native-monolinguals (see *General Methods*, p. 47-50 for demographic information). Although the present study focuses on the processing of their late-L2 English, we continue to refer to this group as "Attriters" rather than simply as "late L2 learners of English" for two reasons: (1) to emphasize that, unlike most studies of L2 processing, these individuals are living in the *L2* environment, and are not only advanced in L2 proficiency but also dominant in the L2, with little/no exposure or use of the L1 (in contrast to the majority of late L2 learners still dominant in their L1), and (2) to make the link clearer between this paper and our other (forthcoming) papers involving the same populations, in both Italian and English.

2.2. Behavioral measures

Although Study 3 was conducted in English, the behavioral measures were analogous to the Italian measures reported in Studies 1 and 2 (see *General Methods*, p. 50-54). Group means are provided in **Table 2.**

Table 2 about here

Particularly relevant to the present study was the bilingual groups' performance on the verbal translation task described in *Section 3.9.2* in the *General Introduction* (p. 46-48). Although a detailed description of these experimental stimuli and results is beyond the scope of this dissertation (but see Kasparian, Vespignani & Steinhauer, 2013b for accuracy data and a description of errors), given the purposefully cross-linguistic nature of the production task, bilingual participants' scores were used as an additional measure of proficiency in the present study. In doing so, our aim was to assess whether a task tapping into a *specific* area of these bilinguals' language processing (rather than overall proficiency measures) would predict their ERP responses during the processing of similar structures, despite several differences in task (i.e., production vs. comprehension, dual-language translation task vs. English reading task; isolated words vs. sentences).

Although our recruitment process targeted Attriters who were highly proficient and dominant in the English-L2, some proficiency measures revealed differences between the Italian-English Attriters and the two groups of native-speakers. Group differences were significant for the *Error-detection task* ($F(2,72) = 18.40$, $p < 0.01$), the *verbal semantic fluency task* ($F(2,72) = 7.60$, $p < 0.05$), as well as the *reading fluency task* ($F(2,72) = 15.38$, $p < 0.01$). Bonferroni-corrected comparisons revealed that, in each case, Attriters as a group differed from native-

English Controls as well as from English-Italian learners (all *p*s < 0.05), while the monolingual and bilingual native-English groups did not differ significantly from each other on any of the measures (*p*s > 0.1). Interestingly, Attriters did not differ from the native-English speaker groups on the *C-test* (*p* > 0.1) – a finding that emphasizes the importance of administering more than one type of proficiency measure in ERP studies of L2 processing. Note, however, that one of our principal aims was to investigate the impact of proficiency level on the "native-like-ness" of processing patterns and the degree of co-activation from the Italian lexicon, and matching the groups on proficiency measures would not have allowed us to achieve this. Attriters in the higher ranges of English-proficiency were within the native-English speaker range, but the most difficult (i.e., most sensitive) measure was the written *Error-detection task* (max. score in Attriters = 84.3%, while in Italian learners = 98.0 % and English-Controls = 96.1%).

To obtain our measure of "*relative proficiency*" we created an overall proficiency score for each language, by adding the standardized scores of the three proficiency measures. We then subtracted the overall English proficiency score from the overall Italian proficiency score.

2.3. Stimuli

The target nouns in this experiment were either **(1)** ***interlingual (non-cognate) homographs*** between English and Italian – i.e., pairs of words that share the same orthography in both languages, but have a different semantic meaning (e.g. *estate* (= "property" in English vs. "summer" in Italian), or **(2)** ***interlingual cognates*** between English and Italian – i.e., pairs of words that share a similar or identical orthography, as well as the same semantic meaning (e.g. *music/musica, dune/dune*).

2.3.1. Homograph target nouns

The set of interlingual homographs contained *only* "false-friends" or false-cognates that were identical in orthography. In order to avoid overlap with the "false-friends" production task that was administered at the start of the experimental session, we avoided false-friends that differed in orthography (e.g. *joke* (joke) / *gioco* (game)). We also avoided pairs of nouns that overlapped in their semantic fields, as it would be difficult to create a sentence context that highly constrained the meaning of only one word in the homograph pair (e.g. *marina* ("*dock for boats*" in English, "*seashore/navy*" in Italian), or *crude* ("*vulgar/unrefined*" in English, "*raw*", in

Italian). We restricted our selection to pairs where the homograph was a noun in both languages (rather than selecting noun-verb or noun-adjective pairs), so as not to create word-category violations in addition to lexical-semantic violations (e.g. *prove* (= "confirm" in English, "trials" in Italian). These rather stringent selection criteria permitted us to select 30 homograph targets.

2.3.2. Cognate target nouns

A set of 30 interlingual cognates was selected by matching homograph nouns pair-wise on length and English lemma frequency (English Lexicon Project, Balota et al., 2007). Half of the selected cognates were also homographs (thus, identical in their orthography *and* semantic meaning, e.g. *diva/diva, idea/idea*), whereas the other half of the cognates differed slightly in their orthographic form between English and Italian due to language-specific properties such as the fact that the majority of English nouns end in a consonant, whereas Italian nouns end in a vowel (e.g., *concert/concerto, ocean/oceano*). This information was coded in order to eventually be able to analyze the effect of identical or non-identical orthography within the set of cognates.

In addition to length and English lemma frequency, we obtained the Italian lemma frequency information (CoLFIS database, Bertinetto et al., 2007) as well as orthographic neighborhood measures in English (English Lexicon Project, Balota et al., 2007) and Italian (Mulatti & Andriolo, in prep) for each target word. Overall, homographs and cognates did not differ in English word length ($p = 1$), English lemma frequency ($p = 0.25$), Italian word length ($p = 0.18$), Italian lemma frequency ($p = 0.94$) and Italian orthographic neighborhood ($p = 0.14$). The only dimension on which we were not able to match the two types of target nouns was on English orthographic neighborhood ($p < 0.05$), where the homographs had a larger number of orthographic neighbors ($M = 14.13$) than the cognates ($M = 9.3$). We attribute this to the fact that many of the homographs in English were four letter, monosyllabic words ending in "e" (e.g. cane, lane, pane), which makes them rhyme with many other words and, therefore, results in a higher number of orthographic neighbors. Despite these inevitable differences, however, English native-Controls showed no target-word effects in their ERP responses.

2.3.3. Experimental sentences

The goal was to create either perfectly plausible or semantically anomalous English sentences, which may or may not have been plausible in Italian. Our design was a 3x2 design, depending on combinations of *Context* (3 levels: priming either the English (E) or Italian (I) meaning of the homograph, or the cognate (C) meaning) and *Target* (2 levels: homograph (H) or cognate (C)). For sentence examples of the six experimental conditions, please refer back to **Table 1**.

Given that the experiment was conducted in English, only conditions EH and CC can be considered semantically congruent, while the other 4 conditions are lexical-semantic violations in English. However, the IH condition, while a violation in English, would be semantically congruent in its Italian reading and potentially acceptable if a bilingual speaker were co-activating the Italian lexicon (thus balancing out the proportion of semantically plausible vs. anomalous sentence conditions).

A set of 90 sentences was created without repetitions of any verbs or nouns. Note that although we had only 30 homographs, they could be considered pairs given their different English and Italian meanings (thus resulting in 60 homograph sentences, half priming the English (EH) meaning and half the Italian (IH) meaning). Lexical-semantic violations were created by first pair-wise matching each homograph pair with a cognate (e.g., *cane* (walking stick vs. dog) and *idea*), and then by replacing the target-noun that was to appear in a specific sentence context (EH → EC; CH → CC and so on). Sentences were eight words long and were identical in grammatical structure. The target noun was consistently the fifth word in the sentence, preceded by the determiner "the", and followed by a prepositional phrase (two function words and a noun), in order to (1) avoid sentence wrap-up effects associated with the presentation of target-words in sentence-final positions, and (2) to minimize any baseline or post-target differences that could be confounded with ERP effects of the target-word. In all sentences, the verb preceding the target noun phrase was a transitive verb in the simple past tense. Although the sentence contexts were not identical across the 6 conditions, we attempted to minimize lexical differences on the verb preceding the target noun by ensuring that the verbs did not differ in lemma frequency or length across context types (ps > 0.1). All sentences were created by the first author of the study and double-checked by a second English native-speaker who did not take part in the study.

Although investigating effects of cloze probability was not an explicit aim of the current study, we obtained and coded cloze probability values for our experimental sentences for two reasons: (1) cloze probability has been viewed as the most important factor in determining N400 amplitudes of content words; and (2) as discussed in the introduction, studies on homograph- and cognate-processing have shown that sentence-constraint influences observed effects of co-activation. A sentence-completion task was administered to 24 English native-speakers who did not participate in the study. These individuals were asked to complete each of the "correct" (EH, IH, CC) sentence contexts with the first noun that came to mind. Sentences had been truncated at the determiner in order to parallel their presentation in the ERP experiment. In order to obtain the cloze probability of the Italian reading of the homograph, we considered the expected word to be the translation of the Italian version of the homograph. For example, for the homograph *"estate"* (= "property" in English vs. "summer" in Italian"), the word we considered the expected target word for our cloze probability calculations was "summer". Cloze probability was operationally defined as the percentage of times an English word was provided, and the target nouns were coded as high or low cloze probability, as determined by a median split (*Median* = 25%; *Mean* = 35 %). Cloze probability did not differ significantly across contexts ($p > 0.1$).

Each participant was presented with 180 experimental sentences (30 in each condition). Each sentence context (E, I or C) occurred once as a correct trial and once as a violation. Similarly, each target word (H or C) appeared once as a correct trial and once as a violation in each language. Despite these unavoidable repetitions (given our strict selection criteria for target nouns), the repeated words were embedded in different contexts, and the design was meticulously controlled. Trials were distributed across four blocks such that there was never a repetition of a context or of a target word within any single block. Nouns belonging to the same homograph-cognate triplet never occurred within less than 5 intervening trials.

Each participant also saw 124 filler sentences, which were part of the larger English study (testing relative clause word-orders) and which will be reported in forthcoming papers. Out of the total of 304 pseudorandomized stimuli (180 experimental and 124 fillers) per participant, 122 sentences (~ 40%) were acceptable in English (grammatically and semantically), while 182 sentences (~ 60%) were expected to receive a rating of 3 or lower on a five-point rating scale.

2.4. Procedure

Although the session was conducted in English (with English instructions), the experimental procedure was identical to the Italian session (see *General Methods*, p. 56-57).

2.5. EEG recording and analysis

For details on EEG recording, see *General Methods* (p. 57). Trials containing artifacts due to blinks, eye-movements and excessive muscle activity were rejected prior to averaging, using a moving-window standard deviation of 30 microvolts. Two participants (one Control and one Attriter) were excluded from ERP analyses due to excessive artifacts. On average, participants contributed 28 artifact-free trials out of 30 per condition, with no differences across conditions (ps > 0.1).

ERPs were time-locked to the onset of the target noun with a baseline correction of -200 to 0 ms. Three time-windows of interest were examined: (1) 350-500 ms (N400); (2) 650-850 ms (P600); (3) 850-1000 ms (late P600). Repeated-measures ANOVAs were performed separately for 3 midline electrodes (Fz, Cz, Pz) and 12 lateral electrodes (F3/4, C3/4, P3/4 and F7/8, T3/4, T5/6). Global ANOVAs for the midline sites included within-subject factors *Context* (English, Italian, Cognate) and *Target* (Homograph, Cognate). Follow-up ANOVAs for midline sites either included factor *Correctness* (Correct, Violation) or *Condition* (e.g., EH vs. CC), depending on the comparison of interest. Lateral ANOVAs additionally included factors *Hemisphere* (left, right) and *Laterality* (lateral, medial). As a default, reported analyses are restricted to the midline only, except in cases where the lateral ANOVAs revealed additional effects. Where appropriate, Greenhouse-Geisser correction was applied to analyses with more than two levels (e.g., *Ant-Post*). In these cases, the corrected p values but original degrees of freedom are reported.

Given the complexity of the design (i.e., the possibility of 5-way interactions with within-subject factors alone), we have opted to first present our ERP results *within* each participant group separately, in order for readers to better understand the ERP patterns for each group of participants before having to interpret complex interactions. In a subsequent step, in order to determine whether the observed ERP patterns differed *between* groups, we conducted global repeated-measures ANOVAs with *Group* as a between-subjects factor (in *Section 1.3.4*). This "bottom-up" way of presenting the data aims to improve the readability of our paper.

3. RESULTS

3.1. Acceptability judgments and reaction times

Acceptability ratings (on a scale from 1-5) and reaction times for each sentence condition are presented in **Table 3**. Repeated-measures ANOVAs were performed with within-subjects factors *Context* (E, I, C) and *Target* (H, C) and between-subjects factor *Group* (Controls, Italian-English Attriters, English-Italian learners).

Table 3 about here

In the ANOVA for acceptability ratings, a significant *Context* x *Target* x *Group* interaction was found ($F(4,140) = 4.58$, $p < 0.01$), and follow-up analyses indicated that the groups differed on EH ($F(2,72) = 10.03$, $p < 0.01$) and CC ($F(2,72) = 6.15$, $p < 0.05$) conditions only (i.e., on their ratings of correct sentences). Bonferroni-corrected post-hoc comparisons between groups revealed that Attriters gave significantly lower ratings to both types of correct sentences, compared to English-Controls and English-Italian learners (all *p*s < 0.05), while the two groups of native-English speakers did not differ significantly on their EH nor CC ratings (*p*s > 0.1). As expected, violation conditions were rated significantly lower than correct sentences ($F(1,70) = 1006.18$, $p < 0.001$). With all three English violation conditions (IC, EC, CH) collapsed into one, the groups did not differ significantly in their rating (*M* <u>Controls</u>: 2.01; *M* <u>Attriters</u>: 2.14; *M* <u>Italian learners</u>: 2.11; $F(2,72) = 0.540$, $p > 0.1$).

In order to address questions of homograph-interference and cognate facilitation in relation to the previous literature, we assessed whether there were group differences on *differential* ratings between *IH – EH* (where we expected a smaller negative difference for bilinguals, reflecting interference of the correct Italian homograph reading) and *EH – CC* (where we expected a larger negative difference for the bilingual groups, reflecting cognate facilitation). The difference in acceptability ratings between *IH – EH* was indeed significantly different between groups ($F(2,72) = 9.223$, $p < 0.0001$), where rating differences were significantly smaller in Attriters (*M* diff = -1.60) compared to Italian learners (*M* diff = -2.11; $p = 0.04$) as well as to Controls (*M* diff = -2.39; $p < 0.001$). Group differences were also significant for *EH –*

CC ratings ($F(2,72) = 5.14$, $p < 0.01$), reflecting the larger difference in ratings for Attriters (*M* diff = - 0.19) compared to Italian learners (*M* diff = - 0.06) and Controls (*M* diff = + 0.57). These results seem partly in line with our hypotheses, despite the finding outlined in the paragraph above that (1) the groups did not differ on their IH ratings (when differential ratings were not considered), and (2) Attriters rated the CC condition *lower* than the other two groups did (which, in and of itself, seems to counter the notion of facilitation).

Much of the behavioral research to date has highlighted differences in bilinguals' reaction times in processing interlingual cognates and homographs. Our 3x2 (*Context* x *Target*) ANOVA for reaction times revealed a main effect of *Group* ($F(1,70) = 4.90$, $p < 0.05$) reflecting the slower overall reaction times for both bilingual groups relative to English monolingual Controls ($ps < 0.05$). However, the significant *Context* x *Target* effect ($F(2,140) = 21.5$, $p < 0.01$) did not interact with factor *Group* ($p > 0.1$). Violation conditions were responded to more slowly than correct conditions ($F(1,70) = 20.6$, $p < 0.01$).

Despite these non-significant group differences in the global ANOVA, given our strong a-priori hypotheses and the wealth of literature suggesting that bilinguals would show cognate-facilitation (i.e., faster RTs) and homograph-interference (i.e., slower RTs) in their behavioral responses, we checked for group differences on differential RTs between *IH – EH* and *EH – CC*. Group differences for *IH – EH* reached significance ($F(2,72) = 3.925$, $p < 0.05$). Follow-up group comparisons revealed that RT differences were largest in the Italian-learners' group (*M* diff = 945.43 ms) who differed significantly from Attriters (*M* diff = 62.43 ms; $p < 0.05$) but not from Controls (*M* diff = 240.55 ms; $p > 0.1$), potentially indicating an interference effect. For the *EH – CC* comparison, contrary to our hypothesis, *Group* interactions were not significant ($p > 0.1$), although reaction times in Attriters and English-Italian learners were numerically faster for the CC condition than the correct EH condition.

Taken together, our behavioral data did not show robust effects of facilitation or interference in the bilingual groups, in contrast to much of the previous literature. However, given the "offline" nature of the acceptability judgment task – performed at the end of a given sentence – we expected ERPs to be more sensitive to effects of co-activation during real-time language comprehension.

3.2. ERP responses

3.2.1. Native-English Controls

Grand-average ERP waveforms time-locked to the onset of the target noun are presented for all six conditions in **Fig. 1a**. Controls elicited a large centro-parietal N400 effect (350-500 ms) in response to the four violation conditions (EC, IC, CH and IH), relative to the two correct control conditions (EH and CC). The N400 was followed by a posterior P600 effect (maximal at Pz), between 650-800 ms. As expected for English Controls, neither the two correct conditions nor any of the four violation conditions (**Fig. 1b**) appeared to differ from one another in either of the time-windows.

Figure 1a and 1b about here

Fig. 2 illustrates the overall pattern of ERP responses elicited in native-Controls for violation sentences relative to correct sentences, by collapsing across the four violation conditions as well as the two correct conditions. The voltage maps confirm the centro-parietal scalp distribution of the N400 (which appears slightly right-lateralized) and the posterior prominence of the P600.

Figure 2 about here

A global ANOVA on midline electrodes was first conducted with factors *Context* (E, I, C), *Target* (H, C) and *Ant-Post* (Fz, Cz, Pz) in the 350-500 ms time-window. A significant *Context* x *Target* interaction ($F(2,56) = 22.89$, $p < 0.0001$) as well as a significant *Context* x *Target* x *Ant-Post* interaction ($F(4,112) = 3.94$, $p < 0.05$) was found. A follow-up ANOVA comparing the four violation conditions (EC, IC, CH, IH) indicated that the conditions did not differ in the N400 effect they elicited ($p > 0.5$). Multiple pair-wise comparisons revealed that each violation condition elicited a significant N400 effect relative to each correct condition (ps < 0.01), and that the two correct conditions (EH and CC) did not differ significantly from each other ($p > 0.1$).

The six conditions were therefore collapsed to form two levels of *Correctness* (violation vs. correct), as illustrated in **Fig.2**. A repeated-measures midline ANOVA revealed a significant main effect of *Correctness* ($F(1,28) = 22.07$, $p < 0.0001$) and a significant interaction with *Ant-Post* ($F(2,56) = 13.42$, $p < 0.0005$), reflecting the centro-parietal distribution of the N400 (Fz: $F(1,28) = 4.56$, $p < 0.05$; Cz: $F(1,28) = 10.10$, $p < 0.005$; Pz: $F(1,28)= 13.25$, $p < 0.005$) . To confirm that the N400 effect was primarily prominent at the midline and medial sites, a lateral ANOVA was also performed. The main effect of *Correctness* was significant ($F(1,28) = 16.51$, $p < 0.0005$), as were two-way interactions with *Ant-Post* ($F(2,56) = 10.43$, $p < 0.005$) and with *Laterality* ($F(1,28) = 18.75$, $p < 0.0005$). Follow-up ANOVAs demonstrated that F values were larger at medial sites (*Correctness*: $F(1,28) = 20.37$, $p < 0.0001$; *Correctness* x *Ant-Post*: $F(2,56) = 13.0$, $p < 0.0005$) than at lateral sites (*Correctness*: $F(1,28) = 10.47$, $p < 0.005$; *Correctness* x *Ant-Post*: $F(2,56) = 6.63$, $p < 0.01$). Despite the seemingly right-lateralized distribution on the voltage map, interactions with factor *Hemisphere* did not reach significance (ps > 0.1).

A similar approach was taken to analyze the P600 effect. First, in the early P600 time-window between 650-850 ms, the global ANOVA revealed a significant *Context* x *Target* interaction ($F(2,56) = 3.46$, $p < 0.05$) as well as a *Context* x *Target* x *Ant-Post* interaction ($F(4,112) = 3.04$, $p < 0.05$). Follow-up analyses at each midline electrode confirmed the posterior prominence of the P600. (Fz: $p > 0.1$; Cz: $F(2,56) = 3.36$, $p < 0.05$; Pz: $F(2,56) = 5.98$, $p < 0.01$). The ANOVA comparing the four violation conditions (EC, IC, CH, IH) confirmed that the violations did not differ from one another ($p > 0.5$). The two correct conditions (EH and CC) also did not differ significantly ($p > 0.1$). Subsequently, the six conditions were collapsed to form two levels of *Correctness* (violation vs. correct) and a repeated-measures midline ANOVA revealed a significant main effect of *Correctness* ($F(1,28) = 10.78$, $p < 0.005$).

Next, in the late P600 time-window between 850-1000 ms, the global midline ANOVA with within-subject factors *Context*, *Target* and *Ant-Post* did not reveal any significant effects (*Context* x *Target*: $p = 0.12$; *Context* x *Target* x *Ant-Post*: $p = 0.14$). Collapsing the conditions and performing an ANOVA with *Correctness* as a two-level factor, however, did reveal a significant main effect of *Correctness* ($F(1,28) = 6.88$, $p < 0.05$) as well as a significant interaction with *Ant-Post* ($F(2,56) = 4.11$, $p < 0.05$). Follow-up analyses by electrode confirmed that the positivity was more frontally-distributed, as depicted in the voltage maps in **Fig. 2** (Fz: $F(1,28) = 10.35$, $p < 0.005$; Cz: $F(1,28) = 6.80$, $p < 0.05$) rather than being prominent at Pz ($p >$

0.1). Given this atypical distribution of the positivity in this late time-window and the fact that it did not emerge as statistically significant in the global ANOVA with the six separate conditions, the effect should be interpreted with some caution.

3.2.2. Italian-English Attriters

Grand-average ERP waveforms for Italian-English Attriters (highly-proficient and dominant in their L2-English) are presented for all six conditions in **Fig. 3a**. Similar to English controls, Attriters elicited a large centro-parietal N400 effect in response to those conditions which are outright lexical-semantic violations in English (EC, IC, CH, **Fig. 3b**). However, contrary to native-Controls, the N400 elicited in response to IH violations – which would *not* be semantically anomalous in Italian – appeared reduced relative to the other violation conditions. Moreover, the two correct conditions also appeared to differ, as the CC condition – correct in both languages – showed a smaller N400 effect compared to EH sentences (which are correct in English but would be lexical-semantic violations in Italian). In the P600 window, Attriters elicited a posterior P600 between 650-850 ms in response to English violations, as did native-English controls. However, visual inspection suggested that there was an additional effect of target-type, as the correct "EH" condition (a violation in Italian) also elicited a posterior P600 in the same time-window as the violation conditions. The posterior P600 effects seemed to persist longer than in native-Controls (until 1000 ms).

Figure 3a and 3b about here

The global ANOVA on midline electrodes conducted with factors *Context* (E, I, C), *Target* (H, C) and *Ant-Post* (Fz, Cz, Pz) in the 350-500 ms time-window revealed a significant *Context* x *Target* interaction ($F(2,42) = 17.16$, $p < 0.0001$). The interaction between *Context* x *Target* x *Ant-Post* interaction was marginal ($F(4,84) = 2.31$, $p = 0.06$). We then conducted an ANOVA comparing the four violation conditions (EC, IC, CH, IH). Given our hypothesis that IH violations would stand out from the other three violations for Italian-English Attriters, we expected that, unlike in Controls, the ANOVA should reveal a significant main effect of *Condition*. Our prediction was confirmed ($F(3,63) = 2.96$, $p < 0.05$). We then conducted an

ANOVA comparing only the *English* violation conditions (i.e., without IH) in order to show that the *Condition* main effect no longer reached significance ($p > 0.2$). Multiple pairwise comparisons revealed that each violation condition elicited a significant N400 effect (*p*s < 0.05), except for IH ($p = 0.08$). Based on our hypotheses and upon visual inspection of the ERP waveforms (**Fig. 4**), we also expected Attiters' N400 response to the two correct conditions (EH, CC) to differ significantly. While the N400 difference did not reach significance in the whole 350-500 ms window ($p = 0.1$), our prediction was supported in a narrower N400 time-window between 360-460 ms, although the strength of the effect was not as robust as we had anticipated, and the difference was mainly visible at Cz (*Condition*: $F(1,21) = 3.80$, $p = 0.06$; Cz: $F = 4.87$, $p < 0.05$; Pz: $F(1,21) = 2.67$, $p = 0.1$). Still, comparing EH vs. CC conditions in native-English Controls in this same narrow time-window between 360-460 ms confirmed that the pattern seen in Attriters did not even approach significance in Controls ($F(1,28) = 1.55$; $p > 0.2$).

Figure 4 about here

The global ANOVA conducted on the midline in the early P600 time-window between 650-850 ms revealed interesting results in Italian-English Attriters that were not observed within the native-English Controls. First, Attriters showed a significant main effect of *Target* ($F(1,21) = 12.32$, $p < 0.005$) as well as a *Target* x *Ant-Post* interaction ($F(2,42) = 8.19$, $p < 0.005$), confirming the pattern observed in the ERP waveforms (**Fig. 3b**) that homographs elicited a larger posterior P600 than cognates (Fz: $F(1,21) = 4.46$, $p < 0.05$; Cz: $F(1,21) = 11.65$, $p < 0.05$; Pz: $F(1,21) = 22.25$, $p < 0.0001$). To further corroborate these patterns, pairwise comparisons were performed within each level of *Target*. As expected, no significant differences in P600 amplitudes were found within *Homographs* (EH vs. CH vs. IH, all *p*s > 0.1). Within *Cognates* (EC vs. CC vs. IC), as visible on the ERP plots, the CC condition elicited a posterior P600 amplitude of a significantly smaller amplitude compared to EC ($F(2,42) = 5.93$, $p < 0.05$) and IC ($F(2,42) = 10.82$, $p < 0.005$), which in turn did not differ from each other (*p*s > 0.1). A significant *Context* x *Ant-Post* interaction was also found in Attriters ($F(4,84) = 4.01$, $p < 0.05$), follow-ups of which indicated that *Cognate* contexts (CC + CH) differed significantly from English contexts (EC + EH; ($F(2,42) = 3.17$, $p < 0.05$) and Italian contexts (IC + IH; $F(2,42) =$

9.30, $p < 0.005$). However, as can be seen on the plots in **Fig. 3**, this difference is primarily driven by the least-positive amplitude of the CC condition. These significant differences between the CC condition and other cognate conditions reflects the least positive-going amplitude for CC sentences which are semantically-acceptable in both Italian and English. Although correct in English, the EH condition elicited a significantly larger posterior positivity than the CC condition ($F(2,42) = 10.23$, $p < 0.001$).

Finally, in the late P600 window between 850-1000 ms, only the main effect of *Target* emerged as significant in the global ANOVA ($F(1,21) = 9.64$, $p < 0.01$), supporting the visible pattern in the ERP waveforms that homographs elicited a larger P600 than cognates also in this later stage of the P600.

3.2.3. English-Italian Learners

Grand-average ERP waveforms for English-Italian L2 learners are presented for all six conditions in **Fig. 5a**. Similar to native-Controls, English-Italian L2 learners elicited a large centro-parietal N400 effect (350-500 ms) in response to the four violation conditions (EC, IC, CH and IH), relative to the two correct control conditions (EH and CC), as shown in **Fig. 5b**. Unlike Italian-English Attriters, the N400 elicited in response to IH sentences did not appear to be reduced compared to the other English violations. Also in contrast to Attriters, the CC condition – which is correct in both languages – did not appear to show a reduced N400 effect relative to the correct EH condition. In fact, if anything, the pattern appeared to be the opposite (CC > EH).

Figure 5a and 5b about here

In short, in the N400 time-window, the English-Italian bilinguals appeared to be more similar to the native-Controls than to the Italian-English bilinguals (i.e., Attriters). In the P600 window, however, English-Italian L2 learners elicited a posterior P600 effect in response to all violations as well as to correct EH sentences – a pattern reminiscent of what was observed in the Attriters. Moreover, the P600 at posterior electrodes appeared to be modulated by target-type, with EH, IH and CH eliciting larger P600s than their cognate counterparts. Similar to Attriters, the P600 in English-Italian learners lasted beyond 850 ms.

In the N400 time-window between 350-500 ms, the global midline ANOVA conducted with factors *Context* (E, I, C), *Target* (H, C) and *Ant-Post* (Fz, Cz, Pz) revealed a significant *Context* x *Target* interaction ($F(2,38) = 7.19$, $p < 0.005$). The interaction between *Context* x *Target* x *Ant-Post* interaction did not reach significance ($p > 0.1$). We then conducted an ANOVA comparing the four violation conditions (EC, IC, CH, IH). As in native-English Controls (and as predicted based on visual inspection of the ERP waveforms), the violation conditions did not differ in the N400 effect they elicited ($p > 0.5$). Multiple pairwise comparisons revealed that each violation condition elicited a significant N400 effect (*p*s < 0.01). The two correct conditions (EH and CC) differed marginally ($F(1,19) = 3.89$, $p = 0.06$), although in the opposite pattern (CC > EH) from what we had expected based on the difference observed in Attriters. We used the narrow N400 time-window between 360-460 ms to ascertain whether English-Italian bilinguals showed a difference between the two correct conditions in the same time-interval where the Attriters had. The difference between the two correct conditions was marginal ($F(1,19) = 3.25$, $p = 0.08$), although recall that this marginal difference reflected the N400 was larger (rather than reduced) for CC than for EH.

For the early P600 between 650-850 ms, the global midline ANOVA confirmed a similar pattern of a larger P600 in response to homograph targets as in Italian-English Attriters, by demonstrating a significant main effect of *Target* ($F(1,19) = 6.07$, $p < 0.05$) as well as a *Target* x *Ant-Post* interaction ($F(2,38) = 4.28$, $p < 0.05$). The *Context* x *Target* x *Ant-Post* interaction also reached significance in the English-Italian learners ($F(4,76) = 5.05$, $p < 0.005$). Follow-up analyses at each electrode indicated that the P600 was prominent at centro-parietal rather than frontal sites (<u>Fz</u>: $p > 0.1$; <u>Cz</u>: $F(1,19) = 5.95$, $p < 0.05$; <u>Pz</u>: $F(1,19) = 8.96$, $p < 0.01$).

Pairwise comparisons were performed within each level of *Target* and, as was the case for Attriters, no significant differences in P600 amplitudes were found within *Homographs* (EH vs. CH vs. IH, all *p*s > 0.1 except for EH vs. CH where $p = 0.09$). Within *Cognates* (EC vs. CC vs. IC), also following the same pattern as in Italian-English Attriters, the CC condition differed significantly in P600 amplitude from EC ($F(2,38) = 12.17$, $p < 0.005$) and IC ($F(2,38) = 5.78$, $p < 0.05$), while EC and IC did not differ from each other (*p*s > 0.1). As was the case in Attriters, the CC condition was also the least positive condition in the early P600 window for English-

Italian learners, and it differed significantly from the correct EH condition ($F(2,38) = 6.03$, $p < 0.01$).

Lastly, in the late P600 time-window (850-1000 ms), only the main effect of *Target* was significant in the global ANOVA ($F(1,19) = 7.32$, $p < 0.05$), as was the case in Italian-English Attriters, reflecting larger P600 amplitudes for homograph target words than for cognates. Within *Cognates*, we confirmed that, as was the case for the Attriters, the CC condition elicited the least positive-going waveform, as it differed significantly from the EC condition ($F(2,38) = 4.74$, $p < 0.05$), and marginally from IC ($F(2,38) = 3.81$, $p = 0.06$), whereas the two other *Cognate* conditions (EC and IC) were not statistically different from one another (ps > 0.1).

To summarize all individual group patterns, English-Italian learners were similar to the native-English Controls in the N400 window and elicited a centro-parietal N400 in response to all violation conditions, including IH. Attriters, on the other hand, showed reduced N400 amplitudes when sentences were either acceptable in Italian (IH) or in both languages (CC). The CC condition did not have a facilitatory effect on English-Italian learners relative to correct EH sentences in the N400 time-window. In the P600 window, English-Italian bilinguals showed processing patterns that resembled those observed in Attriters, namely a larger P600 effect for homograph target-words than for cognates, persisting beyond 850 ms, including a P600 in response to EH sentences, although they were semantically correct in English. For both bilingual groups, the CC condition had the least positive P600 amplitude. As expected in English-Controls, neither the violation conditions nor the different correct conditions differed from one another, and the P600 effect was significant only in the early time-window (650-850 ms).

3.2.4. Group comparisons

Despite the striking patterns that emerged from our within-group analyses, particularly for the two critical IH and CC conditions where we expected cross-linguistic effects for the bilingual groups, we performed global ANOVAs including *Group* as a between-subjects in order to draw any conclusions as to whether ERP response patterns truly differed across groups.

First, with respect to the N400, we wished to confirm that (1) the groups did not differ in the amplitude and distribution of the N400 elicited in response to English lexical-semantic violations (EC, IC, CH), relative to the correct conditions (EH, CC); (2) *Group* was a significant factor in modulating the N400 in response to IH violations relate to EH; and (3) *Group* was also a significant factor in modulating N400 differences between the two correct conditions (CC vs. EH).

Relative to the P600, we wished to ascertain whether (1) *Group* was a significant factor both early and late P600 time-windows; (2) the two bilingual groups did not differ from each other in their P600 response patterns; but that (3) Controls differed significantly from Bilinguals in both P600 time-windows.

3.2.4.1. N400 group comparisons

Fig. 6 illustrates the N400 effect elicited in response to all English violations (EH, IC, CH) relative to correct English sentences (EH, CC) for all three groups. The global ANOVA on the midline confirmed that Italian-English Attriters' showed an N400 response in their L2-English that was indistinguishable in amplitude and scalp distribution from the N400 response of native-English speakers (i.e., the monolingual Controls and English-Italian bilinguals). The groups shared a main effect of *Correctness* ($F(1,67) = 63.80$, $p < 0.0001$) as well as a significant interaction between *Correctness* x *Ant-Post* ($F(2,124) = 13.87$, $p < 0.0001$), reflecting the centro-parietal distribution of the N400 (Fz: $F(1,67) = 28.36$, $p < 0.0001$; Cz: $F(1,67) = 71.57$, $p < 0.0001$; Pz: $F(1,67) = 82.38$, $p < 0.0001$). Crucially, however, no interactions with factor *Group* were close to statistical significance (*Correctness* x *Group*: $p = 0.81$; *Correctness* x *Ant-Post* x *Group*: $p = 0.61$).

Figure 6 about here

Fig. 7 compares the groups on their N400 response to IH violations relative to correct EH sentences. Comparing conditions IH vs. EH across the three groups, the midline ANOVA revealed a main effect of *Condition* ($F(1,68) = 36.70$, $p < 0.0001$) as well as a marginal *Condition* x *Group* interaction ($F(2,68) = 2.29$, $p = 0.08$), which was in the direction of our hypothesis, namely that Italian-English Attriters showed a reduced N400 effect relative to the other violation conditions and so the IH condition was only marginally different from the correct control condition.

Figure 7 about here

Fig. 8 illustrates the comparison between the two correct conditions EH vs. CC, where Italian-English Attriters are the only group to show a reduced N400 effect for CC sentences (i.e., EH sentences, though correct in English, elicited a larger N400 than CC sentences in Attriters). The global ANOVA confirmed this pattern and revealed a significant *Condition* x *Group* interaction ($F(2,68) = 3.80$, $p < 0.05$) in the absence of a significant main effect of *Condition* ($F(1,68) = 0.42$, $p > 0.5$).

Figure 8 about here

Given that the EH condition is semantically anomalous in Italian (as the English context primes the English meaning of the interlingual homograph and would therefore not be consistent with the Italian meaning of the same word), while the CC condition is semantically plausible in both languages, it is worth exploring whether any potential EH vs. CC differences in the bilingual groups of speakers are driven by differences in the EH condition (suggestive of "interference" from Italian), and/or rather by differences in the CC condition (suggesting that the dual-acceptability of these sentences may facilitate their processing if the N400 for CC is reduced as in the case of Attriters). Interestingly, a comparison of the groups on each of these conditions seemed to suggest that group differences in our EH vs. CC analyses were driven by differences on the CC condition ($F(2,68) = 5.27$, $p < 0.01$) with Attriters differing from the two other groups, whereas the groups were indistinguishable in their N400 responses to EH sentences ($F(2,68) = 0.50$, $p > 0.1$).

3.2.4.2. P600 group comparisons

As predicted, the global ANOVA comparing the three groups on the P600 effect in response to English violations (EC, IC, CH) elicited between 650-850 ms revealed a significant *Correctness* x *Group* interaction ($F(2,67) = 4.64$, $p < 0.05$), in addition to a marginal main effect of *Correctness* and a significant interaction between *Correctness* x *Ant-Post* ($F(2,134) = 12.80$, $p < 0.0001$). Group was also a significant factor in the late P600 time-window (*Correctness* x *Group*: $F(2,67) = 4.30$, $p < 0.05$; *Correctness* x *Ant-Post* x *Group*: $F(4,134) = 5.94$, $p < 0.005$). However, as we know from the P600 patterns already discussed within each group, the more appropriate and meaningful analysis for the P600 effect is not by "*Correctness*" but rather by *Context* and *Target*, given that the bilingual groups also showed a P600 in response to *correct* EH sentences (and larger P600s to homograph conditions in general). **Fig. 9** illustrates the *Target* main effect (homographs > cognates) that we expected to be modulated by *Group*.

Figure 9 about here

First, we compared Italian-English Attriters and English-Italian learners in order to confirm that the groups did not differ in either P600 time-window. The global ANOVA found a significant main effect of *Target* ($F(1,40) = 17.86$, $p < 0.0001$) and significant interactions between *Target* x *Ant-Post* ($F(2,80) = 11.76$, $p < 0.0001$) and *Context* x *Ant-Post* ($F(4,160) = 5.23$, $p < 0.001$), as well as a three-way interaction between *Context* x *Target* x *Ant-Post* ($F(4,160) = 6.85$, $p < 0.001$). As predicted, however, none of the interactions with factor *Group* approached significance (ps > 0.1). In the late P600 window, the main effect of *Target* persisted ($F(1,40) = 16.57$, $p < 0.0005$), as well as the interactions between *Target* x *Ant-Post* ($F(2,80) = 3.92$, $p < 0.05$) and *Context* x *Ant-Post* ($F(4,160) = 3.36$, $p < 0.05$). Once again, none of the interactions with *Group* approached significance (ps > 0.1), confirming that the two groups of bilinguals (Italian-English Attriters dominant in English vs. English learners of Italian) did not differ in the amplitude, latency nor scalp distribution of the P600 effects in either time-window.

Given that Italian-English Attriters and English-Italian learners were indistinguishable in their P600 responses, we collapsed the two groups into one bilingual group. In an ANOVA comparing the bilinguals to the monolingual Controls, we expected to find significant interactions with *Group*, particularly involving the factor *Target*. Our predictions were confirmed. The midline ANOVA between 650-850 ms revealed a significant main effect of *Target* ($F(1,69) = 9.28$, $p < 0.005$), two-way interactions between *Target* x *Ant-Post* ($F(2,138) = 7.03$, $p < 0.005$) and *Context* x *Ant-Post* ($F(4,276) = 4.06$, $p < 0.005$), as well as a significant three-way interaction between *Context* x *Target* x *Ant-Post* ($F(4,276) = 8.64$, $p < 0.0001$). Most importantly, a number of interactions with *Group* were also significant, namely *Target* x *Group* ($F(1,69) = 3.98$, $p < 0.05$) and *Context* x *Target* x *Group* ($F(2,138) = 3.24$, $p < 0.05$), and *Target* x *Ant-Post* x *Group* (marginal: $F(2,138) = 3.16$, $p = 0.06$), indicating that Controls differed significantly from bilinguals in their P600 responses. This pattern was also true of the later P600 time-window between 850-1000 ms, and *F* values for group interactions were even larger in this later interval (*Target* x *Group*: $F(1,69) = 8.56$, $p < 0.005$; *Context* x *Target* x *Group*: $F(2,138) = 2.99$, $p < 0.05$; *Target* x *Ant-Post* x *Group*: $F(2,138) = 3.54$, $p < 0.05$; *Context* x *Target* x *Ant-Post* x *Group*: $F(4,276) = 3.01$, $p < 0.05$). Recall that, in our bottom-up approach of reporting our Results, we already "followed-up" these interactions in the analyses conducted within each group, and thus we do not repeat the patterns here.

3.2.5. The influence of proficiency level

3.2.5.1. Proficiency measures and behavioral responses

We assessed whether "global" English proficiency (as measured by accuracy on our various proficiency tasks) would significantly correlate with task-specific proficiency (as measured by our acceptability judgment task on experimental sentences). As detailed in Table 4, English proficiency level was found to influence participants' acceptability judgments of English lexical-semantic violations (= EC, IC, CH) relative to correct control sentences (EH, CC), such that participants with higher scores gave lower acceptability ratings to semantically incongruent sentences. The difference between IH and EH acceptability ratings (i.e., the discriminability of IH as a violation) was also negatively correlated with English proficiency and reading measures.

Table 4 about here

For the two bilingual groups, relative proficiency level (Italian – English) was significantly correlated with differential ratings for *IH – EH* conditions as well as *EH – CC* conditions, indicating that, when relative proficiency was increasingly asymmetrical in favor of *English* (i.e., more *negative* relative proficiency scores), ratings for IH and CC sentences were less favorable (i.e., more *negative* IH – EH differential ratings and more *positive* EH – CC differential ratings). Note that this pattern was also true of Attriters alone, both for *IH – EH* (r = 0.723, $p < 0.0001$) and *EH – CC* (r = -0.692, $p < 0.05$), therefore allowing us to rule out the possibility that the significant results in the direction of our hypotheses were due to group characteristics.

We also examined whether behavioral performance on the false-friends translation (production) task was similar to bilinguals' acceptability ratings of sentences (comprehension). An average accuracy score for critical items (false-friends) was calculated over both sessions to have an overall score reflecting how well these bilinguals could control cross-linguistic interference during the translation task. However, keeping in mind that interference effects could differ in strength depending on the direction of the translation, we also considered each session (i.e., translation direction) separately. An interesting finding was that ratings on IH sentences were negatively correlated with accuracy in translating the false-friend items, but *only* on the session where translation was from Italian to English (i.e., task performed in English as in our ERP study (r = -0.320, $p < 0.05$). Ratings on CC sentences were negatively correlated with average accuracy (both sessions) on the translation task (r = -0.308, $p < 0.05$). These results indicated that bilingual participants who were better able to control competition of non-target semantic meanings rated IH and CC sentences less favorably than bilinguals who produced incorrect translations of the false-friend items (thus less inhibition and less facilitation on acceptability ratings, respectively).

3.2.5.2. Proficiency measures and ERPs

Table 4 also reports the significant correlations between proficiency measures and ERP amplitudes. Only the N400 and P600 modulations due to co-activation effects (rather than English violation – correct ERP effects) were significantly correlated with proficiency measures. In the case of *IH – EH*, negative correlations indicate that individuals with higher English-proficiency scores elicited larger negative amplitudes in the N400 time-window than individuals with lower-proficiency scores. The opposite was true for *EH – CC*, namely the higher the English proficiency level, the smaller the N400 difference (and thus, the CC facilitation effect).

We expected N400 differences to be smaller for *IH – EH* sentences but larger for *EH – CC* sentences in individuals with a larger asymmetry in favor of Italian over English proficiency (i.e., with more positive relative proficiency scores). These predictions were largely confirmed, although the correlation for the IH contrast was only marginally-significant.

Finally, bilinguals with a higher accuracy in translating false-friend items elicited a larger P600 response for *Homograph* vs. *Cognate* target-words, and – as was the case for acceptability ratings – only when translation involved the direction from Italian to English, suggesting perhaps that bilinguals who are better at controlling interference from Italian-to-English engage in an additional conflict-monitoring process (eliciting a P600 effect for homograph target-words, including correct EH conditions).

3.2.6. The influence of background factors on ERP responses

Attriters' ERP response patterns were not significantly influenced by their age of immigration / AoA of English ($p > 0.1$). The N400 for IH vs. EH sentences was negatively correlated with length of residence (LoR), such that the longer Attriters' LoR, the larger the negative amplitude of the N400 in response to IH violations ($r = - 0.421$, $p < 0.05$). Thus, the longer the LoR and, presumably, the more time they had to become immersed in the L2 environment and disconnected from their L1 environment), the less Attriters showed co-activation of the Italian reading of the homograph (i.e., they were more English native-like). Conversely, for EH vs. CC correct sentences, the correlation was positive ($r = 0.391$, $p < 0.05$), such that the shorter Attriters' LoR, the larger the N400 effect for EH sentences which, though correct in English, constitutes a violation in Italian. Finally, amount of L1-Italian exposure was positively correlated with the N400 effect elicited by English violations relative ($r = 0.414$, $p < 0.05$), indicating that the more L1-Italian exposure Attriters had, the less negative their N400 amplitudes. No correlations were found between ERP patterns and age at testing or number of years of education.

3.2.7. Effects of sentence constraint

Given our panoply of research questions, exploring effects of sentence constraint (i.e., cloze-probability) was not one of the main goals of our study. However, we felt it necessary to check whether our study was in line with previous work on homographs and cognates in sentence contexts. Based on the literature, our hypotheses were threefold:

(1) Cloze-probability would affect correct sentences, such that high-cloze (i.e., highly constraining) correct sentences would elicit a reduced N400 effect compared to low-cloze correct sentences. We expected this pattern to hold for all groups, given that Italian-English Attriters are highly-proficient in English and presumably able to make use of contextual cues to a similar degree as native-speakers of English.

In line with this prediction, a repeated-measures ANOVA with 2 levels of *Correctness* (Correct, Violation) and 2 levels of *Cloze* (High, Low) was conducted on midline electrodes for the same N400 time-window as in the main analysis (350-500 ms). Both the main effect of *Cloze* ($F(1,68) = 16.35$, $p < 0.0001$) and the interaction of *Cloze* x *Ant-Post* ($F(2,136) = 11.60$, $p < 0.0005$) reached significance, and were qualified by a significant interaction between *Corr* x *Cloze* x *Ant-Post* ($F(2,136) = 4.22$, $p < 0.05$), in the absence of any significant interactions with *Group* (*p*s > 0.1). Follow-up comparisons between each level of *Correctness* indicated a *Cloze* effect in Correct sentences, with Low-Cloze correct sentences eliciting a larger N400 effect than High-Cloze correct sentences (<u>*Cloze*</u>: $F(1,68) = 10.22$, $p < 0.005$; <u>*Cloze* x *Ant-Post*</u>: $F(2,136) = 12.09$, $p < 0.0001$).

(2) The cognate facilitation effect (N400 difference for EH – CC) would be stronger in Low-Cloze sentence contexts than High-Cloze sentence contexts (as found in Duyck et al., 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Hell & De Groot, 2008), at least for Italian-English Attriters who showed more robust facilitation effects in the N400 time-window. When comparing difference waves (EH – CC) for High vs. Low Cloze sentences, the interaction between *Cloze* x *Ant-Post* x *Group* was close to significance ($F(4,136) = 2.29$, $p = 0.06$), and Attriters were the only group to show a main effect of *Cloze* ($F(1,22) = 9.82$, $p < 0.05$). Though marginally-significant in the global ANOVA, the trend followed the direction of our hypothesis, namely that the N400 effect was larger for *Low-Cloze* EH – CC sentences. Note that there was no cloze-probability effect on cognate facilitation in the late P600 in Italian-English Attriters and English-Italian learners (*p*s > 0.1). In other words, regardless of sentence constraint, cognates elicited less positive-going waveforms in the P600 window than homographs.

(3) The homograph interference effect (N400 difference between English violations – IH) should be stronger in High-Cloze sentence contexts than Low-Cloze contexts, as highly-constraining IH sentences are more consistent with the Italian meaning of the homograph (thus eliciting a smaller N400 than other violation conditions if the Italian meaning is co-activated).

In line with this hypothesis, when comparing difference waves (English violations – IH) for High vs. Low Cloze sentences, the interaction between *Cloze* x *Group* approached significance ($F(2,68) = 2.67$, $p = 0.07$). The pattern reflected the absence of *Cloze* effects in Controls, for whom English violations were also not different than IH violations overall. In Attriters, High-Cloze English violations elicited a significantly larger N400 than High-Cloze IH violations ($F(1,22) = 4.56$, $p < 0.05$), whereas there were no significant differences in the N400 elicited for Low-Cloze English violations and Low-Cloze IH violations (ps > 0.1). English-Italian learners showed a similar trend, where the N400 was numerically larger in High-Cloze English violations than High-Cloze IH sentences ($F(1,19) = 2.06$, $p = 0.1$), whereas Low-Cloze violations overlapped with one another. Thus, it appears that sentence-constraint for homograph-interference (when the context favors the non-target-language reading of the homograph) works in the opposite direction than for cognate facilitation, namely that highly-constraining sentences that "prime" the non-target meaning of the homograph reduce the English violation effect, at least for bilinguals for whom the non-target language is the L1.

4. DISCUSSION

The present study tested English lexical-semantic processing in Italian-English Attriters (highly-proficient in and predominantly exposed to English as a late-acquired L2), compared to English-Italian bilinguals and English native-monolinguals. Our main goal was to determine whether L1-to-L2 lexical co-activation implicitly occurs in L1-Attriters while reading English sentences during an English task performed in their L2 environment and, if so, whether the degree of co-activation was modulated by factors such as proficiency (in L2 but also in L2 relative to L1), age-of-acquisition of English (AoA) and/or other factors such as length of residence (LoR) and amount of L1/L2 exposure. We compared Attriters to L1-English late learners of Italian (thus, bilingual speakers of the same language pair but opposite AoA profiles) to elucidate whether Italian co-activation had a similar effect on the *L1* processing of the Italian learners, despite the difference in direction of cross-linguistic influence. To answer these questions, we explored the processing of English sentences with interlingual homographs (+ form, - meaning) and cognates (+ form, + meaning) embedded in sentence contexts that were either semantically-congruent in English only (EH), Italian only (IH), both languages (CC) or neither (EC, IH, CH). Our ERP findings will be summarized and interpreted separately for each of the ERP components of interest (for a discussion of behavioral results, see *Section 4.3*).

4.1. Main findings on the N400 (350-500 ms)

In accordance with the vast ERP literature on lexical-semantic processing in sentence contexts, native-English Controls elicited a large, centro-parietal N400 effect between 350-500 ms in response to each of the four violation conditions (EC, IC, CH and IH) relative to the two correct conditions (EH and CC) and, as predicted, neither the violation conditions nor the correct conditions differed from one another, given that these sentence conditions are equally implausible for native-English speakers who have no knowledge of Italian.

Italian-English Attriters also showed a large, centro-parietal N400 effect in response to sentences that were lexical-semantic violations in English (EC, IC, CH). Moreover, this response was indistinguishable in amplitude and scalp distribution from both native-English groups, suggesting that L1-Italian Attriters who were immersed and highly-proficient in an L2 they used predominantly on a daily basis were native-like in their real-time processing of lexical-semantic anomalies. In response to IH sentences – where the sentence context primes the Italian meaning of the homograph and is thus incongruent with the English meaning – the N400 effect was reduced in Attriters, compared to other violation conditions. This reduction in the N400 can be construed as evidence for co-activation of the L1-Italian meaning of the homograph while reading English sentences – a finding that is in line with models of non-selective lexical access in bilinguals (BIA+, Dijkstra & Van Heuven, 2002; RHM, Kroll & Stewart, 1994). In addition, examining effects of sentence-context revealed that the IH condition differed from other violation conditions in its N400 amplitude when the Italian homograph meaning was more strongly predictable from the sentence context (i.e., high cloze-probability for the Italian reading), but not when the sentence context was more neutral.

Lexical co-activation from L1-to-L2 was also found to occur for correct cognate sentences (CC) relative to correct homograph sentences (EH), particularly in a narrower N400 time-window between 360-460 ms. Although both sentences were correct and were indistinguishable in native-English monolinguals, Italian-Attriters showed a reduced N400 effect in the CC condition – a pattern that replicates the widely described *cognate facilitation* effect (e.g., Dijkstra, Van Jaarsveld & Ten Brinke, 1998; Midgley, Holcomb & Grainger, 2011; Van Hell & Dijkstra, 2002). A closer look at the EH – CC comparison revealed that group differences on this effect were not due to N400 differences on the EH condition, but rather due to N400 differences on the CC condition. Consistent with findings in previous studies conducted with sentences (Duyck et al., 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Hell & De Groot, 2008), this facilitation effect was also (marginally) modulated by sentence constraint – cognates seemed to more readily facilitate integration of a target-word when it was less predictable from the preceding sentence context, whereas highly-predictable target-words had little to gain from sharing their semantic meaning with an Italian noun. Modulations of the N400 effect in Italian-English Attriters were consistent with our hypotheses and indicated that L1-meanings are activated in parallel during L2-reading in those instances where experimental conditions foster L1 co-activation (e.g., sentences that are plausible if translated into the L1, or sentences where L1 and L2 semantic meanings converge). These results extend prior findings to a unique subgroup of bilinguals (i.e., Attriters) who are immersed in the L2 and experiencing changes in their L1.

In contrast, in English-Italian learners, the N400 elicited in response to IH sentences was not significantly reduced compared to the other English violation conditions, nor was the N400 for the CC condition. In fact, albeit marginally significant, the pattern appeared to be in the opposite direction (CC > EH), as though cognate words were recognized as such but caused interference rather than facilitation. With respect to effects of sentence constraint, Italian learners showed a similar trend as Attriters where the N400 difference between English violations and IH violations was qualitatively larger in conditions where the sentence context was highly-constraining (i.e., high cloze-probability). Overall, however, the English-Italian bilinguals were more similar in their ERP response patterns to the native-Controls than to the Italian-English bilinguals, at least in the N400 time-window.

The lack of a robust co-activation effect in the N400 time-window for English-Italian learners, compared to Italian-English Attriters, may be due to a number of possible factors. Previous studies have reported stronger co-activation effects in tasks where the direction of the transfer is from the stronger to weaker language, which in the majority of cases means from the L1 to the L2 (Costa et al., 2000; Elston-Güttler, Paulmann & Kotz, 2005; Libben & Titone, 2009; Poarch & Van Hell, 2012). This can either be related to constructs of L1-L2 status (i.e., AoA), proficiency or dominance. In the present case, AoA (of Italian) did not affect any of the ERP responses elicited in English-Italian learners. However, additional correlations with *Italian* proficiency level revealed that the N400 effect for IH – EH sentences was larger in amplitude in Italian-learners with higher Italian C-test scores (r = - 0.462, p < 0.05). Similarly, the N400 effect for EH – CC sentences was larger in amplitude for Italian-learners with higher scores on the Italian version of the Error-detection test (r = - 0.387, p < 0.05). Thus, co-activation in terms of both facilitation and interference in late L2 learners of Italian was likely modulated by Italian proficiency level. It has been suggested that a minimum level of L2 proficiency is required for co-activation effects to emerge on the L1 (e.g., Van Hell & Dijkstra, 2002). A closely related point is that our English-Italian learners were dominant in their L1-English and tested in their L1-environment; as a next step, it would be important to test advanced English-Italian late learners in an immersion setting (i.e., the mirror image group of our Attriters) at even higher levels of L2-Italian proficiency, in order to clarify whether the lack of co-activation effects on the N400 in this group were due to insufficient L2 proficiency for L2-to-L1 transfer effects to come into play during real-time comprehension.

Another possible factor, which is also likely to interact with proficiency/dominance, is that the nature of the experimental task and stimuli did not promote lexical co-activation from L2 to L1 as strongly as it may have in other tasks. For example, it has been repeatedly found that lexical co-activation effects occur in tasks where words are presented in isolation, but become attenuated or disappear altogether when presented in sentence contexts (Duyck et al., 2007; Elston-Guttler, 2000; Elston-Guttler, Paulmann & Kotz, 2005; Van Hell & De Groot, 2008). It may be that reading sentences in one's L1 and rating their overall acceptability does not automatically co-activate the Italian-L2, except in cases where the sentence-context directly pushes in favor of L1-meanings (i.e., the IH condition in high cloze-probability conditions, where Italian learners approached the reduced N400 pattern seen in Attriters). It has also been suggested that the intermixing of cognates and homographs within a same experiment may affect patterns of co-activation (as discussed in Van Hell & Tanner, 2012). More specifically, including homographs in the mix of stimuli is predicted by the RHM to enhance the ambiguity in the reliability of form-meaning links (which converge in the case of cognates but diverge for homographs), such that only bilinguals with strong form-meaning mappings will be able to benefit from cognates. Bilinguals with weak form-meaning links, on the other hand, may be slowed down by the enhanced lexical ambiguity created by the occurrence of homographs in the same experimental paradigm, and may instead show competition/interference effects for cognates as well. This possibility seems to fit with the finding of a larger N400 for CC sentences compared to EH sentences in English-Italian learners – a pattern that should be further explored in future studies.

4.2. Main findings on the P600 (650-850 ms and 850-1000 ms)

In addition to the N400, violation conditions also elicited a posterior P600 effect, which has been typically associated with morphosyntactic processing and has been found in response to ungrammatical, ambiguous or complex sentences where re-analysis or revision is required to rescue the interpretation of the sentence (Friederici et al., 2001; Hagoort, Brown & Groothusen, 1993; Kaan & Swaab, 2003; Osterhout & Holcomb, 1992; 1993). However, the P600 has also more generally been attributed to controlled processes such as "*conflict monitoring*" (van de Meerendonk, Kolk et al., 2009), even at the level of processing lexical representations (van de Meerendonk, Kolk et al., 2010; Vissers et al., 2006), and is therefore not restricted to morphosyntactic processing.

Native-English Controls elicited a posterior P600 in response to lexical-semantic violations compared to semantically-correct sentences in the time-window between 650-850 ms. Neither the four violation conditions nor the two correct conditions differed significantly from one another. Controls did not elicit a reliable P600 in later time-window beyond 850 ms, as the positivity was frontal in distribution (maximal at Fz) and only emerged as significant when all violation conditions were collapsed together and compared to correct conditions, but not in the 3 x 2 ANOVA with factors *Context* x *Target*. This positivity with an atypical distribution must therefore be interpreted with caution.

Both bilingual groups, on the other hand, showed interesting and unexpected P600 effects that were not observed in the English monolinguals, and which were therefore indicative of a bilingualism effect. Importantly, Attriters and Italian learners were indistinguishable from each other in the P600 time-window. Both groups showed an effect of target-type on the posterior P600, where homograph-targets elicited larger P600 effects than cognate-targets, not only in the early time-window between 650-850 ms but also in the later time-window between 850-1000 ms. These target-effects were positively correlated with accuracy in translating false-friend items in the production task, but only when the task involved translating items from Italian to English and not in reverse. Higher accuracy on this task is reflective of a stronger ability to control interference and resist the temptation to rely on word-form to arrive at the correct English meaning (e.g., the translation of "*fabbrica*" is not "*fabric*" but "*factory*"). Proponents of non-selective models of lexical access such as the RHM (Kroll & Stewart, 1994) would explain this in terms of stronger vs. weaker form-meaning links in the bilingual lexicon, leading to more reliable activations of correct semantic meanings in cases where there is competition in the word-form. This controlled inhibitory process may be what is reflected by the larger P600 effect for homographs. Unlike cognates, which converge on their semantic meaning, homographs require more "conflict-monitoring" in bilinguals, especially once they recognize a word-form as belonging to both languages.

Consistent with this theory was the finding of a P600 effect for correct EH sentences (which are incongruent with the non-target meaning of the homograph), as well as the finding that CC sentences (semantically-congruent in both languages) elicited the least positive waveform. The P600 effect therefore seems to correlate with the degree of conflict encountered by the reader. However, with our current design, we are not able to determine whether the CC condition showed the least positive waveform in the P600 window due to a cross-linguistic facilitation effect, or simply because it was the only non-homograph correct condition we tested. A follow-up study should include a correct control condition where the target-word is neither a homograph nor a cognate but a matched control-word. Our prediction is that the P600 response would be graded according to the degree of conflict (EH > Control > CC).

It is interesting that the recognition of a word-form as a homograph (i.e., target-effect) did not occur in the N400 time-window but further downstream. A possibility is that the P600 represents the process of selection of the target-language reading of the cross-linguistically ambiguous word. The BIA+ model (Dijkstra & Van Heuven, 2002) posits that "language-membership representations" (i.e., "language nodes" that identify the language that the word belongs to) become activated relatively late, after an initial non-selective / parallel activation of lexical candidates across the two languages. In order to arrive at a correct sentence interpretation, readers must zero-in on the contextually-appropriate interpretation of the target-word (see Bruijn, Dijkstra, Chwilla & Schriefers, 2001; Dijkstra, Timmermans & Schriefers, 2000). It would be of interest to examine whether P600 amplitudes would be larger in language-identification tasks that tap explicitly into this recognition/selection/assignment process. One could also compare P600 effects elicited by interlingual homographs to processing patterns observed for *intralingual* homographs (e.g., *bank* (financial institution vs. shore). If the P600 reflects a general process of "conflict monitoring", then one would expect it to be elicited in instances of intralingual competition as well. Conversely, if the P600 we observed reflects processes of language-identification in order to arrive at a language-selective interpretation of the sentence, this effect should presumably be absent in sentences containing ambiguous words within a language.

A related possibility is that the P600 effect reflects the integration of the target-word into the preceding context – a process in which homographs cause additional conflict or require additional revision because of their ambiguity. In this case, P600 effects might be greater in sentence contexts than in paradigms with isolated words. In contrast to this prediction, however, one ERP study that reported a P600 effect (larger for non-cognates than cognates) was a lexical-decision paradigm with single words (Peeters, Dijkstra & Grainger, 2013).

Although a language-assignment node is part of non-selective models of bilingual lexical access, the temporal dynamics of the parallel word-identification and subsequent language-selection systems are not described in detail. To date, only a few studies have commented on the time-course of bilingual lexical access and semantic integration. For example, the eye-tracking study by Libben and Titone (2009) reported cognate facilitation effects in sentence contexts only during early stages of comprehension. By late stages of comprehension, non-selective lexical access had been resolved and the target representation had been selected. Taken together, such findings highlight the value of studying the time-course of lexical access, not just modulations on early effects such as the N400. Such studies would have interesting implications for models of bilingual lexical access which should attempt to integrate more information about the temporal dynamics of co-activation, selection and semantic integration, and how this time-course may be modulated by methodological factors such as stimuli, task and sentence contexts, or participant-factors such as proficiency and language dominance.

4.3. Proficiency, attrition and lexical-semantic processing

The present study included two groups of bilingual speakers of the same language pair but with different L1s, which allowed us to investigate whether effects of non-selective, parallel activation of crosslinguistic lexical items occurred in both directions (L1-to-L2 and L2-to-L1).

Our inclusion of L1-Attriters also allowed us to investigate how proficiency might affect ERP responses in a study where the L1 was not the language with the most exposure and use, and was subject to negative changes in proficiency compared to other native-speakers.

Our results showed that English-proficiency level predicted participants' acceptability ratings for violation conditions such that ratings were lower in participants with higher proficiency scores. Importantly, in the bilinguals, both IH and CC sentences were rated more favorably (relative to control EH sentences) by individuals with more positive relative-proficiency scores (i.e., larger gap between Italian and English proficiency levels). Bilinguals' accuracy on the false-friends translation task – a production measure that specifically tested vocabulary in a crosslinguistic paradigm that promoted lexical co-activation – was associated with more negative ratings for the IH condition. Thus, the better bilinguals were at inhibiting "false-friend" semantic meanings during the translation task, the more English-like they were in the comprehension task in rejecting IH sentences despite their semantic plausibility in Italian. As briefly discussed above in the context of our N400 and P600 findings, both English proficiency and relative proficiency were found to modulate ERP responses in those conditions where we expected crosslinguistic co-activation.

Furthermore, we found that Italian-English Attriters, though late learners of their L2, elicited N400 responses indistinguishable from both monolingual and bilingual native-English speakers, except in cases where we explicitly pushed for co-activation of L1 semantic meanings to occur in parallel (i.e., in highly-constraining IH sentences and CC sentences). Native-like-ness in English was also correlated with factors such as length of residence and amount of L1-Italian exposure, such that Attriters with a longer LoR and more reduced L1-exposure were more similar to native-English speakers in their processing. These results are in favor of ongoing neuroplasticity for L2-learning in adulthood and show that brain responses elicited in real-time lexical-semantic processing is strongly predicted by proficiency level and can be native-like, even despite a late AoA.

In contrast with the majority of studies conducted on bilingual lexical access, our study used several proficiency tasks, administered both in participants' L1 and L2 *and* in monolinguals. First, the use of several measures is advantageous as it may reveal that some tasks are more difficult for highly-proficient L2 learners who are fully immersed in the L2 environment, and are therefore more sensitive to group differences. For example, not only did our Error-detection task reveal differences between Italian-English attriters and native-speakers that the C-test did not, but it was also more sensitive to individual proficiency differences among English monolinguals (Range: 56 - 96%) than the C-test was (Range: 73 – 100%). Secondly, it has been shown that tasks that are more closely related to the linguistic elements specifically being tested are typically better predictors of ERP responses (see Kasparian, Bourguignon, Drury & Steinhauer, 2010). One example of this in our current study is that the target-word effect on the P600 in the two bilingual groups (where homograph targets > cognates) was not correlated with any proficiency measures *except* participants' accuracy on the false-friends translation task.

In addition to examining effects of English-proficiency level, we also examined relative Italian-English proficiency levels, which taps into the balance between the bilinguals' two languages, given that (1) Italian proficiency is also relevant in a task involving crosslinguistic co-activation, and (2) Attriters may be similar on their English proficiency level but may experience variability in their L1. Indeed, relative-proficiency proved to be a sensitive predictor of the degree of co-activation effects. Given that attrition can be described as a progressive shift from L1 proficiency towards L2 proficiency, it may be of interest for future studies to use such a relative-proficiency measure and to determine cases in which relative-proficiency might be a stronger predictor of behavioral or ERP differences than L2 proficiency alone.

4.4. Context and task effects

The majority of ERP studies conducted on bilingual lexical access so far have involved priming paradigms using single words rather than sentence contexts (but see Elston-Güttler, Gunter & Kotz, 2005; Elston-Güttler, Paulmann & Kotz, 2005). In addition to embedding the homograph and cognate target-words into L2 sentence contexts, we also included a critical condition where the sentence-context favored the non-target semantic meaning of the homograph (IH). This condition directly contrasted with the correct EH condition where sentence-context primed the English reading of the homograph. We found that IH sentences, though semantically-incongruent in English, experienced a homograph interference effect in the N400 time-window, particularly in (1) high-cloze probability contexts that highly-constrained the Italian meaning of the homograph, (2) bilinguals with a larger gap in relative Italian-English proficiency, and (3) bilinguals with lower English proficiency scores.

Sentence context was also found to modulate N400 responses to correct cognate sentences, but cloze-probability operated in the opposite direction than for IH sentences; cognates were found to be more facilitatory in low cloze-probability contexts where the target-word was not readily predictable from its preceding context. In such cases, convergence of both languages on a common semantic meaning proved to facilitate comprehension, leading to a reduction in the N400 relative to correct EH sentences.

Interestingly, these co-activation effects were only detectable with ERPs and our behavioral results were not fully in line with those from previous studies. We found an overall slower reaction time in bilingual groups compared to monolinguals, possibly reflecting a more elaborated process of word-identification and language selection due to the ambiguity of words that are shared crosslinguistically. However, contrary to much of the literature, we did not find longer RTs in response to homographs and shorter RTs for cognates. Another surprising finding, for example, was that Attriters rated CC conditions less favorably than the other two participant groups (although they did rate them more favorably than EH sentences). Acceptability judgments are provided at the *end* of a given sentence, after the process of lexical access and semantic integration has unfolded and the lexical system has already selected the target-language meaning of the ambiguous word. Thus, it is likely that the initially parallel activation of crosslinguistic meanings has been resolved and has decayed by the end of the sentence, especially since our target-words were not sentence-final and were followed by three words plus a response prompt before participants made their judgment. It is not surprising, then, that ERP patterns were more sensitive to co-activation effects than our behavioral data (see Thierry & Wu, 2007 for a similar discussion). However, it is also likely that the embedding of our words into sentence contexts and the task of judging their acceptability put our bilingual participants into more of a semantic-interpretation mode (which was indeed our goal), and did not create the same type of decisional conflict as a lexical decision task. The degree to which task might influence the emergence of co-activation effects in behavior vs. at different stages during the time-course of processing (e.g., N400 vs. P600) is a question that would benefit from future investigations.

4.5. Future directions

Although many aspects of our study are novel, a number of questions would benefit from follow-up studies in the near future. The most obvious step forward would be to conduct the same experiment but in Italian rather than English (L1 for Attriters, L2 for Italian-learners), with the same aims and research questions in mind. Examining lexical access in the L1 of Attriters would clarify whether co-activation in the L2-to-L1 direction is more likely to occur in Attriters than in our current group of non-dominant L2 learners, and would therefore replicate Attriters' anecdotal complaints of lexical-semantic intrusions from the L2 onto their L1. Our false-friends production task is one approach to these questions, but it would be crucial to conduct such an experiment with ERPs.

Another question concerns the role of word-frequency and crosslinguistic frequency between homograph / cognate meanings). In designing our experiment, we obtained and coded the cross-linguistic frequency information our target-words (e.g., whether relative-frequency was high in English and in Italian, *High-Low*, *Low-Low* or *Low-High*). A next step would therefore be to examine frequency effects and possible interactions between frequency, cloze probability and language proficiency and/or exposure. Some studies have shown that frequency modulates competition during bilingual lexical access (e.g., Dijkstra, Van Jaarsveld & Ten Brinke, 1998; Dijkstra et al., 2010; Kerkhofs et al., 2006), and that frequency effects may be fully accounted for by language proficiency (Diependaele, Lemhöfer & Brysbaert, 2013) and/or language exposure (Whitford & Titone, 2012). Extending these investigations to the realm of L1 attrition could shed light on whether frequency effects in an L2 may parallel those observed in an attriting L1, given that the language of more frequent exposure for attriters is actually the L2.

Finally, although studies involving homographs and cognates have their merit, it will be necessary to move away from investigations of conditions that explicitly promote a dual-language processing mode (see arguments discussed in Wu & Thierry, 2010), especially since highly-proficient L2 learners but particularly Attriters are consciously aware of these pitfalls in their own production.

5. CONCLUSIONS

The present study examined real-time lexical-semantic processing in Italian-English Attriters (dominant in L2) and English-Italian learners (dominant in L1), in an attempt to determine which factors influence the "native-like-ness" of L2 processing and the degree of crosslinguistic co-activation of the non-target language. Italian-English Attriters were found to elicit N400 effects that were indistinguishable in amplitude, latency and scalp distribution from both monolingual and bilingual native-speakers of English for conditions that constituted lexical-semantic violations in English. On those conditions where we created a conflict between English and Italian readings of the target-word or created a sentence context that was congruent with *both* languages, we observed parallel activation of Italian meanings in the real-time processing mechanisms of both bilingual groups, albeit with differences in the time-course of co-activation. While Attriters showed Italian-homograph interference and cognate facilitation effects on both the N400 and P600, English-Italian learners' responses only converged on those of Attriters in the P600 time-window. Instead, English-Italian learners showed a tendency towards a cognate inhibition effect in the N400 window. Co-activation effects on the N400 were found to be modulated not only by target-language (English) proficiency, but by bilinguals' Italian proficiency *relative* to English.

In the P600 interval, both bilingual groups showed increased P600 amplitudes for homograph conditions compared to cognate conditions, whereas the correct cognate (CC) condition was least positive in amplitude. ERPs proved to be more sensitive to cross-linguistic effects than an end-of-sentence acceptability rating task. Our study is among the first to report P600 effects in an examination of non-selective bilingual lexical processing, and to find that homograph sentences that were correct in the target-language also elicited a P600 effect. We discussed this finding in the context of current models of non-selective bilingual lexical access and in line with the "conflict monitoring" theory of the P600.

Ours is also the first ERP study to examine how a shift in L1 dominance and proficiency brought on by changes in linguistic environment and language exposure (i.e., attrition) affect mechanisms of bilingual lexical access during sentence processing. Although we did not explicitly test L1 attrition effects in the present study, we found that L1 co-activation effects during L2 comprehension were influenced by background factors such as proficiency, language exposure and length of residence, with more English native-like ERP responses (i.e., less co-activation) associated with increased English proficiency, decreased L1 language exposure, and increased length of residence. Crucially, ERP response patterns were not predicted by AoA for either of the bilingual groups.

Our results are consistent with existing research and theories of non-selective lexical activation in bilinguals, while contributing to the field with a number of novel elements in our experimental design. We also advocate in favor of ongoing neuroplasticity for language, even in adulthood, and provide additional evidence for the crucial role of proficiency in shaping the brain's responses to language.

ACKNOWLEDGEMENTS

The authors thank Dr. John Drury for valuable feedback on experimental design, as well as research assistants Linna Jin and Kristina Maiorino for their invaluable help with data collection. K.K was supported by a Vanier Canada Graduate Scholarship from the Canadian Institutes of Health Research (http://www.vanier.gc.ca/en/home-accueil.html), as well as a Richard Tomlinson Doctoral Fellowship awarded by the Faculty of Medicine of McGill University (https://www.mcgill.ca/gps/funding/students-postdocs/students/tomlinson). This research was supported by grants awarded to K.S. by the Canada Research Chair program and the Canada Foundation for Innovation (CRC/CFI; project # 201876; http://www.chairs.gc.ca/), the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPGP 312835-05 and RGPIN 402678-11; http://www.nserc-crsng.gc.ca/index_eng.asp), the Social Sciences and Humanities Research Council of Canada (SSHRC; # 435-2013-2052; http://www.sshrc-crsh.gc.ca/Default.aspx), and the Fonds de Recherche Société et Culture, Québec (FQRSC; # 2010-SE-103727; http://www.fqrsc.gouv.qc.ca/fr/accueil.php). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Study 3: Tables and figures

Table 1. *Experimental sentences are provided each of the six conditions. The abbreviations of the condition names describes the levels of Context (English (E), Italian (I), Cognate (C)) and Target word (Homograph (H), Cognate (C)). The target noun is underlined. The asterisk marks a lexical-semantic violation in English. Acceptability between English and Italian is compared in the last two columns. Note that cognates share their meaning across languages, whereas homographs only share their form (parole: "prisoner's release" vs. "words")*

CONDITION	Example	English acceptability	Italian acceptability
EH (English homograph context + Homograph)	The prisoner violated the **parole** after a month.	✓	✗
EC (English homograph context + Cognate)	The prisoner violated the ***cabin** after a month.	✗	✗
IH (Italian homograph context + Homograph)	The poet rhymed the ***parole** of the verse.	✗	✓
IC (Italian homograph context + Cognate)	The poet rhymed the ***cabin** of the verse.	✗	✗
CH (Cognate context + Homograph)	The hikers rented the ***parole** in the mountains.	✗	✗
CC (Cognate context + Cognate)	The hikers rented the **cabin** in the mountains.	✓	✓

Table 2. *Group means (standard deviation) for English proficiency and control tasks*

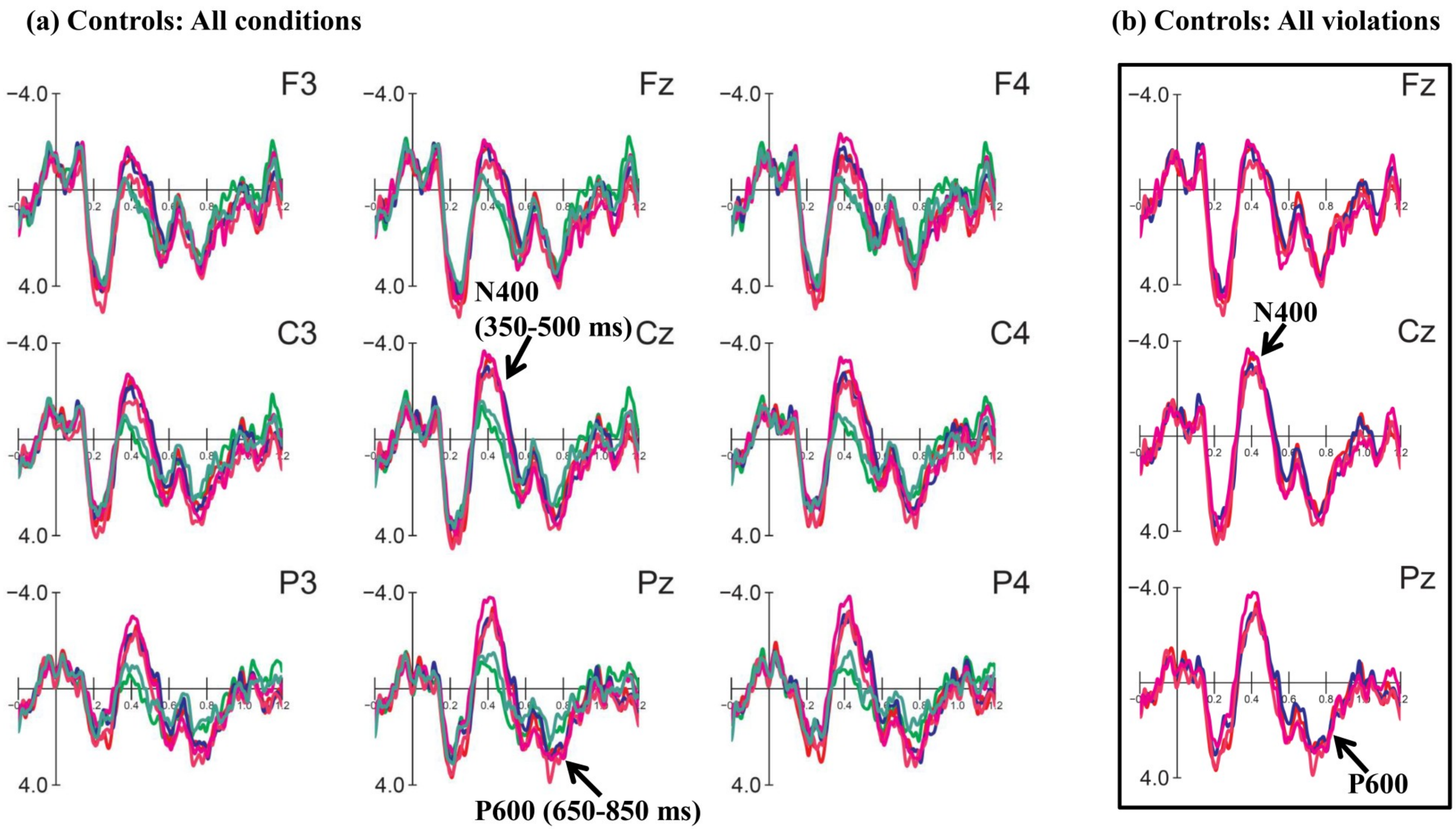
BEHAVIORAL MEASURES	ENGLISH CONTROLS (n = 30)	ITALIAN-ENGLISH ATTRITERS (n = 24)	ENGLISH-ITALIAN LEARNERS (n = 20)
Self-report of proficiency (7 point scale)	7 (0)	5.7 (0.8)	6.9 (0.1)
Listening comprehension	7 (0)	6.0 (0.8)	7 (0)
Reading comprehension	7 (0)	6.4 (0.6)	7 (0)
Pronunciation	7 (0)	5.0 (1.0)	7 (0)
Fluency	7 (0)	5.7 (1.0)	7 (0)
Vocabulary	7 (0)	5.3 (1.1)	6.9 (0.2)
Grammar	7 (0)	5.5 (1.1)	6.9 (0.3)
C-test (%)	90.6 (7.3)	86.3 (10.6)	89.3 (9.0)
Error-detection test (%)	80.2 (8.5)	61.0 (16.2)	80.7 (12.9)
Verbal semantic fluency (average of 2 categories)	23.9 (5.7)	19.1 (3.8)	22.8 (3.5)
Reading fluency (# correct in 3 minutes)	89.9 (7.8)	70.4 (16.8)	84.3 (17.8)
Working memory			
Correct	11.2 (2.9)	11.6 (3.7)	12.1 (2.7)
Span	5.7 (1.2)	5.8 (1.8)	5.8 (1.0)

Table 3. *Participants' average ratings (from 1 (unacceptable) to 5 (perfect)) and average reaction times (in milliseconds) for each of the six conditions. Standard deviation is reported in parentheses. The asterisk marks a lexical-semantic violation in English.*

CONDITION *Average rating*	**ENGLISH CONTROLS (n = 30)**	**ITALIAN-ENGLISH ATTRITERS (n = 24)**	**ENGLISH-ITALIAN LEARNERS (n = 20)**
EH	4.49 (0.39)	3.96 (0.59)	4.45 (0.31)
***EC**	2.13 (0.48)	2.28 (0.58)	2.23 (0.54)
***IH**	2.14 (0.47)	2.36 (0.61)	2.33 (0.58)
***IC**	1.94 (0.44)	2.09 (0.57)	2.02 (0.60)
***CH**	1.99 (0.50)	2.07 (0.62)	2.09 (0.60)
CC	4.43 (0.40)	4.15 (0.35)	4.50 (0.26)
CONDITION *Average RT*	**ENGLISH CONTROLS (n = 30)**	**ITALIAN-ENGLISH ATTRITERS (n = 24)**	**ENGLISH-ITALIAN LEARNERS (n = 20)**
EH	726.00 (488.92)	1328.61 (934.32)	1101.80 (627.17)
***EC**	952.39 (493.82)	1443.04 (783.95)	1423.50 (855.32)
***IH**	966.55 (506.55)	1391.04 (891.45)	1572.60 (861.78)
***IC**	802.75 (412.02)	1326.78 (837.53)	1246.10 (618.77)
***CH**	894.32 (480.82)	1338.04 (906.70)	1342.55 (774.81)
CC	699.65 (547.66)	1190.00 (687.68)	1018.05 (648.31)

Table 4. *Pearson coefficients for significant correlations between proficiency measures and behavioral ratings, as well as with ERPs amplitudes. N400 amplitudes for each relevant difference-wave were calculated at Cz between 350-500 ms. P600 amplitudes were calculated at Pz between 650-900 ms. * p < 0.05; ** p < 0.01; ***p < 0.001; ns: p > 0.1*

	ENGLISH				**ITA - ENG**
Ratings	**C-test**	**Err.-Detection**	**Sem. fluency**	**Reading fluency**	**Rel. proficiency**
English violations	- 0.460***	- 0.482***	- 0.373***	- 0.472***	ns
IH – EH	- 0.525***	- 0.541***	- 0.414***	- 0.478***	0.565***
EH – CC	0.413***	0.443***	0.205*	0.219*	- 0.403**
ERP amplitudes	**C-test**	**Err.-Detection**	**Sem. fluency**	**Reading fluency**	**Rel. proficiency**
N400 English violations - Correct	ns	ns	ns	ns	ns
N400 IH – EH	- 0.424***	- 0.247***	- 0.255***	ns	0.214 (*p* = 0.08)
N400 EH – CC	0.299**	0.228*	0.280**	0.233*	- 0.372**
P600 English violations - Correct	ns	ns	ns	ns	ns
P600 Hom. targets – Cogn. targets	ns	ns	ns	ns	ns



EH: The prisoner violated the parole after a month.
EC: The prisoner violated the *cabin after a month.
IH: The poet rhymed the *parole of the verse.
IC: The poet rhymed the *cabin of the verse.
CH: The hikers rented the *parole in the mountains.
CC: The hikers rented the cabin in the mountains.

Figure 1. ERPs elicited in Controls in response to each of the six conditions. ERPs are time-locked to the onset of the target noun, which is underlined in the legend. The asterisk marks the point of a lexical-semantic violation in English. The two correct conditions (EH, CC) are depicted in shades of green. The four conditions which constitute lexical-semantic violations in English are depicted in shades of red and pink (EC, IC, CH). The IH condition which is a violation in English but semantically correct in Italian is depicted in blue. Time ranges (in milliseconds) depicted on the x-axis are relative to the onset of the verb (0 ms). Negative values are plotted up. Native-English Controls show a large centro-parietal N400 effect in response to all violation conditions followed by a posterior P600 effect, relative to the two correct conditions. **Figure 1b** focuses on ERPs elicited on the midline in response to the four violation conditions. As expected for English-Controls, the violation conditions did not differ from one another in either of the time-windows.

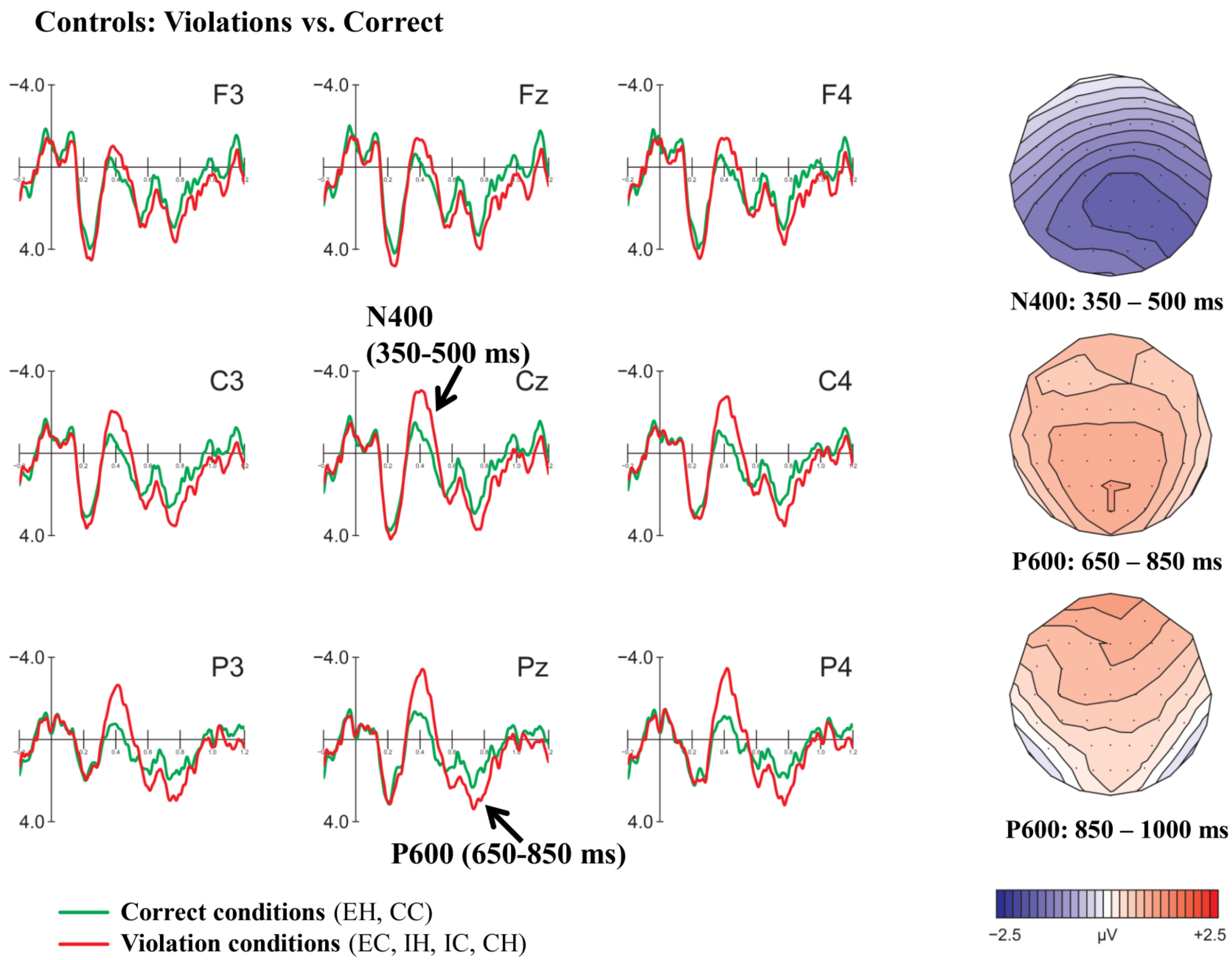


Figure 2. An illustration of the overall "violation effect" elicited in Controls on the target noun. All four violation conditions are collapsed into one violation condition (red) and compared to the two correct conditions, also collapsed into one correct condition (green). The voltage maps confirm the centro-parietal scalp distribution of the N400 (350-500 ms) and the posterior prominence of the P600 (650-1000 ms).

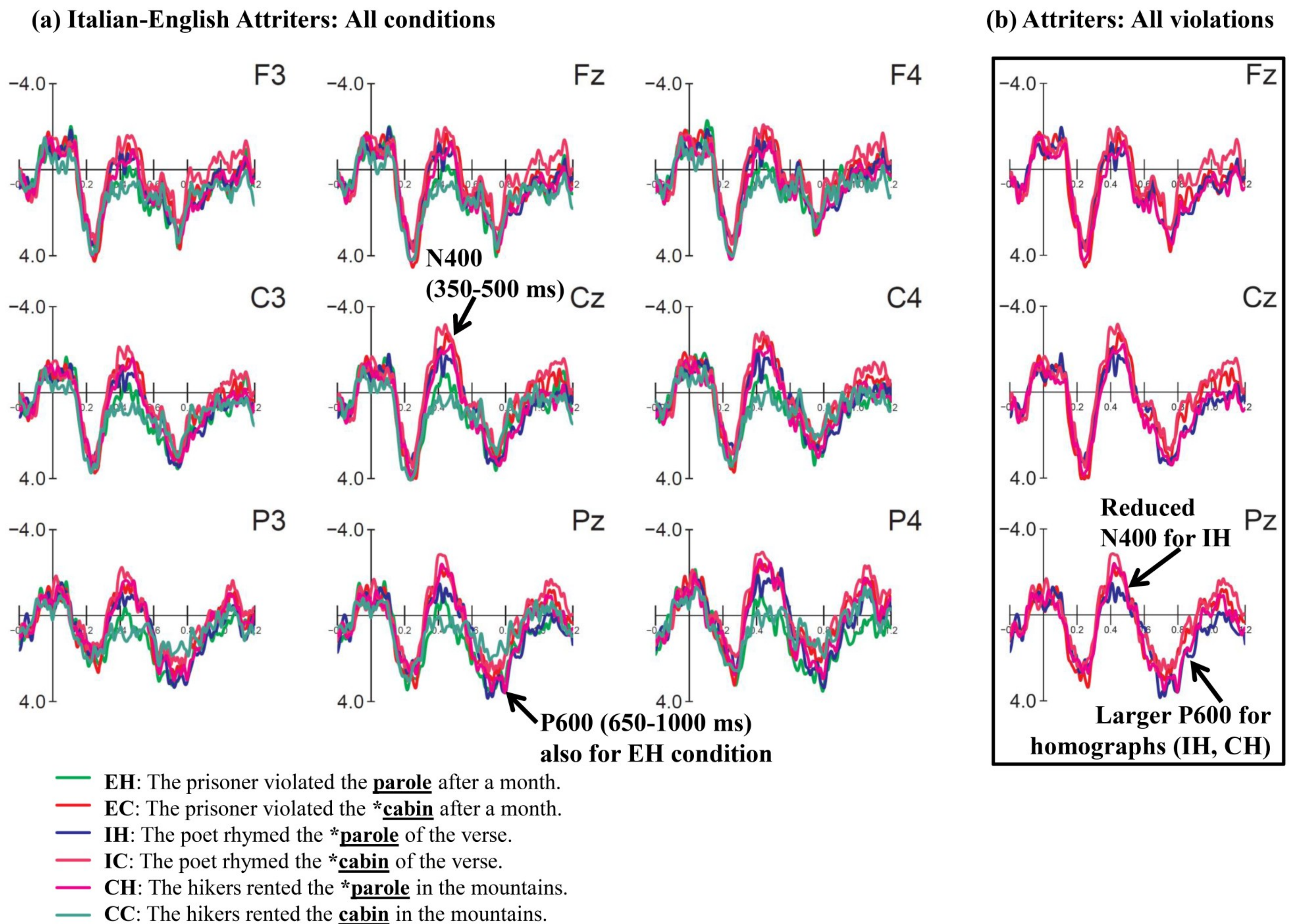

(a) Italian-English Attriters: All conditions
(b) Attriters: All violations
F3
Fz
F4
C3
Cz
C4
P3
Pz
P4
-4.0
4.0
N400
(350-500 ms)
P600 (650-1000 ms)
also for EH condition
Reduced
N400 for IH
Larger P600 for
homographs (IH, CH)
EH: The prisoner violated the parole after a month.
EC: The prisoner violated the *cabin after a month.
IH: The poet rhymed the *parole of the verse.
IC: The poet rhymed the *cabin of the verse.
CH: The hikers rented the *parole in the mountains.
CC: The hikers rented the cabin in the mountains.

Figure 3. ERPs (time-locked to the target noun) elicited in Attriters in response to each of the six conditions. The asterisk marks the point of a lexical-semantic violation in English. The two correct conditions (EH, CC) are depicted in shades of green. The four conditions which constitute lexical-semantic violations in English are depicted in shades of red and pink (EC, IC, CH). The IH condition which is a violation in English but semantically correct in Italian is depicted in blue. Attriters show a large centro-parietal N400 effect in response to all English violation conditions, but a reduced N400 for IH violations (see **Fig 3b**). The two correct conditions also differ, as the CC condition (correct in both English and Italian) shows a reduced N400 relative to the EH condition. The P600 was modulated by "target type", with larger P600 amplitudes for homograph targets (including the correct EH condition) than for cognate targets.

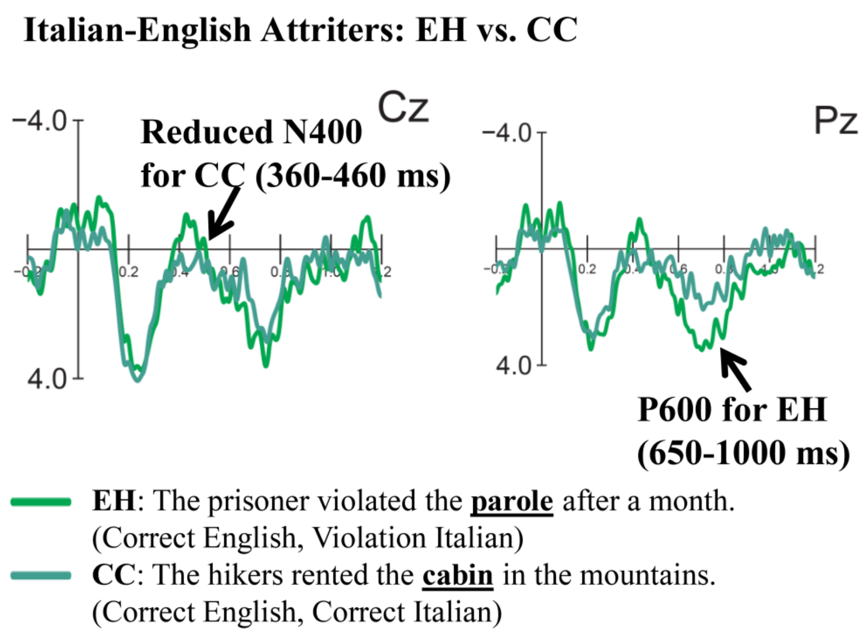


Figure 4. ERPs elicited in Attriters in response to the two correct conditions: EH (light green) and CC (blue-green). Between 360-460 ms at Cz and Pz, the N400 effect for CC sentences is reduced relative to EH sentences, although both are correct sentences. The correct EH condition also elicits a P600 in Attriters, relative to the correct CC condition.

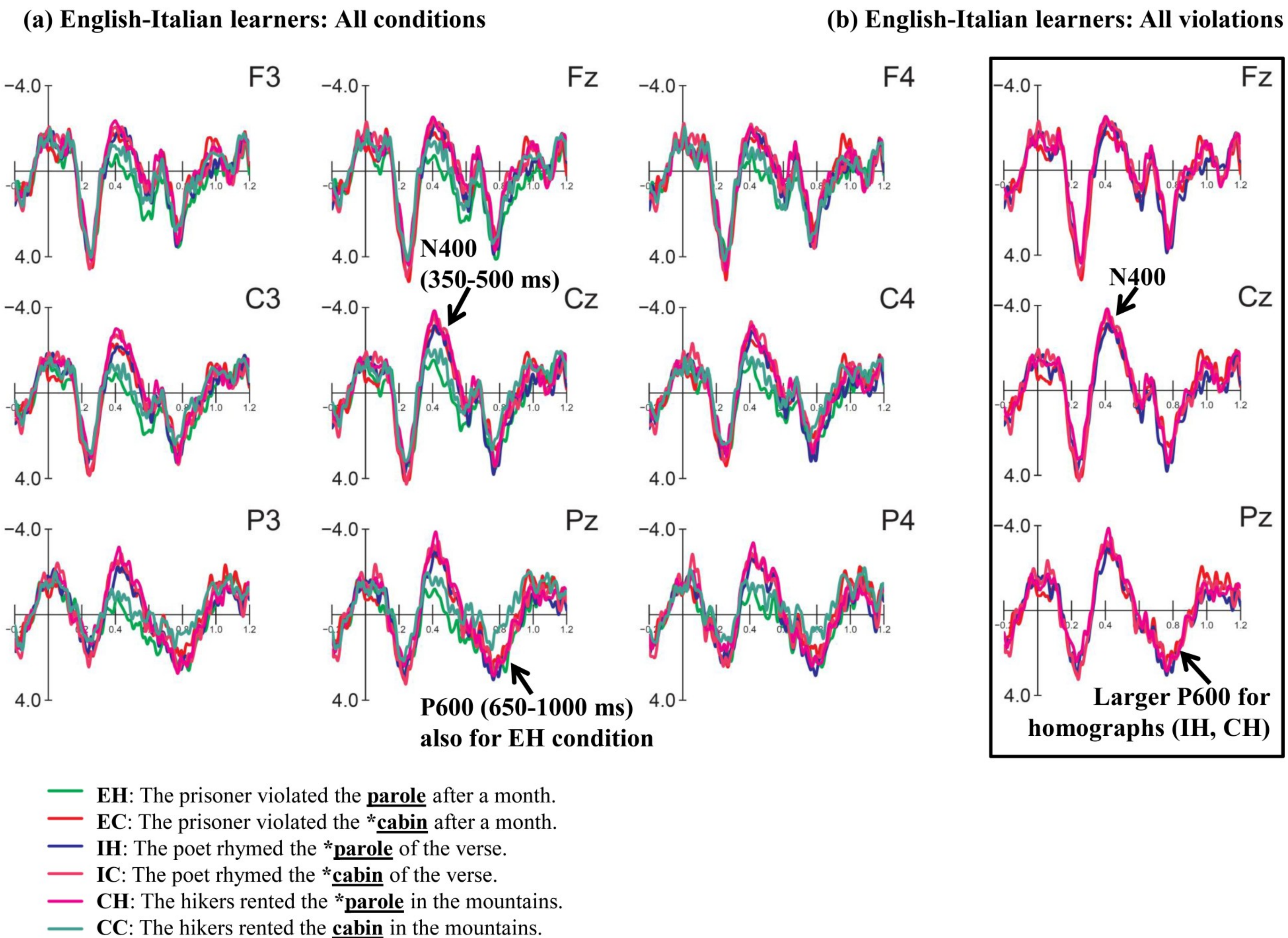

(a) English-Italian learners: All conditions
(b) English-Italian learners: All violations
F3
Fz
F4
C3
Cz
C4
P3
Pz
P4
-4.0
4.0
0.4
1.0
1.2
N400 (350-500 ms)
P600 (650-1000 ms) also for EH condition
N400
Larger P600 for homographs (IH, CH)
EH: The prisoner violated the parole after a month.
EC: The prisoner violated the *cabin after a month.
IH: The poet rhymed the *parole of the verse.
IC: The poet rhymed the *cabin of the verse.
CH: The hikers rented the *parole in the mountains.
CC: The hikers rented the cabin in the mountains.

Figure 5a. ERPs elicited in English-Italian bilinguals (Italian L2 learners) in response to each of the six conditions. ERPs are time-locked to the onset of the target noun, which is underlined in the legend. The asterisk marks the point of a lexical-semantic violation in English. The two correct conditions (EH, CC) are depicted in shades of green. The four conditions which constitute lexical-semantic violations in English are depicted in shades of red and pink (EC, IC, CH). The IH condition which is a violation in English but semantically correct in Italian is depicted in blue. Like Controls, English-Italian bilinguals show a large centro-parietal N400 effect in response to all violation conditions. Unlike Attriters, the CC and IH conditions do not show a reduced N400 effects. A P600 effect follows for all violation conditions **(*Figure 5b*),** with a larger posterior P600 elicited for homograph target words (EH, IH and CH), as was the case for Attriters.

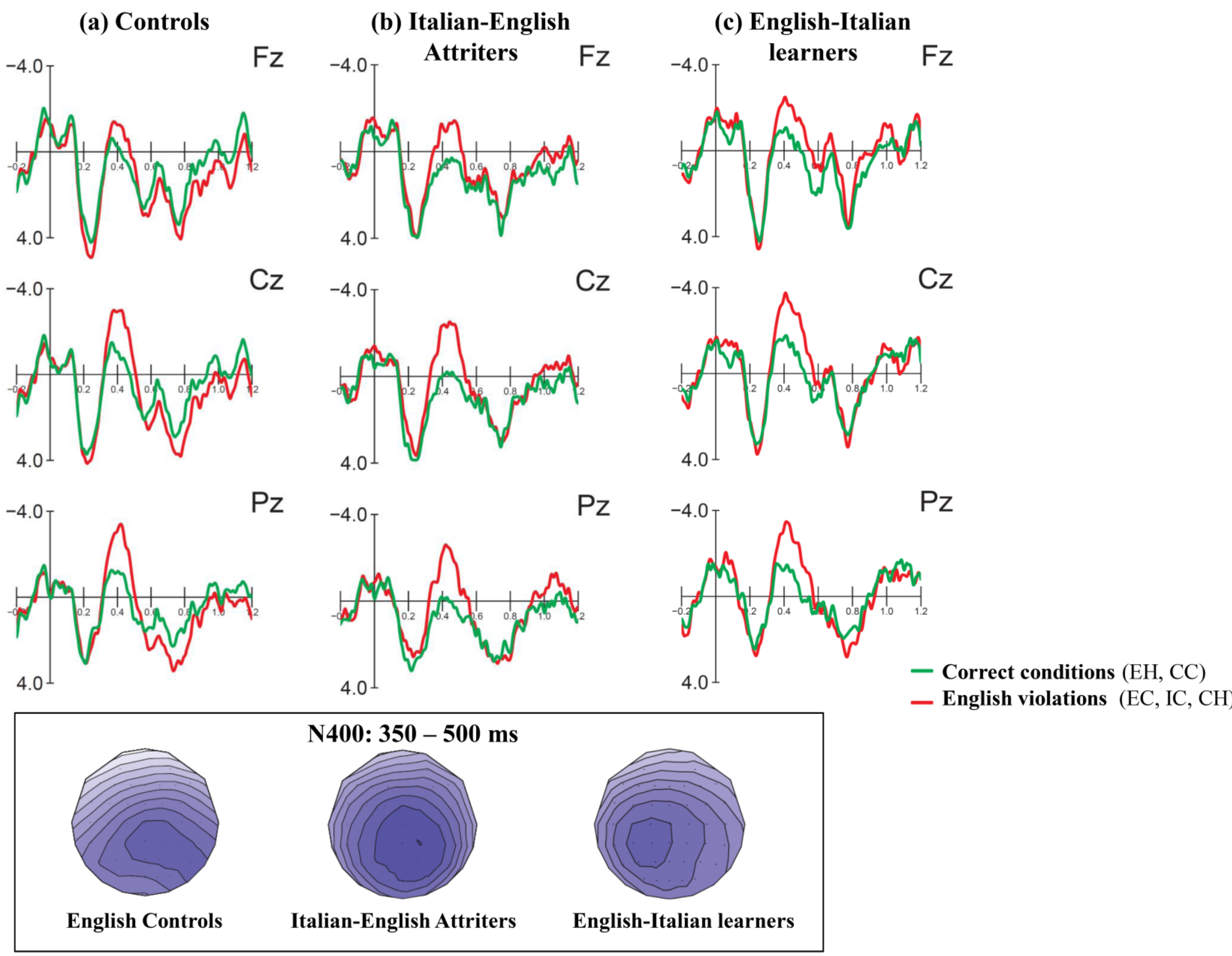


Figure 6. A comparison of all three groups on English violations (EC, IC, CH collapsed) relative to correct sentences (EH, CC collapsed), to illustrate that Italian-English Attriters elicited a statistically indistinguishable N400 in response to English lexical-semantic violations as native-speakers of English ((a) monolingual Controls and (c) bilingual Italian-learners). Note that the P600 appears smaller for the Attriters and Italian-learners because, in those two groups, the correct EH condition also elicited a P600 effect. The voltage maps confirm the centro-parietal scalp distribution of the N400 (350-500 ms) for all three groups.

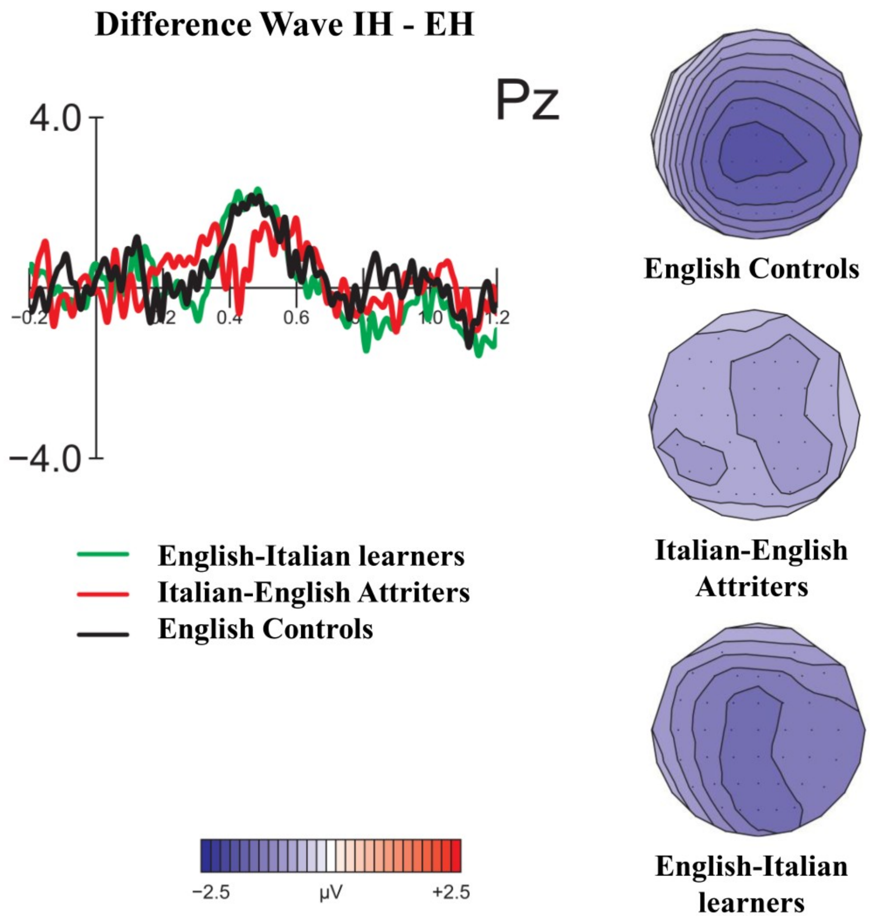


Figure 7. A comparison of the difference waves and voltage maps for the N400 effect elicited by IH violations relative to correct EH sentences, for all three groups. English Controls are represented in black, Italian-English Attriters in red and English-Italian learners in green. Italian-English Attriters only showed a marginally significant N400 effect relative to the correct control condition (EH), while the two other groups showed a robust N400 response for IH sentences (that was indistinguishable from the other violation conditions).

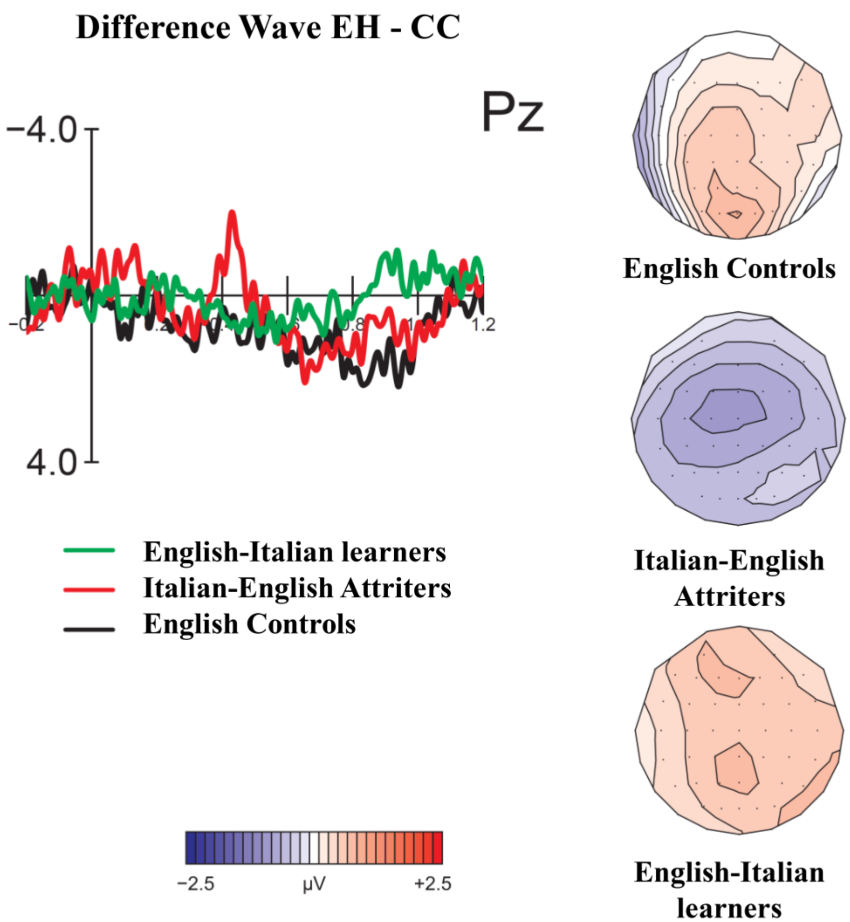


Figure 8. A comparison of all three groups on their N400 responses to the two correct conditions EH vs. CC. Italian-English Attriters were the only group to show an N400 difference between the two correct conditions, namely an enhanced N400 for the EH condition (a reduced N400 for CC). Thus, EH sentences, although correct in English, elicited a larger N400 than CC sentences in Attriters.

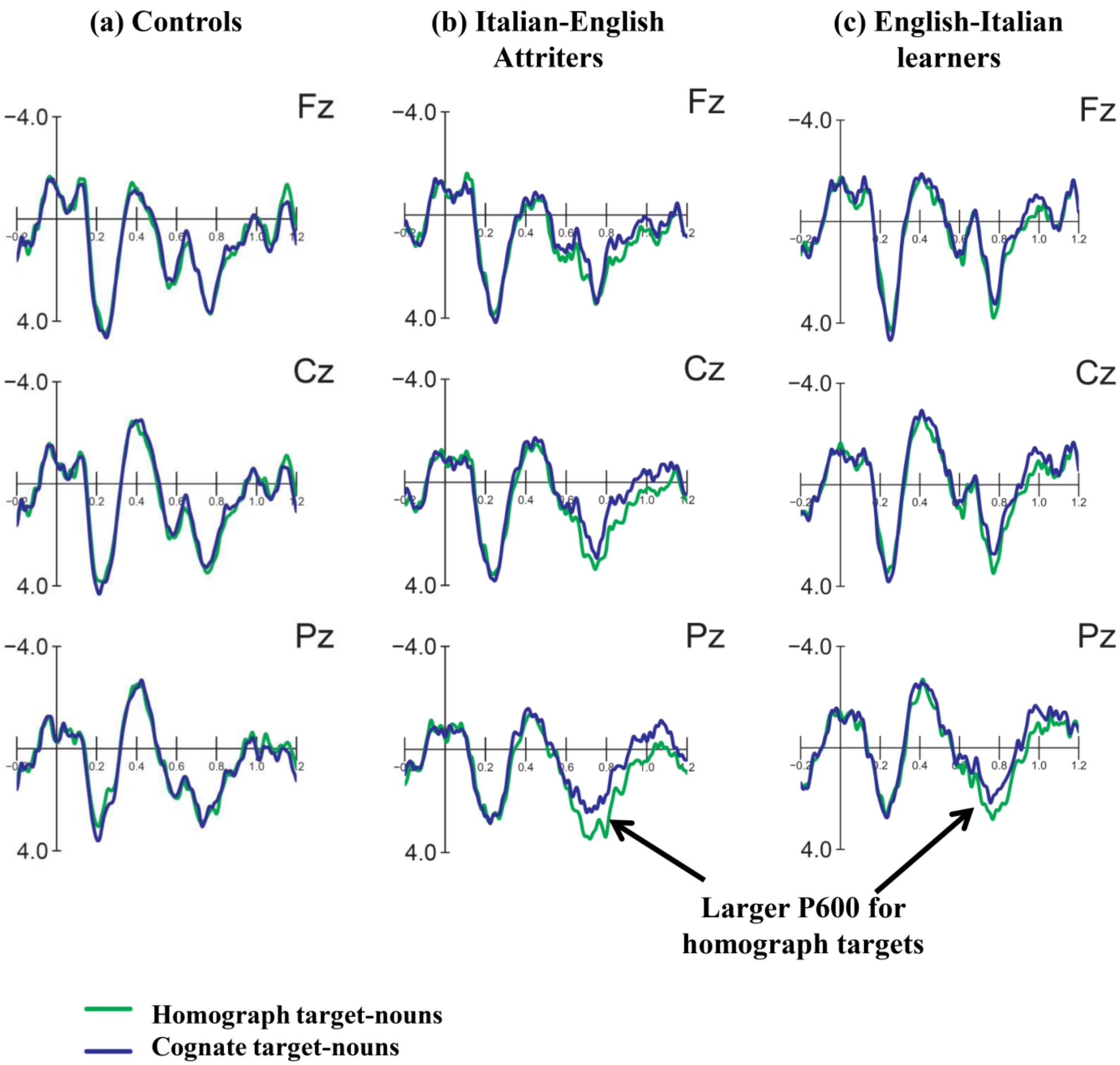


Figure 9. A comparison of all three groups' P600 responses based on target type (homographs vs. cognates). Both bilingual groups showed a larger P600 in response to homograph targets than cognate targets, while this pattern was absent in monolingual Controls. The two bilingual groups did not differ significantly from one another. The effect of target type persisted into the late P600 time-window between 850-1000 ms.

Study 3: References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39* (3), 445-459.

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition, 35*, 953–965.

Bergmann, Berends, Brouwer, Meulman, Seton, Sprenger, Stowe & Schmid (2013). Processing of gender in L2 acquisition and L1 attrition of German: Evidence from event-related potentials. Talk presented at the Workshop on Neurobilingualism, Groningen, The Netherlands, August 25-27, 2013.

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C. and Thornton, A. M. (2005). Corpus e Lessico di frequenza dell'Italiano Scritto (CoLFIS).

Birdsong, D. (2006). Dominance, proficiency, and second language grammatical processing. *Applied Psycholinguistics*, *27*(01), 46–49.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*(13), 2492–2511.

Brenders, P., Van Hell, J. G., & Dijkstra, A. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology, 109*, 383–396.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(01), 3–42.

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

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1283–1296.

De Bot, K. (1996). Language loss. In H. Goebl, P. Nelde, Z. Stary, & W. Wölk (Eds.), *Contact linguistics: An international handbook of contemporary research*, Vol. 1 (pp. 579 – 585). Berlin: Walter de Gruyter.

De Bruijn, E., Dijkstra, T., Chwilla, D., & Schreifers, H. (2001). Language context effects on interlingual homograph recognition: Evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition, 4,* 155–168.

De Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and non-cognates in compound bilinguals. *Journal of Memory and Language, 30*, 90-123.

Diependaele K., Lemhöfer K., Brysbaert M. (2013). The word frequency effect in first and second language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology, 66*(5), 843-863.

Dijkstra, A., De Bruijn, E., Schriefers, H. J. & Ten Brinke, S. (2000). More on interlingual homograph recognition: Language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition, 3*, 69–78.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*, 496–518.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language, 62*, 284–301.

Dijkstra, T., Timmermans, M., & Schreifers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*, 445–464.

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*, 175–197.

Dijkstra, T., & Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*, 51–66.

Duñabeitia, J.A., Perea, M., & Carreiras, M. (2010). Masked translation priming with highly proficient simultaneous bilinguals. *Experimental Psychology, 57*, 98-107.

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 663–679.

Elston-Güttler, K. E. (2000). An enquiry into cross-language lexical–conceptual relationships and their effect on L2 lexical processing. PhD Dissertation, University of Cambridge.

Elston-Güttler, K. E., Gunter, T. C. (2008). Fine-tuned. Phonology and Semantics Affect First- to Second-language Zooming In. *Journal of Cognitive Neuroscience, 21*(1), 180-196.

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research, 25*, 57–70.

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, *17*, 1593–1610.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001).Syntactic parsing preferences and their on-line revisions: A spatio temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*, 305–323.

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*(1), 529–534.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439-483.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266.

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

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, *44*, 2–31.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98-110.

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. *Poster presentation at the Donostia Workshop on Neurobilingualism, Basque Center on Cognition, Brain and Language*, Donostia-San Sebastián, Spain, Sept 30 – Oct 2, 2010.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013a). When the second language takes over: ERP evidence of L1-attrition in morphosyntactic processing. Talk presented at the International Conference on Multilingualism, Montreal, Canada, October 24-25, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013b). My Italian is not what it used to be: Investigating the neural correlates of L1 attrition and late L2 acquisition. Poster presented at the Workshop on Neurobilingualism, Groningen, The Netherlands, August 25-27, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014a). The case of the non-native-like first language: ERP evidence of L1-attrition in lexical and morphosyntactic processing. Symposium presentation, The 17th World Congress on Psychophysiology, Hiroshima, Japan, September 23-27, 2014.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014b). Neurophysiological correlates of L1 attrition and L2 acquisition: A continuum based on proficiency. Poster presented at the Society for the Neurobiology of Language, Amsterdam, Netherlands, August 27-29, 2014.

Kasparian, K., Vespignani, F., & Steinhauer, K. (under review). First-language attrition induces changes in online morphosyntactic processing and re-analysis: An ERP study of number agreement in complex Italian sentences.

Keatley, C.W., Spinks, J.A., & de Gelder, B. (1994). Asymmetrical cross-language priming effects. *Memory and Cognition, 22*, 70–84.

Keijzer, M. C. J. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. Utrecht: LOT publications.

Kerkhofs, R., Dijkstra, T., Chwilla, D. J., & de Bruijn, E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research, 1068,* 170–183

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

Köpke, B. (1999). *L'attrition de la première langue chez le bilingue tardif : Implications pour l'étude psycholinguistique du bilinguisme*. Unpublished Doctoral Dissertation, Toulouse: Université de Toulouse-Le Mirail.

Köpke, B. (2002). Activation thresholds and non-pathological L1 attrition". In *Advances in the Neurolinguistics of Bilingualism. Essays in Honor Of Michel Paradis*, F. Fabbro (ed), 119-142. Undine: Forum.

Köpke, B., & Schmid, M. S. (2004). Language Attrition: The Next Phase. In M. S. Schmid, B. Köpke, M. Keijzer & L. Weilemar, L. (Eds.), *First Language Attrition: Interdisciplinary perspectives on methodological issue*s (pp. 1-43). Amsterdam: John Benjamins.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Kroll J. F., Van Hell J. G., Tokowicz N., Green D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism, 13*, 373–381

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science, 207, 203–205.

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

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

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition, 34*(4), 533-550.

Libben, M. R, & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 381–390.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immerse in a second language. *Psychological Science*, *20*, 1507–1515.

Midgley, K., Holcomb, P. J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience*, *23*, 1634–1647.

Montrul, S. (2008). *Incomplete acquisition in bilinguals: Re-examining the age factor.* Amsterdam: John Benjamins.

Mulatti, C., e Andriolo, S. (in preparation). Algoritmo calcola vicini. http://dpss.psy.unipd.it/claudio/vicini.php

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-1228.

Opitz, C. (2011). *First language attrition and second language acquisition in a second language environment.* PhD dissertation. Centre for Language and Communication Studies. Trinity College Dublin. Dublin.

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

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech, *Language and Cognitive Processes, 8*(4), 413-437.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning*, *56*(1), 199–230.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*(6), 509–521.

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

Pavlenko, A. (2000). L2 influence on L1 late bilingualism. *Issues in Applied Linguistics, 11*(2), 175–205.

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language, 68*, 315-332.

Penfield, W., & Roberts, L. (1959). *Speech and Brain Mechanisms.* New York: Athenaeum.

Perani, D., Paulesu, E., Sebastián-Gallés, N., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain, 121*, 1841-1852.

Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across language with highly proficient bilinguals. *Journal of Memory and Language, 58*, 916–930.

Poarch, G. J., & Van Hell, J. G. (2012). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *111*, 419–438.

Potter, M. C, So, K.-F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*, 23-38.

Prat, C. S. (2011). The brain basis of individual differences in language comprehension abilities. *Language & Linguistics Compass, 5,* 635–649.

Rossi, S., Gugler, M., Friederici, A., & 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*(12), 2030–2048.

Sanchez-Casas, R. M, Davis, C. W, Garcia-Albea, J. E. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology, 4*, 293–310.

Scarborough, D. L., Gerard, L., & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior, 23*, 84–99.

Schmid, M. S. (2011). *Language attrition*. Cambridge: Cambridge University Press.

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, 1–20.

Schmid, M. S., & Keijzer, M. (2009). First language attrition and reversion among older migrants. *International Journal of the Sociology of Language, 200*, 83–101.

Schmid, M. S, & B. Köpke. (2008). L1 attrition and the mental lexicon, in: Aneta Pavlenko (ed.), *The Bilingual Mental Lexicon*. Clevedon: Multilingual Matters, pp. 209 – 238.

Schwartz, A. I. & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*, 197–212.

Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and bilingual speech mode: The effect on lexical access. *Memory and Cognition, 12*, 380-386.

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.

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*(02), 277–293.

Thierry, G., & Wu, Y. J. (2007). Brain potential reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*, 12530–12535.

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.

van Assche, E., Drieghe, D., Duych, W., Welvaert, M., & Hartsuiker, R. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language, 64*(1), 88-107.

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., and Vissers, C. T. W. M. (2009). Monitoring in language perception. *Lang. Linguist. Compass* 3, 1211–1224.

van de Meerendonk, N., Kolk, H. H., Vissers, C. Th. W. M, & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82.

Van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*(3), 431–451.

Van Hell, J. G., & Dijkstra, A. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*, 780–789.

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning, 62* (2), 148–171.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. *Brain Research, 1106*, 150–163.

Waas, M. (1996). *Language attrition downunder: German speakers in Australia*. Frankfurt: Peter Lang.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early settings of grammatical processing in the bilingual brain. *Neuron*, *37*, 159–170.

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.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale: Fourth Edition (WAIS–IV)*. New York: Psychological Corporation.

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain Responses before and after Intensive Second Language Learning: Proficiency Based Changes and First Language Background Effects in Adult Learners. *PLoS ONE*, *7*(12), 52318.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review, 19*(1), 73-80.

Wu, Y. J., & Thierry, G. (2010). Investigating bilingual language processing without artificial activation of participants' two languages. *Frontiers in Language Sciences, 94, 236-*

Yağmur, K. (1997). *First language attrition among Turkish speakers in Sydney*. Tilburg: Tilburg University Press.

Yilmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon, 7*(3), 249-274.

Zhao, X., Li, P., Liu, Y., Fang, X., & Shu, H. (2011). Cross-language priming in Chinese-English bilinguals with different second language proficiency levels. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

8. GENERAL DISCUSSION

The series of studies reported in this dissertation, as well as those conducted but not reported here, were designed to address two main questions pertaining to the neurocognition of language: (1) is there neurophysiological evidence of first-language (L1) attrition in a group of adult first-generation immigrants, as a result of predominant use and exposure to the late-acquired L2? (Attriters in **Study 1** on Italian morphosyntax and in **Study 2** on Italian lexical-semantics); and (2) to what extent are the neurocognitive mechanisms underlying real-time processing in late L2 learners similar to those observed in native-speakers? (Italian L2-learners in **Study 2** on Italian lexical-semantics and Attriters in **Study 3** on English lexical-semantics).

At the core of these questions is the still-controversial notion that a maturational decrease of neuroplasticity in childhood restricts the brain's ability to change with experience and to recruit the same neurocognitive substrates for L2 processing as those used for the L1. A crucial point to reiterate is that it is generally agreed upon that L2 learning and processing are subject to age-of-acquisition (AoA) effects, such that younger learners are typically more likely to exhibit native-like neurocognitive responses to the L2 than older learners. However, the debate centers around the *explanation* behind such AoA-effects; while it is widely claimed that maturational limits on neuroplasticity are the cause for differences in the processing patterns of late L2 learners, a late AoA is typically confounded with a slew of other important experiential factors, such as less exposure, less use/practice and lower proficiency in the L2 relative to one's dominant L1. The advantage of using the phenomenon of attrition as a lens into investigations of L1/L2 processing is that potential AoA-effects are not confounded with other factors such as less frequent exposure, infrequent use/practice and low proficiency relative to the L1, contrary to the majority of prototypical late L2 learners.

Overall, our studies provided insights into several key themes relevant to the study of both L1 and L2 processing: (1) the crucial role of proficiency-level in modulating the brain's responses to language, irrespective of whether the language was acquired as an L1 or an L2; (2) the nature of manifestations of L1 attrition (i.e., in which areas and to what degree?), and how the phenomenon may be operationalized for future studies; (3) the possibility of parallels between L1 attrition and L2 acquisition, and the viability of a proficiency-based continuum regardless of L1/L2 status, at least for some aspects of language-processing; (4) the potential for a dissociation between effects observed at the behavioral level and those reflected by ERPs during real-time processing, and the theoretical and methodological implications of such a discrepancy; (5) the functional significance of the P600 and its modulations in L1 attrition and L2 acquisition; and finally (6) the likelihood for continued neuroplasticity for language in adulthood, whereby the neurocorrelates of both L1 and L2 processing are amenable to change with experience.

These main themes, which tie directly into the research questions initially outlined in the *General Introduction*, will be discussed in turn below, in light of the relevant findings from each study. New questions raised by our studies, as well as areas of improvement for future work, will be highlighted at the end of the discussion. In order to avoid redundancy, the specific aims and findings of each individual study will not be detailed anew prior to the discussion of these major themes (please refer to pages 42-46 of the *General Introduction* for such an overview).

8.1. The role of proficiency-level in modulating ERP profiles of language processing

In line with research emphasizing the impact of proficiency on native-like L2 processing patterns (e.g., Bowden et al., 2014; Friederici, Steinhauer & Pfeifer, 2002; Morgan-Short et al., 2012; Newman et al., 2012; Osterhout et al., 2006; Rossi et al., 2006; Steinhauer, White & Drury, 2009), our findings confirmed that proficiency scores predicted the amplitude, scalp distribution, latency and/or duration of ERP correlates of language processing, not only in the case of late L2 acquisition, but *also* in native-speakers processing their L1.

In **Study 2**, which explored lexical access and integration of semantically-anomalous sentences and confusable Italian nouns, English-Italian late L2-learners were shown to be indistinguishable from native-speakers in the N400 responses they elicited in response to two kinds of lexical-semantic violations when they were in the high proficiency range; conversely, when L2 learners had lower Italian proficiency scores, N400 effects were smaller and more broadly-distributed over the scalp (*Mismatch* condition) or overlapping with correct control target-words (*Swap* condition).

In Study 1, which examined number agreement processing at two points in time during the comprehension of Italian sentences, we showed that Italian (L1) proficiency modulated the amplitude of the early P600 effect on both the verb (650-1000 ms) and on the modifier (650-900 ms), as well as of the N400-like negativity (300-500 ms) elicited on the modifier in response to the *xxy* violation conditions, where the modifier clashed in number with the two preceding elements (subject and verb) in a given sentence. Across both groups of Italian native-speakers, proficiency was found to influence both the amplitude (larger P600 in higher proficiency) and scalp distribution (more frontal N400/LAN in higher proficiency) of these effects.

Proficiency effects in L1 processing were also highlighted in **Study 2** (lexical-semantic processing in Italian), where N400 responses of native-Italian attriters to *Mismatch* and *Swap* violations were significantly dependent on their Italian proficiency scores – while high-proficiency attriters elicited a significant N400 effect in response to both *Mismatch* and *Swap* violations that was indistinguishable from monolingual L1-Italian speakers, Attriters who fell in the lower L1-proficiency range converged on the pattern shown by late L2 learners, and revealed more broadly-distributed N400 effects for *Mismatch* sentences (whereas the effect was focal at Cz/Pz in higher-proficiency individuals), and no significant N400 effects for the *Swap* condition relative to correct sentences. Higher L1-Italian proficiency scores were also associated with a more focal/posterior P600 distribution, a distributional difference that has been reported in L2 learners in previous studies (see Steinhauer et al., 2009). The role of proficiency in shaping the neurocognitive response patterns of individuals processing their native-L1 is not a new finding (Pakulak et al., 2004; Pakulak & Neville, 2010; Prat, 2011), although ours was one of the first ERP studies to explore L1-proficiency effects in a group of attriting L1-speakers, compared to non-attriting monolingual controls.

In **Study 3** – examining lexical-semantic processing of English sentences that contained Italian-English interlingual homographs and cognates – we showed that proficiency not only affects L1 and L2 processing during language-specific experimental tasks, but also modulates the dynamic interplay between the two language systems during online comprehension. English proficiency scores, as well as bilingual participants' relative Italian-English proficiency, modulated the degree of cross-linguistic influence or parallel-activation of the non-target language during both L1 (English-Italian L2 learners) and L2 (Italian-English attriters) processing.

Proficiency was also found to be associated with environmental factors particular to L1-attriters – their scores on Italian proficiency measures were found to be positively correlated with their amount of L1-Italian exposure (as reported in **Study 1**). Furthermore, the longer their length of residence (LoR), the lower their Italian proficiency scores on the *Error-detection* task – a task that was specifically designed for this dissertation with the purpose of more closely mirroring the task performed during ERP recording (i.e., reading sentences and detecting errors, many of which were deliberately similar to the structures we tested in our experiments). Thus, proficiency differences within the L1-attrition group were associated with experiential factors unique to their circumstances.

A final note about the impact of proficiency concerns its measurement. The studies in this dissertation explored the influence of proficiency with a methodological approach where (1) multiple proficiency measures were administered to participants, in order to obtain an informative picture of potential proficiency differences and of areas/tasks which may be more susceptible to group differences than others; (2) the same proficiency measures were also administered to native-speakers in their L1 (including monolingual controls); (3) proficiency measures were used to create subgroups of higher/lower proficiency individuals, with the purpose of first visualizing potential subgroup differences in ERP response patterns, and analyzing these effects categorically in order to relate to vast majority of L2 processing studies conducted to date; and (4) proficiency measures were also used as continuous scores in correlations that sought to determine how individual scores might predict ERP profiles.

Our results revealed that, across all studies, our subgroup analyses (i.e., by median split on multiple proficiency measures) largely coincided with results of correlational analyses where proficiency was treated as a continuous variable. Although this finding may be surprising if one thinks of all the methodological limitations of a median-split approach, it may be that first creating a composite measure of proficiency (based on multiple tests) on which to split participants into higher- and lower-ranges may already be a less arbitrary way to create proficiency categories than by basing oneself on one measure or, worse, solely on self-report measures.

What are the potential advantages of having employed a more comprehensive series of proficiency measures in bilingual and monolingual individuals? While our proficiency measures were correlated with one another and, for the large part, correlations of similar strength were reported between ERP patterns and each of the various proficiency measures, we found that, in some instances, a certain measure of proficiency proved to be more sensitive to differences at the neurocognitive level. For example, in **Study 3**, the only proficiency measure that significantly predicted the target-type effect on the P600 (*homograph* targets > *cognate* targets) in both groups of bilinguals was their accuracy on the experimental "false-friend" items in a verbal (timed) translation task – in other words, a task that was highly relevant to the type of processing that was demanded of them during the ERP experiment. This finding is related to the notion of "*structure-specific proficiency*" (White, Genesee and Steinhauer, 2012; Steinhauer et al., 2009; Tanner et al., 2009; 2013), although our results did not provide overwhelming support for structure-specific proficiency over overall proficiency (see *Section 8.4* on brain vs. behavior below). The use of multiple proficiency measures in investigations of L1 attrition is also a methodological advantage, as some measures may be more sensitive to group differences among speakers with native language proficiency. Our Italian Error-detection task, for example, revealed differences between Attriters and Italian-controls whereas the C-test did not. Our English version of the test was also less prone to ceiling effects in English monolinguals than the C-test (see discussion in **Study 3**).

The next logical step is to employ multiple regression analyses to examine questions pertaining to proficiency effects in L1 attrition and L2 processing, including which measures are more/less sensitive in predicting certain ERP patterns.

8.2. Parallels between L1 attrition and L2 acquisition – two sides of a same continuum?

One of the primary reasons for which we compared L1-attriters and L2-learners in this dissertation was to assess the likelihood of a continuum between L1 and L2 processing, where response profiles in the brain are dependent on proficiency, irrespective of whether the language being processed in one's L1 or L2 (i.e., regardless of AoA). Attrition is a phenomenon whose contribution is key to investigating the idea of such a continuum; not only could we examine proficiency modulations across different language-acquisition profiles, but we could also gain insight as to whether difficulties observed in late L2 learners may mirror those observed in attrition. It has been suggested in recent work on *L2* attrition that stages of attrition may parallel stages of L2 learning but in reverse (Pitkänen, Tanner, McLaughlin & Osterhout, in prep).

Study 2 provided a first step towards a model of a proficiency-based continuum between L1 processing (and attrition) and L2 processing. We showed that the bilingual groups did not differ from one another in the amplitude, scalp distribution and latency of their N400 responses to lexical-semantic anomalies when they were both part of the higher- or lower- proficiency range. A more thorough scrutiny of overlap in proficiency scores between individuals in all three groups is warranted, as is the recruitment of additional participants to fill gaps that currently exist on this potential continuum (e.g., attriters with lower L1-proficiency, late L2 learners with native-like proficiency, bilinguals with dominant (non-attriting) L1 despite high L2-English proficiency).

High-prof. native L1	High-prof L1-dominant bilinguals	Lower-prof native-L1 bilinguals	Lower-prof L1-dominant	Higher-prof L1-attriters	Native-like L2 learners	Lower-prof L1-attriters	High-prof L2 learners	Lower-prof L2 learners

It is of value to compare the processing of the *same* language (L1 for natives, L2 for late learners) across these different AoA and proficiency profiles, as we did with our three groups in **Study 2**. Moreover, examining both the L1 *and* L2 of the bilingual individuals on the spectrum, as well as conducting longitudinal studies of the same individuals, would be ideal to not only capture proficiency changes over time, but especially to understand the impact of shifts in exposure and dominance, and the special role these circumstances might play in attrition compared to more prototypical situations of L2 learning. For example, it may be that the degree of cross-linguistic co-activation from L2 to L1 is not contingent on high-L2 proficiency per se, but on a shift towards greater exposure to the L2 than the L1. Although our investigation was not longitudinal, our studies did investigate both languages within the same bilingual individuals and were suggestive overall of a possible continuum in attriters' L1 vs. L2 processing patterns, with more native-like L2 and less native-like L1 processing profiles with increasing length of residence, decreasing L1 exposure and increasing L2 proficiency. More specifically, prolonged periods of residence in the L2-dominant environment led to less native-like N400 responses (smaller amplitudes than native-speakers) in response to *Mismatch* lexical-semantic anomalies (**Study 2**), as well as to a lesser degree of L1-to-L2 transfer during L2 processing (**Study 3**). The same pattern held between ERP profiles and amount of L1-exposure: lower amounts of L1 exposure were linked to more native-like N400 responses to English (L2) lexical-semantic violations and less crosslinguistic transfer from the L1 to the L2 (**Study 3**). In addition, **Study 1** showed that lower L1 exposure was associated with smaller amplitudes in the late-P600 phase (repair/re-analysis) in response to number agreement errors on the modifier relative to correct sentences. Lower L1 exposure also predicted smaller N400 amplitudes to *Mismatch* vs. *Correct* conditions in **Study 2.** An open question that we are currently addressing is whether such a proficiency-based continuum of responses regardless of L1/L2 status may hold for complex areas of morphosyntactic processing as well, which have been deemed to show evidence of difficulty and "non-native-like" processing mechanisms even in high-proficiency late-learners (e.g., Clahsen & Felser, 2006).

In sum, the possibility of a proficiency-based continuum spanning across L1 and L2 processing and encompassing attrition is likely to be a promising avenue to understand the neurocognitive correlates of language-processing and to determine whether continued neuroplasticity for language in adulthood allows brain mechanisms underlying L1 and L2 processing to change with increasing or decreasing proficiency, exposure and/or dominance.

8.3. What is "attrition" and is there evidence of attrition in L1 comprehension at the neurocognitive level?

Our findings from Study 1 and 2 were taken as evidence for attrition in L1 comprehension. In **Study 1**, we argued that Attriters showed crosslinguistic influence from English in their processing of subject-verb agreement mismatches, and also differed in their morphosyntactic re-analysis mechanisms as reflected by differences in the late P600 (and behavioral response times) from non-attriting native-monolinguals. In **Study 2**, we claimed that lower-proficiency attriters differed from native-controls in their diagnosis of *Swap* violations, showing the low-proficiency L2 learner pattern in the N400 time-window (*Swap* likely to be processed as *Correct*), but engaging in a controlled, conscious "second-thought" process further downstream (larger early P600s). In sum, we found significant group differences in amplitude, scalp distribution and/or duration of several ERP effects (N400 in morphosyntax and lexical-semantics; P600s at different temporal stages), and interpreted these findings as some of the first evidence of attrition at the neurophysiological level during real-time L1 comprehension.

Can these effects be accounted for differently than being attributed to attrition? First, one might observe that, as a group, attriters were more advanced in age than native-Controls (*M* Attriters = 36; *M* Controls = 31). However, there were no significant correlations between ERP patterns elicited in any of the experiments and age-at-testing. A second argument might be that differences in working memory (WM) and/or reading speed may account for group differences. However, the groups did not differ significantly on either of these measures (see **Study 1**). The groups may also have been differentially sensitive or experienced with the rapid-serial-visual-presentation (RSVP) of sentences on the computer screen. However, we had constructed the WM task to mirror the ERP experiment and, thus, letter and number sequences were presented one character at a time at the center of the screen. Given that the two groups did not differ in their WM scores, it seems unlikely that there should be an underlying difference in their abilities (or, at the very least, comfort level) in processing input in an RSVP paradigm.

Another viable interpretation, however, is that attentional differences contribute to group differences in ERP response patterns. Given that we explicitly set out to recruit individuals who were reporting changes/difficulties in their L1, participants in our attrition group were aware that they were being selected for a study assessing their native-language. We chose not to hide this information from them in order for the individuals we recruited to be more likely candidates of attrition. However, this also inherently meant that attention, focus and motivation to show maintenance of the L1 may have played a greater role in these individuals than in native-speakers who were tested in Italy. Attriters are more likely to be self-conscious and to want to perform well in their native-language. That said, I believe that the potentially heightened attention in these individuals is an intrinsic characteristic of attrition, and an influencing factor of its own right. We will return to this point in the context of the P600 below.

Is it fair to label these differences as "attrition"? Some may argue that differences in amplitude, distribution or duration of otherwise "similar" ERP components may not constitute compelling evidence in favor of attrition. Such a criticism raises important issues relevant to the operationalization of "attrition" as well as for ERP research in general. Modulations in amplitude, scalp distribution, latency and duration have long been interpreted in L2-processing studies as differences in processing mechanisms between (late) L2 learners and native-speakers, (e.g., Hahne & Friederici, 2001; Osterhout et al., 2006; Rossi et al., 2006; Weber-Fox & Neville, 1996). Such differences (e.g., modulations in N400/LAN amplitude or distribution, smaller/absent P600s, etc.) have been explained as non-native processing due to transfer effects from the entrenched L1 and/or reliance on differential processing mechanisms for the late L2 compared to the L1. Would it not be a double-standard to interpret modulations of similar kinds as merely subtle differences in otherwise "native-like" processing patterns in Attriters? Such an argument also potentially undermines the functional significance of ERP components – for example, the view that a frontal positivity is P600 with a slightly different distribution would be overly simplistic and actually inaccurate.

If it *is* attrition, then why were participants' anecdotal reports of changes or difficulties not reflected in their L1 proficiency scores which, though numerically lower, were not statistically different from those of native-controls in Italy? It may be argued that these individuals are not “attriters”, or at least may be in the earlier (rather than advanced) stages of attrition. Potential reasons for the dissociation between anecdotal reports and actual proficiency scores is addressed in more detail in the *brain vs. behavior* section below. However, after having conducted the studies in this dissertation, it is not clear to me whether low-proficiency scores on proficiency tests constitute a necessary prerequisite for attrition. First, attrition effects could potentially be found in some areas and tasks but not others (see p. 12 of Schmid, 2011), though it is an open empirical question whether our current group experienced minimal attrition and that is why effects were not detectable “across the board”. However, an important point to raise is that this view of attrition and the evidence deemed necessary to support it are highly circular – researchers ask whether a group with "more advanced attrition" would show lower scores on L1 behavioral proficiency tasks than individuals with "less advanced attrition", yet the very means with which the degree of attrition is quantified are often those same measures of proficiency and behavior.

I must admit that, when this series of studies was first designed, I also subscribed to this view that attriters would (and must) be considerably lower in their L1 proficiency profiles than Italian native-speakers still living in Italy. Such individuals surely do exist; however, the argument here is that proficiency and behavioral differences need not be the sole defining factor of attrition. The studies in this dissertation have highlighted that attrition effects are not always entirely describable as profiles of lower-proficiency. In **Study 1** and **3**, we showed that, in addition (and independent of) proficiency effects, several *group* differences were found between Attriters and Controls. Recall that, in **Study 1**, these group differences involved a more robust N400-like negativity in response to subject-verb number agreement violations, and shorter P600 effects when processing agreement violations both on the verb and the modifier. In **Study 3**, group differences were found on the amplitude and distribution of the P600 in response to *Swap* and *Mismatch* lexical-semantic anomalies, with larger and more broadly (less posterior) P600-effects in Attriters than in Controls. Given that proficiency and group effects were additive in our studies, it seems necessary to adjust our operational definition of attrition in order to encompass effects that go beyond proficiency effects per se.

Our studies have permitted us to describe and quantify L1 attrition in terms of three main parameters: (1) *proficiency effects,* where L1 processing patterns are more native-like at higher ranges of L1 proficiency (**Study 1** and **2**), and L2 processing is more native-like at higher levels of L2 proficiency (**Study 3**); (2) *crosslinguistic transfer effects,* where increased attrition effects are characterized by a decrease in L1-to-L2 transfer (**Study 3**) and an increase in L2-to-L1 transfer (**Study 1** and **relative clauses**, see Section 1.9.2 of *General Introduction*); and (3) *special experiential circumstances*, where increased attention, motivation to perform well, amount of L1 vs. L2 exposure (dominance, see **Study 3**), number of years spent in the L2-environment, and other factors that are characteristic of this group but neither to L1-dominant controls nor L2-learners.

The question as to what extent attrition may be different from "normal" proficiency variation within native-speakers and from "normal" bilingualism effects – and how this variability may fit the continuum discussed earlier – is an important avenue for research to pursue, with the use of appropriate comparison groups (see Section 8.7 on *Future directions*).

8.4. Attrition effects observed in the brain vs. behavior

Another research question explored in this dissertation was whether attrition effects would be observed both at the level of behavior (proficiency tasks, acceptability judgment task) and at the level of the brain (ERP correlates of online language-comprehension). It may perhaps be perceived as a counterintuitive finding that the differences we observed at the level of ERPs were not directly mirrored in participants' (especially attriters') behavioral performance.

First, it was already noted above that attriters did not perform significantly worse than native-controls on the behavioral measures of proficiency that we administered. One possibility is that, at least for the written measures, Attriters arrived at the same responses but after spending more time on the task. A potential shortcoming of this work is that we did not time participants while they completed the offline (pen-paper) proficiency tasks. The rationale behind this decision was that, contrary to the timed measures of production (e.g., verbal semantic fluency and false-friends translation task) as well as the acceptability judgment task performed during ERP recording (where participants were encouraged to respond as quickly as possible), we wished to administer several untimed measures in order to have an idea of the "upper limit" of individuals' language abilities, particularly for the group of Italian L2 learners. Though this choice still seems valid, in hindsight, we may have wanted to at least record the amount of time spent on the task, such that we could correlate participants' scores with the amount of time they invested in performing the written task. Attriters (and L2 learners) did seem to take longer on these written proficiency measures than native-speaker controls in Italy. These slower response times would be in line with the longer reaction times we observed during the acceptability judgment task overall. As mentioned earlier, it is likely that Attriters who were explicitly asked to describe how their L1 fluency may have changed since immigration were more self-conscious, more alert and aware, and motivated to perform well. Although our background questionnaires do contain some questions about motivation to maintain Italian and their comfort-level in using Italian with native-speakers in Italy, it might have been useful to correlate the amount of time spent on the tests with their accuracy and ERP profiles. This is certainly an aspect I would improve if I were to collect these data again.

Next, with respect to the acceptability judgment task performed at the end of each sentence during ERP recording, we found that group or proficiency differences in ERPs were often not reflected in acceptability ratings. For example, although Attriters' ERP responses to number agreement violations in **Study 1** were suggestive of a weaker tendency to engage in online sentence-repair processes compared to native-controls, there were no significant differences between Attriters and Controls in their acceptability judgment ratings, nor were there differences between the ratings for the three kinds of violation conditions. The only significant difference between groups was revealed for response times, where Attriters were found to be slower overall (thus potentially consistent with an account of less efficient processing). Similarly, in **Study 2**, Attriters and native-monolinguals did not differ in their ratings for either of the lexical-semantic violation conditions relative to the correct sentences, although we had expected Attriters (along with L2 learners) to give higher ratings to *Swap* sentences than *Mismatch* sentences. Only in our study of **Italian relative-clauses** did we find acceptability ratings to parallel the pattern we had observed in ERPs, namely that Attriters rated *and* processed the word-orders that were acceptable in Italian but not in English as morphosyntactic violations.

We initially hypothesized that "structure-specific proficiency" (i.e., performance on a task that specifically examines the same structures that we are examining with ERPs) would be the most sensitive predictor of ERP responses, as has been suggested in several studies (White, Genesee and Steinhauer, 2012; Steinhauer, 2014). However, a number of L2 processing studies have instead shown that behavioral responses did not reflect the patterns observed at the ERP level (McLaughlin, Osterhout & Kim, 2004; Morgan-Short et al., 2012; Thierry & Wu, 2007).

What may account for such a discrepancy, if behavior is ultimately a result of brain patterns? Moreover, why might we find a dissociation between ERPs and behavior for some linguistic structures but not others?

A first likely possibility centers on the degree of sensitivity of the acceptability judgment task. Although we predicted that our judgment scale (1-5) would be more suited to reveal subtle differences before groups than binary yes/no decisions, it is conceivable that in some linguistic areas (such as number agreement computation across three positions within a given sentence), the task was unable to reflect differences in specific processes such as online repair. In **Study 1**, our results did not suggest that attriters were unable to *detect* the number agreement errors on the verb or modifier for the violation conditions. Rather, we argued that what differed in Attriters compared to non-attriting native-controls was (a) the robustness of the violation effect on the verb as soon as it disagreed with the subject-noun, and (b) the repair mechanisms reflected by amplitude differences in the late P600 effect, as well as Attriters' tendency to elicit a larger P600 than controls in the "repair condition" (*xyy*). The first differences (on the verb) occurred early on during sentence processing and may have been better reflected in behavioral performance if the task tapped into processes unfolding at that immediate stage prior to integration of subsequent sentence material, rather than at the end of the sentence.

Differences on the modifier, on the other hand, were mirrored only by longer response times in Attriters than native-controls. Response times, interestingly, were longer for conditions where the verb agreed with the preceding subject, indicating that participants were faster to make up their mind about the acceptability of the sentence if the violation occurred early in the sentence (on the verb). In hindsight, a comprehension task that tapped into sentence interpretation rather than participants' ability to categorize a sentence as "good" or "bad" (and sometimes quite early on in the sentence) may have proven more sensitive to processing differences related to input-revision and re-analysis mechanisms. Specifically, comprehension questions assessing readers' interpretation of the number value of the sentences (was the sentence about one worker or two workers?) may more readily detect differences related to sentence-repair.

As a potential interpretation as to why acceptability judgments may be more closely associated with ERP profiles in some linguistic areas (e.g., relative clause word-orders) but not others may be that acceptability decisions – even if on a scale rather than a binary choice – are relatively easy to make when the sentences consist of clear violations. In the case of the Italian relative-clauses, on the other hand, the sentences were ambiguous (unless readers used semantic agent-patient cues to disambiguate them) and potentially awkward in their construction, but they they were all grammatical in Italian. This sort of "grey-zone" may be better reflected by an acceptability judgment task (especially on a scale) than outright violations.

Aside from the potential insensitivity of the end-of-sentence judgment task, another reason for the dissociation between attriters' ERP profiles and behavioral responses may be that attrition effects are detectable in the brain before they emerge in behavior. Although longitudinal studies would be best equipped to study this question, insights from L2 processing studies suggest that a time-lag before behavior shows the patterns predicted their ERP response profiles is indeed a possibility (e.g., McLaughlin et al., 2004). This shift towards attrition effects becoming visible in behavioral tasks may occur first for those areas of language that are more susceptible to attrition effects, such as word-order preferences (rather than ungrammaticality), or areas of cross-linguistic competition or transfer.

8.5. The P600 in L1 attrition and L2 acquisition: Implications for ERP research

The P600 was an ERP component of recurring importance across the three studies reported in this dissertation, and one where group and proficiency effects were most pronounced. The following section considers whether we can reconcile the different P600 effects across studies into a uniform account, in a way that can inform attrition and acquisition.

In **Study 1** on Italian agreement processing, we showed that the P600 is not a monolithic component and that its functional significance is distinct from earlier, more frontal positivities that we discussed in the context of a P3a (although they have often been termed "early P600" – Barber & Carreiras, 2005; Friederici, Hahne & Saddy, 2002; Kaan & Swaab, 2003; Molinaro, Vespignani et al., 2008). We showed that the earlier portion of the posterior P600 effect (on the verb and on the modifier) lasting from 650 ms to about 900 or 1000 ms was modulated by L1 proficiency level but not by L1 group (i.e., Attriters vs. Controls), whereas the later portion of the P600 (1000-1200 ms) was determined by group membership, with the P600 virtually absent in this later time-window for the Attriters. We interpreted the first window of the posterior P600 as reflective of the diagnosis of a violation (see Fodor & Inoue, 1998, discussed in Friederici et al., 2001 for garden-path sentences), whereas we associated the later window with processes related to morphosyntactic repair (Carreiras, Salillas & Barber, 2004; Hagoort & Brown, 2000; Mancini, Vespignani, Molinaro, Laudanna & Rizzi, 2009; Molinaro, Vespignani et al., 2008; Silva-Pereyra & Carreiras, 2007).

In **Study 2** on lexical-semantic processing of confusable and non-confusable words in Italian, Attriters were shown to elicit the largest P600 effects in comparison to native-Italian monolinguals and English-Italian L2 learners. This P600 effect was also larger for *Swap* vs. *Correct* sentences than for *Mismatch* vs. *Correct* sentences, an effect that we interpreted as reflecting enhanced conflict-monitoring when double-checking the preceding input to determine the source of the error (i.e., "It should have been X"; van de Meerendonk, Kolk et al., 2010; Vissers et al., 2006).

Finally, in **Study 3** on interlingual homographs and cognates embedded into English sentences, we showed that Attriters were similar to English-Italian L2 learners and English native-speakers in their P600 responses to English violations (IC, EC, CH) between 650-850 ms. However, in addition to these violation effects, the two bilingual groups showed an enhanced and longer-lasting (until 1000 ms) P600 effect for homograph target words relative to cognate target words. This effect was not found in English native-monolinguals. We interpreted this target-effect in the bilingual groups as suggestive of a more controlled inhibitory process to resolve the competition between the two identical homograph readings. In the context of the conflict-monitoring view (Vissers, Kolk, Van de Meerendonk and Chwilla, 2008; Kolk and Chwilla, 2007; Vissers, Chwilla & Kolk, 2006; 2007, the recognition of interlingual homographs created heightened conflict due to their non-overlapping semantic meanings and the fact that the inappropriate Italian homograph meaning had to be inhibited in the context of English sentences (contrary to cognates which converge on both form and meaning).

Taken together, results from the three studies suggest that the posterior P600 lasting until about 1000 ms is linked to *diagnosis* of an error, whether this error is morphosyntactic or lexical-semantic in nature. This late diagnosis may follow an N400/LAN-like negativity in a biphasic pattern (**Study 1** on verb and on modifier in *xxy*; **Study 2** and **Study 3**) or not (**Study 1** *xyy* and *xyx* violation conditions; **Study 3**: correct English Homograph (EH) sentences). In cases where Attriters differed significantly from native-controls in the amplitude of the P600 they elicited as a *group* (Study 2 and 3), we suggest that Attriters seem to be engaging in a more controlled, explicit diagnosis process, and that this conflict-monitoring process was additionally modulated by proficiency (**Study 1**: larger P600 amplitudes in higher-proficiency; **Study 2**: more posterior distribution of P600 in higher-proficiency; and **Study 3**: larger P600 for homographs vs. cognates in individuals who performed better on the false-friend items of the verbal translation task). Proficiency might explain why English-Italian attriters only show this pattern in *English* in the homograph study, and do not elicit significant P600 effects in neither *Swap* nor *Mismatch* violations in Italian.

However, proficiency alone cannot be the whole story to account for these large P600 effects, as even native controls (in Italian and English) were shown to elicit smaller P600 effects than Attriters (in **Study 2** and **3** respectively). While differences in the earlier time-window of the posterior P600 were not significant between groups in **Study 1** on agreement processing, there were indeed group differences in these effects for **Study 2** and **3** on lexical-semantic processing, suggesting perhaps that the degree of conflict-monitoring and diagnosis may have differed across our experimental paradigms and stimuli. The large P600s we detected in Attriters across all studies may also be connected to the notion of attention raised earlier – it is likely that, having been recruited for a study about their native-language in a non-native environment, Attriters were more alert and metalinguistically aware during the experiment, leading to more controlled "second-thought" processes further downstream from the initial anomaly.

Interestingly, the P600 is the effect that Monika Schmid and collaborators focused on in their large-scale study of gender agreement processing (see Bergmann et al., 2013; Schmid keynote, 2013) and on which they reported no significant amplitude differences between attriters and non-attriting monolingual controls. However, as discussed both in the *General Introduction* and in the introduction to **Study 1**, determiner-noun agreement (without an inflected intervening adjective in Dutch and German) may not be sensitive to group differences in processing at high levels of proficiency among native-speakers. Furthermore, to our knowledge, Schmid and colleagues did not test lexical-semantic processing using ERPs.

Our findings collectively advance our understanding of P600 effects and of the processing routines reflected by different stages of positivities. An unexpected and novel finding was that, in response to number agreement violations, P600 effects in different time-windows were differentially affected by proficiency and group membership, among two groups of *native*-speakers. This differential impact of L1 proficiency and group emphasizes that (1) different stages of the P600 reflect different underlying processes and therefore it would be much too simplistic to consider the P600 as a monolithic component; and (2) attrition cannot be described as merely proficiency variation (i.e., where attriters show "low proficiency" processing profiles). Instead, group differences beyond proficiency differences have indicated that Attriters engage in more controlled, elaborated conflict-monitoring processes that depend both on experimental properties (i.e., stimuli conditions where conflict is enhanced) as well as experiential circumstances that are inherently part of attrition (e.g., increased attention, more cautious approach, second-thoughts). Further studies are needed to continue to investigate the factors that modulate P600 effects in attriters' morphosyntactic and lexical-semantic processing.

8.6. Evaluating the "critical period hypothesis" and related claims of reduced neuroplasticity in adulthood

This dissertation was framed around a theoretical standpoint that posits a neurobiological basis for age-of-acquisition effects in L2 learners (CPH: Lenneberg, 1967; Penfield & Roberts, 1959; Penfield, 1965). While it seems that few, if any, would advocate for the stronger versions of this theory where brain plasticity is thought to abruptly end at a cut-off point in childhood, the notion of a "critical period" and conclusions drawn in support of maturational constraints continue to pervade the literature on the neurocognition of language.

The claims at the core of the CPH and related views that our research specifically attempted to address are: (1) AoA effects are due to maturational constraints on brain plasticity, and non-native-like processing patterns persist at high levels of L2 proficiency; (2) the L1, having had its neuronal connections fixed and "entrenched" within the critical period, is stable in adulthood; and (3) related to the previous point, the more fully-developed the L1-system in the brain, the less those language areas can be modified with late L2 exposure. In other words, attrition effects are unlikely to be found when the change in language-environment occurred post-puberty, and we would expect brain response patterns to be dependent on AoA for L2 learners (and age of immigration for Attriters).

Our studies provided evidence of L1-attrition at the neurocognitive processing level in a group of migrants who had lived in an exclusively monolingual-L1 context until adulthood (age-at-immigration > 28 years). Modulations in the amplitude, scalp distribution and latency/duration of ERP components of interest during online L1 comprehension revealed (1) proficiency effects (that were consistent with those of Italian L1 controls (**Study 1**) or Italian L2 learners (**Study 2)**),

(2) L2-induced changes in L1 processing (negativity on verb in **Study 1**, and **relative clauses**), (3) a decrease in L1 to L2 transfer (**Study 3**) , and (4) other patterns that were characteristic of the Attriters alone (shorter P600 in **Study 1**; largest P600s in **Study 2**) or consistent with those of English-Italian learners (*Homograph-target* P600 effect in **Study 3**)). Evidence of L1-attrition in adulthood and of L2-to-L1 transfer in processing mechanisms are difficult to reconcile with views that the L1 is stable and unlikely to change with L2 experience after childhood. Moreover, age-of-immigration did not predict ERP profiles, and L1-to-L2 transfer decreased with both increasing length-of-residence and decreasing amount of L1 exposure, thus casting doubt on the idea that the L1 is entrenched in adulthood. Although these are some of the first ERP studies on L1 attrition and only an initial step towards clarifying these complex issues, our results on L1 processing seem to argue *against* maturational limits on neuroplasticity for language, at least for the lexical and morphosyntactic phenomena investigated here.

With respect to L2 acquisition, we found no significant correlations between AoA and ERP profiles in our group of English-Italian late L2 learners. Instead, we found that L2 proficiency level was a crucial determinant of processing patterns, both in lexical-semantic processing of confusable words in Italian (**Study 2**) as well as in experiencing L2-to-L1 transfer when reading in English (**Study 3**). Attriters (i.e., late L2 learners of English) also showed native-like N400 profiles on English lexical-semantic violations when their English proficiency level was high (also relative to their Italian proficiency scores). The most crucial finding, however, was late L2 learners were found to be indistinguishable from L1 attriters in their N400 responses in response to Italian lexical-semantic anomalies when both groups belonged to the same proficiency category, and both groups of high-proficiency bilinguals did not differ from native-Italian controls, whereas low-proficiency bilinguals did. The two bilingual groups were also indistinguishable in the P600 effect they elicited in response to lexical-semantic violations containing homographs and cognates, while both groups differed significantly from English-controls. As discussed earlier, these results are in line with the idea of a continuum between L1 and L2 processing (and L1 attrition).

The notion that proficiency-level drives the brain's neurocognitive responses to language regardless of whether the language is the L1 or the L2 (i.e., regardless of AoA) is, in principle, at odds with claims in favor of maturational constraints on native-like language processing. That said, as it has been argued that the lexical-semantic domain may be less affected by a late AoA than morphosyntax, due to greater relative plasticity (Clahsen & Felser, 2006; Newport, Bavelier & Neville, 2001; Weber-Fox & Neville, 1996), it remains to be seen whether a similar continuity between ERP profiles across L1/L2 groups (i.e., irrespective of AoA) would be observed for complex areas of morphosyntactic processing. The latter case would pose a greater challenge for proponents of categorical maturational restrictions on L2 vs. L1 processing.

8.7. Future directions

8.7.1. Limitations

An obvious limitation of this dissertation is that, due to its scope and space- and time-limitations, we were not able to additionally include the data from English-Italian late L2 learners' processing of Italian morphosyntax (Study 1 and Italian relative clauses), nor were we able to include Attriters' data in our experiment of *English* morphosyntactic processing (English relative clauses). As the studies were designed to address, as a whole, the full picture of L1/L2 processing in Attriters and late-learners, in lexical-semantics and morphosyntax, this dissertation must in a way be viewed as a work in progress – a first (big) step in the direction of clarifying questions that are central to our understanding of multilingualism and the brain. The remaining investigations are crucial to fully address our research questions and will be reported in forthcoming manuscripts.

An area of improvement concerns the level of Italian proficiency of our late L2 learners. Although our recruitment process targeted learners in the highly-advanced proficiency-range, our sample of Italian L2 learners scored significantly lower on the proficiency measures we administered. They also stood out in their acceptability ratings of Italian sentences in **Study 2**, rating violation conditions as more highly acceptable than Italian-controls and Attriters. In order to provide even more compelling evidence in favor of a continuum, the aim should be to continue recruiting near-native L2 learners, and to examine individual differences in proficiency to determine which L2 learners may be in the range of higher- and lower-proficiency L1-attriters. Some may argue that the difficulty of recruiting fluent or "native-like" late L2 learners is in and of itself evidence in favor of a critical-period (e.g., Selinker, 1972). However, the challenges we faced in finding highly-advanced late-learners of Italian were largely due to the constraints of our research project. For one, we limited our sample to only L1-English late learners of Italian, in order to allow for the comparison of the same language pair across groups. Moreover, for logistic reasons and time-constraints, we opted to test our Italian-learners in Montreal, rather than returning to Italy and initiating a second collaboration in order to collect data in a more internationally-populated region of Italy where we would find a large number of immersed and fluent late-learners of Italian (such as Rome or Florence, as compared to Rovereto).

Our studies would have also benefitted from the inclusion of a fifth participant group – Italian late-learners of English who, unlike L1-attriters, still live in their native-environment and are therefore dominant in their L1 though highly-proficient in English. Such a bilingual control group would have been crucially informative for several reasons: (1) they can be conceived as the ideal control-group for L1-attriters, as non-attriting Italian-English bilinguals do not experience a shift in exposure and dominance characteristic of Attriters, thus permitting us to directly examine the impact of these factors on L1 and L2 processing; (2) such a group would be the equivalent to our current Italian-controls except that they would be bilingual rather than monolingual, thus giving us the opportunity to determine whether *group* differences we observed in Study 1 were due to attrition or the result of our comparison of bilinguals vs. monolinguals; and finally, (3) in both Attriters and a potential group of non-attriting Italian-English bilinguals, Italian would still be the L1, thus allowing us to address the questions we addressed in Study 1, namely are Attriters different in how they process the *L1*, although AoA is not a factor?

Lastly, another limitation of our work concerns our statistical approach. Given that the vast majority of ERP studies examining the neurocognitive correlates of L2 processing employed ANOVAs to compare groups and often median-split derived subgroups of AoA and/or proficiency (but see Newman et al., 2012; Tanner, Inoue & Osterhout, 2014) for alternative approaches), and as our comprehensive project is among the first to investigate L1 attrition in relation to L2 acquisition, we felt it justifiable to begin with a similar approach. Treating proficiency categorically allowed us to visualize the ERP waveforms of those proficiency subgroups and to conduct analyses that paralleled those visual patterns. However, as was highlighted by Newman and colleagues (2012), a methodological shortcoming of analyses that treat proficiency as a categorical variable is that a median-split can be an arbitrary method of deciding whether an individual should be assigned to a "high" or "low" proficiency group. With this – and our multiple proficiency measures – in mind, we decided to additionally analyze proficiency as continuous scores, as a complement to our traditional ANOVAs. However, adopting a linear mixed-effects (LME) approach would allow for testing of more complex hypotheses than bivariate correlations, and would likely clarify the impact of individual proficiency measures and behavioral judgments on ERP response profiles. While we believe our approach was valid and suitable for addressing our research questions for the first time in a study of such magnitude, we fully realize that the next steps may require more fine-tuned statistical approaches.

8.7.2. New or unanswered questions

A current study's shortcomings are a future study's strengths. The various discoveries and limitations of our current work have permitted us to outline a series of interesting questions that would benefit from future investigations.

A most immediate next step would be to further test the argument that L1 and L2 processing are situated along a same proficiency-based continuum, not only for lexical-semantic processing (as indicated by Study 2) but in areas of morphosyntax as well. Do late L2 learners converge upon native-speakers' neurocognitive processing profiles at very high levels of L2 proficiency? Are there areas of morphosyntax where L1/L2 status affects ERP responses, suggesting that underlying processing mechanisms may differ even in fluent L2 learners, as suggested by Clahsen and Felser (2006)? If so, is predominant L2-use and exposure a modulating factor in determining how native-like L2 learners can become in their online L2 processing, all other things being equal? To test this last point, including a group of late Italian L2-learners fully immersed in Italy (i.e., the reverse profile of our current attrition group) would be highly informative. This reiterates the importance of considering different kinds of "control groups" for different types of research questions, and not simply turning to a group of late L2 learners as the default bilingual control group in future studies of attrition and late L2 processing. It is also vital to continue to examine both the L1 and the L2 of various groups, in order to pinpoint what the mechanisms underlying processing differences might be.

Another important avenue for follow-up work would be to examine the impact of attentional or motivational factors on group differences observed between non-attriting Controls and L1-attriters. For example, we discussed the possibility of our enhanced P600 effects being driven by more conscious "second-thought" processes or increased conflict-monitoring; though "self-consciousness" might be an interesting dimension of advancing attrition, it is an open empirical question to what extent conscious motivation and attentional focus might impact

Attriters' behavioral performance (at least in response times) and ERP profiles. One approach would be to analyze whether there was a relationship between ERP patterns and Attriters' self-ratings on questions that addressed their desire to maintain their native-L1. Another option would be to administer a last questionnaire after the experimental sessions, in order to obtain insights on how Attriters perceived various aspects of our experimental stimuli, and how motivated they were to perform the experimental tasks in each language.

We also discussed the importance of not simply expecting attrition to parallel "low-proficiency" profiles, but rather to adopt a more holistic account that encompasses increases in L2-to-L1 transfer (and decreases in L1-to-L2 transfer) in processing, as well as the impact of factors directly related to Attriters' special environmental circumstances. While Schmid (2011) has argued that crosslinguistic influence may not constitute attrition and may simply be a general "bilingualism effect" and that including a group of late L2 learners of the same language is vital to differentiate between attrition effects and general bilingualism effects, we argue that (1) simply including a group of late L2 learners and examining only their L2 processing is not a fully appropriate solution to this problem (as discussed in detail in Section 1.7.3 of the *General Introduction*); and (2) it is impossible to define "attrition" separately from "bilingualism" as the very core of attrition is an L2-induced change to the L1.

Finally, another methodological recommendation would be to continue to investigate more subtle or complex areas of language processing where we might expect group and/or proficiency differences (e.g., effects of frequency and sentence-constraint on lexical-semantic processing, long-distance dependencies or crosslinguistic differences in morphosyntax), as well as to systematically test how the nature of the task (acceptability judgment vs. comprehension questions vs. priming) may affect the results.

In sum, the present body of work is only the first step on a likely exciting road towards understanding the neurocognitive correlates underlying L1 attrition in relation to both L1 and L2 processing.

8.8. Novelty and implications of this work

The studies in this dissertation are among the first neurophysiological investigations of the phenomenon of L1-attrition. This is also the first work to examine Attriters' processing of both their L1 *and* their L2 in a wide range of lexical-semantic and morphosyntactic properties, to explore attrition effects both *within* and *between* the speakers' two languages, and to systematically assess the impact of proficiency levels in an attempt to describe L1 processing (and attrition) and L2 processing as belonging to a same proficiency-modulated continuum. In order to conduct the most compelling evaluation possible of "critical period" claims that advocate in favor of maturational limits on brain plasticity in adulthood, we opted to test adult L1-attriters (i.e., post-pubescent migrants) and late L2-learners on areas of language that were potentially subtle and more difficult or demanding, in order to maximize the potential for group and/or proficiency differences. The event-related-potential (ERP) studies we conducted all employed meticulously-controlled experimental designs, which contribute to the field of ERP research on L2 acquisition by addressing a number of methodological shortcomings of previous studies. In addition to testing individuals with event-related-potentials (ERPs), we also administered a number of proficiency measures and behavioral tasks. Crucially, these same measures were administered even to native-monolinguals.

Our findings are novel and highly topical, as they provide the first ERP evidence of attrition effects in real-time morphosyntactic and lexical-semantic processing in a population of adult first-generation immigrants, thus challenging the notion that one's first-language is "entrenched" and stable in the brain due to maturational limits on neuroplasticity. Our discussions of attrition highlight a number of important methodological and theoretical arguments that could help pave the path along towards new empirical avenues. Our findings also contribute to second-language processing research, given our exploration of issues such as neuroplasticity, proficiency and exposure, all of which lie at the heart of ongoing debates in the field of neurobilingualism. Thus, first-language attrition can be seen as a unique socio-linguistic phenomenon that bridges the gap between first- and second-language processing research, and sheds new light on an old problem about how language experience can shape the human brain.

9. 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: Language and Cognition, 4*, 179–190.

Ammerlaan, T. (1996). *You Get a Bit Wobbly...: Exploring Bilingual Lexical Retrieval Processes in the Context of First Language Attrition*. PhD dissertation, Nijmegen University.

Andersen, R. W. (1982). Determining the linguistic attributes of language attrition. In Lambert & Freed (Eds.), *The loss of language skills* (pp. 83-118). Rowley, MA: Newbury House Publishers, Inc.

Barber, H.A., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137-153.

Ben Rafael, M (2001). *Contact de langues: le français parlé des francophones israéliens*. Unpublished doctoral dissertation. University of Tel Aviv.

Ben Rafael, M. & Schmid, M. (2007). Language attrition and ideology: two groups of immigrants in Israel. In B. Köpke, M. Schmid, M. Keijzer & S. Dostert, (Eds), *Language Attrition: Theoretical perspectives* (pp. 205-226). Amsterdam: John Benjamins.

Bergmann, Berends, Brouwer, Meulman, Seton, Sprenger, Stowe & Schmid (2013). Processing of gender in L2 acquisition and L1 attrition of German: Evidence from event-related potentials. Talk presented at the Workshop on Neurobilingualism, Groningen, The Netherlands, August 25-27, 2013.

Berman, R. A. (1979). The re-emergence of a bilingual: A case study of Hebrew-English speaking child. *Working papers on bilingualism, 19*, 157-180.

Birdsong, D. (1999). *Second language acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Erlbaum

Birdsong, D. (2006a). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*(1), 9–49.

Birdsong, D. (2006b). Dominance, proficiency, and second language grammatical processing. *Applied Psycholinguistics*, *27*(01), 46–49.

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

Boudreault, P., & Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes, 21*, 608–635.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2014). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*(13), 2492–2511.

Bylund, E. (2008). Maturational constraints and first language attrition. *Language Learning*, *59*(3), 687–715.

Carreiras, M., Salillas, E., & Barber, H. (2004). Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Research*, *20*(1), 98–105.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism, 7*, 71–87.

Chee, M. W. L., Tan, E. W. L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *Journal of Neuroscience, 19*, 3050–3056.

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,* 161–174

Chwilla, D. J., Hagoort, P., & Brown, C. M. (1998). The Mechanism Underlying Backward Priming in a Lexical Decision Task: Spreading Activation versus Semantic Matching. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 51*(3), 531-560.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(01), 3–42.

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

Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: an introduction. In M. D. Rugg & M. G.H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 1-26). New York: Oxford University Press.

Connolly, J. F., Phillips, N. A., Stewart, S. H., Brake, W. G. (1998). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43*(1), 1-18.

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

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology, 39*(2), 131-143.

Curtiss, S. (1977). *Genie: A psycholinguistic study of a modern-day "wild child".* New York: Academic Press.

Datta, H. (2010). *Brain bases for first language lexical attrition in Bengali-English speakers*. City University of New York.

Davis, K. (1947). Final note on a case of extreme social isolation. *American Journal of Sociology, 52*, 432-437.

De Bot, K. (1996). Language loss. In H. Goebl, P. Nelde, Z. Stary, & W. Wölk (Eds.), *Contact linguistics: An international handbook of contemporary research*, Vol. 1 (pp. 579 – 585). Berlin: Walter de Gruyter.

De Bot, K. (2001). Language use as an interface between sociolinguistic and psycholinguistic processes in language attrition and language shift. In J. Klatter-Folmer, & P. van Avermaet (Eds.), *Theories on maintenance and loss of minority languages. Towards an more integrated explanatory framework* (pp. 65–82). Münster: Waxmann.

De Bot, K., & Clyne, M. (1994). A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development, 15*(1), 17– 28.

De Bot, K., Gommans, P., & Rossing, C. (1991). L1 loss in an L2 environment: Dutch immigrants in France. In H. W. Seliger, & R. M. Vago (Eds.), *First language attrition*, (pp. 87 – 98). Cambridge: Cambridge University Press.

De Keyser, R. M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 313-348). Oxford, UK: Blackwell.

De Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(01), 33. De Long et al., 2011;

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations: Dual ERPs to low probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207.

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the 'Late Positive Complex': Mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology*, 41(5), 665–678.

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

Donchin, E., Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*, 357–374.

Doughty, C. J., & Long, M. H. (2003). *The handbook of second language acquisition*. Oxford: Blackwell

Ecke, P. (2004). Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism*, *8*(3), 321–354.

Eubank, L., & Gregg, K. R. (1999). Critical periods and (second) language acquisition: Divide et impera. In Birdsong, D. (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 65–99). Hillsdale, NJ: Erlbaum.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47–65.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language, *The Journal of the Acoustical Society of America, 89*, 395-411.

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

Friederici, A.D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31*(1), 45-63.

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*(1), 529–534.

Gardner, **R. C. (1985).** Social Psychology and Second Language Learning: The Role of Attitudes and Motivation. London: Edward Arnold.

Genesee, F., Paradis, J., & Crago, M. B. (2003). *Dual language development and disorder: A handbook on bilingualism and second language learning*. Baltimore: Brookes.

Goad, H., & White, L. (2008). Prosodic structure and the representation of L2 functional morphology: A nativist approach. *Lingua, 118*(4), 577-594.

Green, D. W. (1986). Control, activation and resource: a framework and a model for the control of speech in bilinguals. *Brain and Language, 27*, 210–223.

Grosjean, F., & Py, B. (1991). La restructuration d'une première langue: l'intégration de variants de contact dans la competence de migrants bilingues. *La Linguistique, 27*, 35– 60.

Grotjahn, R. (1987). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In Grotjahn, R., Klein-Braley, C. and Stevenson, D.K. (Eds.), *Taking their measure: The validity and validation of language tests* (pp. 219-253). Bochum: Brockmeyer.

Gürel, A. (2002). *Linguistic characteristics of second language acquisition and first language attrition: Turkish overt versus null pronouns*. PhD dissertation, McGill University, Montreal, Canada.

Gürel, A. (2004). Attrition in L1 competence: The case of Turkish. In M. S. Schmid, B. Kopke, M. Keijzer and L. Weilemar, (Eds.), *First language attrition. Interdisciplinary perspectives on methodological issues* (pp. 225-242). Amsterdam: John Benjamins.

Gürel, A. (2004). Selectivity in L2-induced L1 attrition: a psycholinguistic account. *Journal of Neurolinguistics*, *17*(1), 53–78.

Gürel, A. (2007). (Psycho)linguistic determinants of L1 attrition. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dosterst (Eds.), *Language Attrition. Theoretical Perspectives* (pp. 99–120). Amsterdam: John Benjamins.

Gürel, A. & G. Yılmaz (2011). Restructuring in the L1 Turkish grammar. Effects of L2 English and L2 Dutch. *Language, Interaction and Acquisition, 2*(2), 221-250.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*(11),1531-1549.

Hagoort, P., Wassenaar, M., & Brown, C. A. (2003). Syntax-related ERP effects in Dutch. *Cognitive Brain Research, 16*(1), 38-50.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266.

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., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*(02), 123–141.

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

Hansen, L. (2001). Language attrition: The fate of the start. *Annual Review of Applied Linguistics, 21*, 60-73.

Harley, B., & Wang, W. (1997). The critical period hypothesis: Where are we now? In A. de Groot and J. Kroll, (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp 19-51). Mahwah, NJ: Erlbaum.

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

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30*(1), 47-61.

Hulsen, M (2000). *Language loss and language processing. Three generations of Dutch migrants in New Zealand*. Unpublished doctoral dissertation. Nijmegen: Katholieke Universiteit.

Hyltenstam, K., & Abrahamsson, N. (2000). Who can become native-like in a second language? All, some, or none? On the maturational constraints controversy in second language acquisition. *Studia Linguistica, 54*, 150– 166.

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.

Jarvis, S. H. (2003). Probing the effects of the L2 on the L1: A case study. In V. Cook (Ed.), *The Effects of the Second Language on the First* (pp. 81-120). Clevedon: Multilingual Matters.

Jaspaert, K. & Kroon, S. (1989). Social determinants of language loss. *Review of Applied Linguistics (I.T.L.), 83,* 75-98.

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.

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: An ERP study. *Journal of Psycholinguistic Research*, *31*(2), 165–193

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98-110.

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. Poster presentation at the Donostia Workshop on Neurobilingualism, Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain, Sept 30 – Oct 2, 2010.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013a). When the second language takes over: ERP evidence of L1-attrition in morphosyntactic processing. Talk presented at the International Conference on Multilingualism, Montreal, Canada, October 24-25, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2013b). My Italian is not what it used to be: Investigating the neural correlates of L1 attrition and late L2 acquisition. Poster presented at the Workshop on Neurobilingualism, Groningen, The Netherlands, August 25-27, 2013.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014a). The case of the non-native-like first language: ERP evidence of L1-attrition in lexical and morphosyntactic processing. Symposium presentation, The 17th World Congress on Psychophysiology, Hiroshima, Japan, September 23-27, 2014.

Kasparian, K., Vespignani, F., & Steinhauer, K. (2014b). Neurophysiological correlates of L1 attrition and L2 acquisition: A continuum based on proficiency. Poster presented at the Society for the Neurobiology of Language, Amsterdam, Netherlands, August 27-29, 2014.

Kasparian, K., & Steinhauer, K. (2015). First-language (L1) attrition in adulthood: New insights on language experience and neurplasticity. Poster presented at the Bilingual Brain Initiative Symposium, Montreal Neurological Institute, May 28-29, 2015.

Keijzer, M. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada.* Unpublished doctoral dissertation, Vrije Universiteit Amsterdam.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205-225.

Kim, K. H., Relkin, N. R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*(6638), 171–174.

Kolk, H. H. J., & Chwilla, D. J. (2007). Late positivities in unusual situations. *Brain and Language, 100*, 257–261.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1-36.

Köpke, B. (1999). *L'attrition de la première langue chez le bilingue tardif : Implications pour l'étude psycholinguistique du bilinguisme*. Unpublished doctoral dissertation. Toulouse: Université de Toulouse-Le Mirail.

Köpke, B. (2000). Effet du pays d'accueil sur le maintien de la langue. Le cas des immigrés d'origine allemande. *Éducation et Sociétés Plurilingues/Educazione e Societa' Plurilingue, 9*, 59-65.

Köpke, B. (2002). Activation thresholds and non-pathological L1 attrition". In *Advances in the Neurolinguistics of Bilingualism. Essays in Honor Of Michel Paradis*, F. Fabbro (ed), 119-142. Undine: Forum.

Köpke, B. (2004). Neurolinguistic aspects of attrition. *Journal of Neurolinguistics, 17*, 3-30.

Köpke, B. (2007). Language attrition at the crossroads of brain, mind and society. In B. Köpke, M. S. Schmid, M. Keijzer & S. Dosterst (Eds.), *Language Attrition: Theoretical Perspectives* (pp. 9-37). Amsterdam: John Benjamins.

Köpke, B., & Schmid, M. S. (2004). Language Attrition: The Next Phase. In M. S. Schmid, B. Köpke, M. Keijzer & L. Weilemar, L. (Eds.), *First Language Attrition: Interdisciplinary perspectives on methodological issue*s (pp. 1-43). Amsterdam: John Benjamins.

Köpke, B., Schmid, M. S., Keijzer, M. & Dosterst, S. (2007). *Language Attrition: Theoretical Perspectives*. Amsterdam: John Benjamins.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*(2-3), 68–74.

Kras, T. (2008). *L2 acquisition of the lexicon-syntax interface and narrow syntax by child and adult Croatian learners of Italian*. Unpublished doctoral dissertation, University of Cambridge.

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

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

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117-129.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(1), 203-205.

Lambert, W. E. (1974). Culture and language as factors in learning and education. In F. E. Abour and R. D.Meade (Eds.), *Cultural factors in learning and education* (pp.91-122). Bellingham, Washington: 5th Western Washington Symposium on Learning.

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

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

Lück, M., Hahne, A., & Clahsen, H. (2006). Brain potentials to morphologically complex words during listening. *Brain Research, 1077*, 144–152.

Major, R. C. (1992). Losing English as a First Language. *The Modern Language Journal*, *76*(2), 190.

Mancini, S., Vespignani, F., Molinaro, N., Laudanna, A., & Rizzi, L. (2009). Number Agreement Processing with Different Persons: An ERP Study. Poster presented at the 15th AMLAP Conference. Barcelona, Spain, September 7-9, 2009.

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

Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research, 36*, 1258-1270.

Mayberry, R. I., & Eichen, E. (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.

Mayberry, R. I., & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language, 87*, 369-384.

Mayberry, R. I., Lock, E., & Kazmi, H. (2001). Linguistic ability and early language exposure. *Nature, 417*, 38.

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

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition*, *23*(4), 477–494.

Meisel, J. (1997). The acquisition of the syntax of negation in French and German: Contrasting first and second language development. *Second Language Research, 13*, 227–63

Molinaro, N., Kim, A., Vespignani, F., & Job, R. (2008). Anaphoric agreement violation: An ERP analysis of its interpretation. *Cognition*, *106*(2), 963–974.

Molinaro, N., Vespignani, F., Zamparelli, R., & Job, R. (2011). Why brother and sister are not just siblings: Repair processes in agreement computation. *Journal of Memory and Language*, *64*(3), 211–232.

Montrul, S. (2008). *Incomplete acquisition in bilinguals: Re-examining the age factor.* Amsterdam: John Benjamins.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological
exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, *22*(2), 205–220.

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). Empirical considerations on the age factor in L2 phonology. *Issues in Applied Linguistics, 15*, 109–129.

Münte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., Johannes, S. (2001). Differences in brain potentials to open and closed class words: class and frequency effects. *Neuropsychologia, 39*(1), 91-102.

Nakuma, C. (1997). A method of measuring the attrition of communicative competence: A pilot study with Spanish L3 subjects. *Applied Psycholinguistics, 18*, 219 –235

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151-165.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, *24*(5), 1205–1223.

Newcomer, P. L., & Hammill, D. D. (1982). *Test of Language Development-Primary*. Austin, TX: Services for Professional Educators.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Doupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: MIT Press.

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*, 119–131.

Norman, D. A. & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz and D. Shapiro (Eds.), *Consciousness and Self-regulation* (pp. 1-18). New York: Plenum Press.

Ojima, S., Nakamura, N., Matsuba-Kurita, H., Hoshino, T., & Hagiwara, H. (2011). Neural correlates of foreign-language learning in childhood: a 3-year longitudinal ERP study. *Journal of Cognitive Neuroscience, 23*, 183–199.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-1228.

Opitz, C. (2011). *First language attrition and second language acquisition in a second language environment.* PhD dissertation. Centre for Language and Communication Studies. Trinity College Dublin. Dublin.

Orsini, A., Pezzuti, L. (2013). WAIS-IV. Manuale. Giunti OS: Firenze.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494-522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 34*, 785-806.

Osterhout, L., Holcomb, P. J., Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20,* 786–803.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In Carreiras M and Clifton C (Eds), *The On-line Study of Sentence Comprehension: Eyetracking, ERP, and Beyond*. London: Psychology Press.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning*, *56*(s1), 199–230.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*(6), 739-773.

Oyama, S. (1976). A sensitive period for the acquisition of a non-native phonological system. *Psycholinguistic Research, 5*, 261-285.

Pakulak, E. & Neville, H. (2004). Individual differences in online syntactic processing in monolingual adults as reflected by ERPs. In Proceedings of the 17th Annual CUNY Conference on Human Sentence Processing, 149.

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

Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience*, *23*(10), 2752–2765.

Pallier, C., Dehaene, S., Poline, J.-B., LeBihan, D., Argenti, A.-M., Dupoux, E., & Mehler, J. (2003). Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex*, *13*(2), 155–161.

Paradis, M. (1989). Bilingual and polyglot aphasia. In F. Boller, and J. Grafman (Eds.), *Handbook of Neuropsychology* (pp. 117–140). Amsterdam: Elsevier.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In A. De Groot and J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 331-354). Erlbaum, Mahwah, NJ

Paradis, M. (2003). The bilingual Loch Ness Monster raises its non-asymmetric head again- or, why bother with such cumbersome notions as validity and reliability? Comments on Evans et al (2002). *Brain and Language*, 87, 441-448.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Köpke, M. S. Schmid, M. Keijzer & S. Dosterst (Eds.), *Language Attrition: Theoretical Perspectives* (pp. 9-37). Amsterdam: John Benjamins.

Pavlenko, A. (2000). L2 influence on L1 late bilingualism. *Issues in Applied Linguistics, 11*(2), 175–205.

Pavlenko A. (2003). L2. influence and L1 attrition in adult bilingualism. In M. Schmid, B. Kopke, M. Keijzer and L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 47-59). Amsterdam: John Benjamins.

Pelc, L. (2001). *L1 lexical, morphological and morphosyntactic attrition in Greek-English bilinguals*. Unpublished doctoral dissertation, City University of New York, United States.

Penfield, W. (1965). Conditioning the uncommitted cortex for language learning. *Brain*, 88, 787-798.

Penfield, W., & Roberts, L. (1959). *Speech and Brain Mechanisms.* New York: Athenaeum.

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, *15*(2), 202–206.

Perani, D., Paulesu, E., Sebastian-Galles, N., Dupoux., E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain, 121*, 1841-1852.

Pierce, L. J., Klein, D., Chen, J., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 111*(48), 17314–17319.

Pitkänen, I., Tanner, D., McLaughlin, J., & Osterhout, L. (in prep). Use it or lose it: Second language attrition in the brain looks like acquisition in reverse.

Polinsky, M. (1997). *American Russian: Language loss meets language acquisition*.Proceedings of the Annual Workshop on Formal Approaches to Slavic Linguistics (pp. 370–406). Ann Arbor: Michigan Slavic Publications.

Polinsky, M. (2000). A composite linguistic profile of a speaker of Russian in the US. In O. Kagan & B. Rifkin (Eds.), *The learning and teaching of Slavic languages and cultures* (pp. 437-465). Bloomington, IN: Slavica.

Prat, C. S. (2011). The brain basis of individual differences in language comprehension abilities. *Language & Linguistics Compass, 5,* 635–649.

Rossi, S., Gugler, M., Friederici, A., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Cognitive Neuroscience, Journal of*, *18*(12), 2030–2048.

Sabourin, L. (2003). *Grammatical Gender and Second Language Processing* (Groningen Dissertations in Linguistics 42.) Groningen: CLCG.

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

Schaufeli, A. (1996). Word order patterns in contact: Turkish in the Netherlands. *Southwest Journal of Linguistics, 15*(2), 153–169.

Schmid, M. S. (2002). First Language Attrition, Use and Maintenance: The case of German Jews in Anglophone Countries. Amsterdam: John Benjamin

Schmid, M. S. (2007). The role of L1 use for L1 attrition. In B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (Eds), *Language Attrition. Theoretical perspectives* (pp. 135-153). Amsterdam/Philadelphia: John Benjamins.

Schmid, M. S. (2009). On L1 attrition and the linguistic system. *EUROSLA Yearbook*, *9*, 212–244.

Schmid, M. S. (2010). Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingualism: Language and Cognition*, *13*(1), 1.

Schmid, M. S. (2011). The theoretical significance of research on language attrition for understanding bilingualism. Retrieved from http://www.enl.auth.gr/symposium19/19thpapers/005_Schmid.pdf

Schmid, M. S. (2013). First language attrition as a window to constraints on bilingual development. Keynote lecture at the International Symposium on Bilingualism (ISB9), Singapore, June 10, 2013.

Schmid, M. S., & Fägersten, K. B. (2010). Disfluency Markers in L1 Attrition: Disfluency Markers in L1 Attrition. *Language Learning*, *60*(4), 753–791.

Schmid, M. S., & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research 26*(1), 125-160.

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, 1–20.

Schmid, M. S. & Köpke, B. (2009). L1 attrition and the mental lexicon. In Pavlenko, A. (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 209-238). Clevedon: Multilingual Matters.

Schmid, M. S., & Köpke, B. (2011). L'attrition de la première langue en tant que phénomène psycholinguistique. *Language, Interaction and Acquisition*: *Special Issue on L1 attrition*, *2*(2), 197-220

Schmitt, E. (2010). When boundaries are crossed: Evaluating language attrition data from two perspectives. *Bilingualism: Language and Cognition*, *13*(01), 63.

Seliger, H. W. (1989). Deterioration and creativity in childhood bilingualism. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan* (pp. 173–184). Cambridge: Cambridge University Press.

Seliger, H. W., & Vago, R. M. (1991). *First language attrition*. Cambridge: Cambridge University Press

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*, 209-231.

Silva-Pereyra, J. F., & Carreiras M. (2007). An ERP study of agreement features in Spanish. *Brain Research, 1185*(14), 201-211.

Slobin, D. I., Dasinger, L., Kyntay, A., & Toupin, C. (1993). Native language reacquisition in early childhood. In E. V. Clark (Ed.), Proceedings 24th Annual Child Language Research Forum (pp. 179–196). Stanford, CA: Stanford University.

Steinhauer, K. (2014). Event-related Potentials (ERPs) in Second Language Research: A Brief Introduction to the Technique, a Selected Review, and an Invitation to Reconsider Critical Periods in L2. *Applied Linguistics*, *35*(4), 393–417.

Steinhauer, K., & Connolly, J. F. (2008). Event-related potentials in the study of language. *Handbook of the Neuroscience of Language*, 91–104.

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., Cornell, S., Genesee, F., & White, L. (2006). The neural dynamics of second language acquisition: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, supplement, 99.

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.

Tanner, D., Osterhout, L., & Herschensohn, J. (2009). Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. In *Proceedings of the 33rd Boston University conference on language development* (pp. 528–539).

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(02), 367–382.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56,* 289-301.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530–12535.

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.

Tsimpli, I. (2007). First language attrition from a minimalist perspective: Interface vulnerability and processing effects. In Kopke, Schmid, Keijzer & Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 86–101). Amsterdam: John Benjamins

Ullman, M. T. (2001a). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726.

Ullman, M. T. (2001b). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, *30*(1), 37–69.

Ullman, M. T. (2001c). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(02), 105–122.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, *92*(1-2), 231–270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In Cristina Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice* (pp.141-178). Georgetown University Press: USA.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., and Vissers, C. T. W. M. (2009). Monitoring in language perception. *Lang. Linguist. Compass* 3, 1211–1224.

van de Meerendonk, N., Kolk, H. H., Vissers, C. T. W., & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82.

van Herten, M., Chwilla, D. J., & Kolk, H. H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18*, 1181–1197.

van Herten, M., Kolk, H. H., Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*, 241-255.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*, 279–293.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190.

Ventureyra, V., & Pallier, C. (2004). In search of the lost language: The case of adopted Koreans in France. In M. S. Schmid, B. Köpke, M. Keijzer, and L. Weilemar (Eds.), *First Language Attrition:Interdisciplinary perspectives on methodological issues* (pp. 207-221). Amsterdam: John Benjamins.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: the effect of misspellings of words in highly constrained sentences. *Brain Research, 1106*, 150–163.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2007). The interplay of heuristics and parsing routines in sentence comprehension: Evidence from ERPs and reaction times. *Biological Psychology, 75*(1), 8-18.

Vissers, C. T. W. M., Kolk, H. H. J., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: Evidence from ERPs in a picture–sentence matching task. *Neuropsychologia, 46*(4), 967-982.

Walsh, T., & Diller, K. (1981). Neurolinguistic considerations on the optimal age for second language learning. In K. Diller (Ed.), *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, *37*(1), 159–170.

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.

Wechsler, D. (1997). *Manual for the Wechsler Intelligence Scale for Children-Fourth Edition (WAIS-IV)*. San Antonio, TX: Psychological Corporation.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12*, 233–365.

White, E. J., Genesee, F., Drury, J. E., & Steinhauer, K. (2007). Before and after: an ERP investigation of late second language learning in an intensive language course. *Journal of Cognitive Neuroscience*, supplement, 290.

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain Responses before and after Intensive Second Language Learning: Proficiency Based Changes and First Language Background Effects in Adult Learners. *PLoS ONE*, *7*(12), e52318.

Yağmur, K. (1997). *First language attrition among Turkish speakers in Sydney*. Tilburg: Tilburg University Press.

10. APPENDICES

Appendix 1a: English C-test (Keijzer, 2007)

Text 1.
We all live with other people's expectations of us. These are a refl__________ of th__________ trying to under__________ us; the__________ are predic__________ of wh__________ we wi__________ think, d__________ and feel. Gene__________, we acc__________ the sta__________ quo, but these expec__________ can be ha__________ to han__________ when they co__________ from our fami__________ and can be diff__________ to ign__________ , especially wh__________ they come from our par__________.

Text 2.
Founded in 1878 by Bishop Isaac Hellmuth and the Anglican Diocese of Huron as "The Western University of London Ontario", Western is one of Canada's oldest and best universities. The fi__________ students grad__________ in ar__________ and medi__________ in 1883. To__________, The University of Western Ontario is a vib__________ centre of lear__________ with 1,164 fac__________ members and alm__________ 29,000 underg__________ and graduate stud__________ _. Through i__________ 12 Facu__________ and Sch__________, and three affi__________ Colleges, the University off__________ more th__________ 60 diffe__________ degree and dip__________ programs to London's comm__________.

Text 3.
The BBC's core purpose is broadcasting. Since the lau__________ of Radio Times in 1923 it h__________ also eng__________ in comme__________ activities. If pur__________ properly, su__________ commercial activities he__________ to rea__________ the va__________ of lic__________ payers' ass__________ and gene__________ income to be plou__________ back in__________ the public ser__________ programming. T__________

commercial Policy Guidelines s__________ out the fram___________ which ens__________ that the BBC's commercial activities supp__________ its public purpose.

Text 4.

The decision to remove soft drinks from elementary and junior high school vending machines is a step in the right direction to help children make better choices when it comes to what they eat and drink. Childhood obe__________ has bec__________ a ser__________ problem in th__________ country a__________ children cons__________ more sugar-based fo__________ and sp__________ less ti___________ getting the nece__________ exercise. Many par___________ have quest__________ schools' deci__________ to al___________ vending machines which disp___________ candy and so__________ drinks. Many schools, th__________, have co___________ to re___________ on the mo__________ these machines generate through agreements with the companies which makes soft drinks and junk food.

Text 5.

In the last federal election, 61% of eligible voters cast a ballot. That's a fright__________ lack of inte__________ by the elect__________, but is not__________ compared to the turn___________ in provi_____________ and munic_________ elections, which s___________ even lo____________ turnouts. It's diff___________ to bel__________ there's so lit__________ interest in elections. In Canada, we're fort___________ to have pol___________ stations wi__________ a short wa___________ or dr__________. There are volun__________ more th___________ willing to pro__________ rides to someone unable to walk or who doesn't have a car.

Answer key:

1. We all live with other people's expectations of us. These are a reflection of them trying to understand us, they are predictions of what we will think, do and feel. Generally we accept the status quo, but these expectations can be hard to handle when they come from our families and can be difficult to ignore, especially when they come from our parents.

2. Founded in 1878 by Bishop Isaac Hellmuth and the Anglican Diocese of Huron as "The Western University of London Ontario", Western is one of Canada's oldest and best universities. The first students graduated in arts and medicine in 1883. Today, The University of Western Ontario is a vibrant center of learning with 1,164 faculty members and almost 29,000 undergraduate and graduate students. Through its 12 Faculties and Schools, and three affiliated Colleges, the University offers more than 60 different degree and diploma programs to London's community.

3. The BBC's core purpose is broadcasting. Since the launch of Radio Times in 1923 it has also engaged in commercial activities. If pursued properly, such commercial activities help to realize the value of license payers' assets and generate income to be ploughed back into the public service programming. The commercial Policy Guidelines set out the framework which ensures that the BBC's commercial activities support its public purpose.

4. The decision to remove soft drinks from elementary and junior high school vending machines is a step in the right direction to help children make better choices when it comes to what they eat and drink. Childhood obesity has become a serious problem in this country as children consume more sugar-based food and spend less time getting the necessary exercise. Many parents have questioned schools' decisions to allow vending machines which dispense candy and soft drinks. Many schools, though, have come to rely on the money these machines generate through agreements with the companies which makes soft drinks and junk food.

5. In the last federal election, 61% of eligible voters cast a ballot. That's a frightening lack of interest by the electorate, but is nothing compared to the turnout in provincial and municipal elections, which show even lower turnouts. It's difficult to believe there's so little interest in elections. In Canada, we're fortunate to have polling stations within a short walk or drive. There are volunteers more than willing to provide rides to someone unable to walk or who doesn't have a car.

Appendix 1b: Italian C-test (Kras, 2008)

Codice partecipante: _________________________ Data: ____________________

ISTRUZIONI

In questi cinque testi mancano parti in alcune parole. Il tuo compito è di completare i testi nell'ordine prestabilito, non dedicando più di cinque minuti ad ogni testo. Non ti preoccupare se non riesci a completare tutte le parole in un testo in cinque minuti. Mentre completi i testi, tieni presente che in alcuni casi ci sono più soluzioni possibili.

Testo 1

La carota era molto invidiosa della cipolla e diceva "per me non piange mai nessuno. Mi tagl_______, mi pel_______, mi frig_______, mi gratt_______, me n_______ fanno d_______ tutti i col_______ e mai u_______ che pia_______ per m_______". Non c_______ le cip_______ abbiano u_______ destino migl_______ della car_______, anche lo_______ vengono affe_______, bollite, arro_______, fritte e soff_______, mangiate cr_______ nell'insalata, ma almeno tutti piangono per loro. Non c'è cuoco o cuoca a cui non vengano gli occhi lucidi mentre si mette a tagliare una cipolla. Che cosa avrà mai la cipolla per essere così compatita?

Testo 2

Ogni volta che lasci in giro un rifiuto, offendi la natura e la vita. I rifiuti s_______ distruggono lentiss_______ (un bara_______ di la_______ impiega 50 an_______, una bott_______ di ve_______ addirittura u_______ milione), e n_______ frattempo soff_______ la nat_______ e rovinano l'amb_______. La lat_______ che og_______ lasci distrat_______ in me_______ al pr_______ sarà anc_______ lì t_______ cento an_______, e avrà tutto il tempo di cedere al terreno l'alluminio di cui è fatta. Il sacchetto di plastica abbandonato durerà 10 o anche 50 anni, o ucciderà una mucca che lo avrà mangiato assieme all'erba del pascolo.

Testo 3

Le zuppe sono un modo semplice ed efficace per incrementare il consumo di verdure, notoriamente salutari. Inoltre, pos_______ aiutare a_______ avere un'alime_______ più bilan_______. Come è eme_______ da u_______ ricerca cond_______ in Francia, i_______ cui s_______ sono esam_______ i dati rela_______ a quasi 5000 adu_______ partecipanti a_______ un am_______ studio epidemi_______ e si è vi_______ che i gra_______ consumatori d_______ zuppe assum_______ meno gra_______ rispetto ai non consumatori e ai consumatori occasionali ed avevano maggiori apporti di carboidrati, folati, beta carotene, vitamina C e, negli uomini, anche di vitamina E.

Testo 4

In Italia ogni anno vengono denunciate 4500 morsicature di cani, solo una piccola percentuale di quelle effettive, e vi sono anche casi di morte della persona aggredita. Il "ris_______ cane" i_______ questi ult_______ periodi sem_______ anzi aume_______. Da u_______ parte s_______ privilegiano ra_______ di tag_______ media e gra_______, il c_______ morso è inevita_______ più dan_______ di que_______ di u_______ cane d_______ piccola tag_______; inoltre c_______ acquista u_______ cane n_______ è sempre adeguatamente informato e capace di tenere a bada l'animale. Il rischio è soprattutto quello del "cane padrone", che non riconosce più nell'uomo il suo "capo" e che, quindi, può aggredire.

Testo 5

Spesso consideriamo le emozioni come debolezze, impulsi da reprimere. Ma sec_______ gli ult_______ studi sar_______ bestie se_______ di lo_______. Gli stud_______ spiegano c_______ sentimenti e_______ emozioni so_______ la ba_______ di que_______ che p_______ millenni g_______ esseri um_______ descrivono co_______ spirito o an_______ dell'uomo. Se_______ emozioni n_______ ci sar_______ l'arte i_______ nessuna forma, non potremmo comunicare con gli altri, neppure apprendere e memorizzare. Certo, è vero che le emozioni in libertà rischiano talvolta di fare danni, quindi, proprio per questo motivo, è importante imparare a riconoscerle e a sintonizzarsi su quelle degli altri per comprenderli.

Answer key:

1. tagliano
2. pelano
3. friggono
4. grattugiano/grattano
5. ne
6. di
7. colori
8. uno
9. pianga (1)/piange (0.5)
10. me
11. che/credo
12. cipolle
13. un
14. migliore
15. carota
16. loro
17. affettate
18. arrostite
19. soffritte
20. crude
21. si
22. lentissimamente/lentissimi
23. barattolo
24. latta/latte
25. anni
26. bottiglia
27. vetro/vernice
28. un
29. nel
30. soffocano
31. natura
32. l'ambiente
33. lattina/latta
34. oggi
35. distrattamente/distratto/distratta
36. mezzo
37. prato
38. ancora
39. tra/tre
40. anni
41. possono
42. ad
43. un'alimentazione
44. bilanciata

45. emerso
46. una
47. condotta
48. in
49. si
50. esaminati
51. relativi
52. adulti
53. ad
54. ampio/ambizioso
55. epidemiologico (1)/epidemico (0.5)
56. visto
57. grandi
58. di
59. assumono/assumevano
60. grassi/grasso
61. rischio
62. in
63. ultimi
64. sembra
65. aumentato/aumentare
66. una
67. si
68. razze
69. taglia
70. grande
71. cui
72. inevitabilmente
73. dannoso
74. quello
75. un
76. di
77. taglia
78. chi
79. un
80. non
81. secondo
82. ultimi
83. saremmo
84. senza
85. loro
86. studiosi/studi
87. che/come
88. ed
89. sono
90. base

91. quello
92. per
93. gli
94. umani
95. come
96. anima
97. senza
98. non
99. sarebbe (1)/sarà (0.5)
100. in

Appendix 2a: English Error-detection task (designed by Kasparian for this dissertation)

Please read the two texts carefully and find all the errors. Please cross out what is incorrect and provide the correction in the space above the line. Errors may occur on several words within the same sentence, so read carefully! Please take no more than 5 minutes per text.

Photography

Photography has been invented officially in the 19th century, although the first camera was described by philosophers and mathematicians in the 5th century BC. Over the past one and a half century, this medium was used to record many aspects of the human life. During this relatively brief story, photography have expanded it's capabilities in recording time and space, thus allowing human vision to being able to capture fleeting moments. The photography has also allow us to visualize both the vast and the minuscule, and has bringed us images from some remote area of the world.

Photography is certainly an art form, and every artists have there own original stile. But one doesn't need necessarily a fancy lens or many money to practice the art; nowadays, with the invention of smart phones and our fascination of social media, many people enjoy very much to snap creative photos on a daily bases, and to share the photos they have took with people they know. Photography has became more popular then it used to be some year ago.

The Tower of Pisa

The Tower of Pisa is a bell tower of the cathedral at Pisa, in Italy. Construction begun in 1178 and the tower leans ever since! The architect that the tower designed did not intended it to lean; it was a flow in the design. First, the foundations was not enough deep. Second, was build on the unstable soil. The tower started sinking by when construction commenced at the third floor. Construction was halted for almost a century, because the Republic of Pisa was engaged with battle. This period allow time for the ground to settle, otherwise the tower could topple over. To compensate for the tilt, the engineers constructed the upper floors with one side taller then the other, and the tower is actually curved. Starting in 1990, many restoration works was made on the tower and the ground around it. In 2008, it has been announced that the tower was perfectly stabilized. Prior the restoration, the tower leaned at an angle of 5.5 degrees, but today it only leans at about 3.99 degrees. This means that the top of it is horizontally displaced with 3.9 meters than it will be if the structure was perfectly vertical.

Answer key:

A) **Photography**

1. was invented
2. officially (moved before invented)
3. centuries
4. has been
5. cross out "the"
6. history
7. has expanded
8. its
9. to be able
10. cross out "the"
11. allowed
12. has brought
13. areas
14. every artist
15. has
16. their
17. style
18. doesn't necessarily need
19. much (or "a lot of")
20. with social media
21. very much enjoy (or "really enjoy")
22. snapping
23. basis
24. have taken
25. has become
26. than
27. some years

B) Tower of Pisa

1. in
2. began
3. has been leaning
4. that designed the tower (or "who")
5. intend (or "had not intended it")
6. flaw

7. foundation (or "foundations were")
8. Deep enough
9. it
10. was built (or "it had been built")
11. cross out "the"
12. the time (or "when" and delete "by")
13. on
14. in
15. allowed
16. could have toppled
17. than
18. much
19. work
20. done
21. was announced
22. prior TO
23. by
24. would be (or "would have been")
25. were

Appendix 2b: Italian Error-detection task (designed by Kasparian for this dissertation)

In questi testi ci sono molti errori. Il tuo compito è di sottolineare e di corregere questi errori (di grammatica o di vocabolario). Per favore non dedicare più di cinque minuti ad ogni testo.

L'utilità dello sport

Fare lo sport ci aiuta a tenere il nostro corpo in forme. Il palestro è il luogo dove ci fa ginnastica, culturismo e aerobica, ma altri sport come il calzino e il tennis si fa nei campi sportivo, all'aperta. Alcune attività sportivi è collettivi, perché i giocatori fanno parte di una squadra, mentre altri sono individuali. In genere i sport individuali come la nuota, lo sci e il ciclismo, se non sono praticato per passione, sono sport di competizione e ci vole molto impegno e molto preparazione per poter partecipare nelle gare. I atleti, per improvare i loro prestazioni, deve fare molto ore di allenamento al giorno.

L'allenamento servono a dare maggiore forza, capacità di concentrazione e resistenza allo sforzo. Gli atleti, inoltre, deve seguire una dieta alimentare appropriato allo sport che praticono, perché non devono augmentare di peso in moda eccessiva. La diete deve essere bilanciato per poter dare all'organismo la giusta quantità di calorie, di vitamine e di sali minerale.

Suonerie e voce alte: Telefonini, multato chi disturba in treno

Basterebbe la buona educazione e quel senso della pudore che ci fareste vergognare se il nostro telefonino si metteva a squillare forte al treno. Se fosse così, averemo già abbassato la suoneria, e se qualcuno si chiamano, risponderemo subito, spostandoci in una posta più tranquillo, o finiriamo in fretta la conversazione. Ma questo non è il gioco del se, e di fatto sui treni italiano impazza le suonerie e le voce alta. Gli italiani non ama il cellulare, ma ci sono dipendente. E la considerano ineliminabile, anche come «salvavita». Adesso però basta tollerare l'incivilità. Il messaggio diffuso dai altoparlanti, di bassare la volume della suoneria e di moderare il tonno della voce, resta inascoltata? Tra poco arriverano le multe. A seconda della gravità del comportamento, il cliente che altri passeggeri disturba dovranno pagare dai 7 ai 23 euro. Non si tratta di una nuove legga, ma di estendere all'uso indiscriminato del telefonino la sanzione già previsto per chi non observa le prescrizione delle ferrovie quando si mettono in viaggio. Per esempio chi disturba o si comporta in moda pericolosa e non rispettava il regolamento.La proposta di legge è stato presentato in Parlamento in gennaio.

Answer key:

L'utilità dello sport

Fare dello sport ci aiuta a tenere il nostro corpo in forma. La palestra è il luogo dove si fa ginnastica, culturismo e aerobica, ma altri sport come il calcio e il tennis si fanno nei campi sportivi, all'aperto. Alcune attività sportive sono collettive, perché i giocatori fanno parte di una squadra, mentre altre sono individuali. In genere gli sport individuali come il nuoto, lo sci e il ciclismo, se non sono praticati per passione, sono sport di competizione e ci vuole molto impegno e molta preparazione per poter partecipare alle gare. Gli atleti, per migliorare le loro prestazioni, devono fare molte ore di allenamento al giorno. L'allenamento serve a dare maggiore forza, capacità di concentrazione e resistenza allo sforzo. Gli atleti, inoltre, devono seguire una dieta alimentare appropriata allo sport che praticano, perché non devono aumentare di peso in modo eccessivo. La dieta deve essere bilanciata per poter dare all'organismo la giusta quantità di calorie, di vitamine e di sali minerali.

1. dello (deleting "lo" also acceptable)
2. forma
3. la palestra
4. si
5. calcio
6. fanno
7. sportivi
8. all'aperto
9. sportive
10. sono
11. colettive
12. altre
13. gli
14. nuota
15. praticati
16. vuole
17. molta
18. alle
19. Gli
20. migliorare
21. le
22. devono
23. molte
24. serve
25. devono
26. appropriata
27. praticano
28. aumentare
29. modo
30. eccessivo
31. dieta

32. bilanciata
33. minerali

Suonerie e voce alte: Telefonini, multato chi disturba in treno

Basterebbe la buona educazione e quel senso del pudore che ci farebbe vergognare se il nostro telefonino si mettesse a squillare forte in treno. Se fosse così, avremmo già abbassato la suoneria, e se qualcuno ci chiamasse, risponderemmo subito, spostandoci in un posto più tranquillo, o finiremmo in fretta la conversazione. Ma questo non è il gioco del se, e di fatto sui treni italiani impazzano le suonerie e le voci alte. Gli italiani non amano il cellulare, ma ne sono dipendenti. E lo considerano ineliminabile, anche come «salvavita». Adesso però basta tollerare l'inciviltà. Il messaggio diffuso dagli altoparlanti, di abbassare il volume della suoneria e di moderare il tono della voce, resta inascoltato? Tra poco arriveranno le multe. A seconda della gravità del comportamento, il cliente che altri passeggeri disturba dovrà pagare dai 7 ai 23 euro. Non si tratta di una nuova legge, ma di estendere all'uso indiscriminato del telefonino la sanzione già prevista per chi non osserva le prescrizioni delle ferrovie quando si mette in viaggio. Per esempio chi disturba o si comporta in modo pericoloso e non rispetta il regolamento.La proposta di legge è stata presentata in Parlamento a gennaio.

1. alta (title)
2. del
3. farebbe
4. mettesse
5. in
6. avremmo
7.ci
8. chiamasse
9. risponderemmo
10. un posto
11. finiremmo
12. italiani
13. impazzano
14. alte
15. amano
16. ne
17. dipendenti
18. lo
19. inciviltà
20. dagli
21. abbassare
22. il
23. tono
24. inascoltato
25. arriveranno
26. dovrà
27. legge

28. prevista
29. osserva
30. prescrizioni
31. mette
32. modo
33. pericoloso
34. rispetta
35. stata
36. presentata
37. a

Appendix 3a: English Reading-fluency test (Woodcock-Johnson et al., 2001)

English Reading Exercise

Examples:

A.	A cow is an animal ……………………………………….….	**T**	**F**
B.	A fish lives on land….………………………………….….	**T**	**F**

Practice:

A.	An apple is blue ....….……………………………………..….	**T**	**F**
B.	The moon is in the sky ....……………………………………..	**T**	**F**
C.	A man has two legs.……………………………………………..	**T**	**F**
D.	Ice is hot..………………………………………………………..	**T**	**F**

Test:

1.	You can eat an apple……………………...………………….	**T**	**F**
2.	A mouse can fly……...…………………………………...…….	**T**	**F**
3.	Dogs have five legs……….....……………………….....…………	**T**	**F**
4.	A hat goes on your head..……………………………...…….	**T**	**F**
5.	A book has pages……………………………………………….	**T**	**F**
6.	A fish has two arms and legs………………………………...	**T**	**F**
7.	The letter B is a number……………………………………...	**T**	**F**
8.	A ring is round……………………………………….............	**T**	**F**
9.	A hen can lay an egg………………………………………....	**T**	**F**
10.	People can see with their eyes……………………...................	**T**	**F**

11. A car flies in the sky…………………….............................. T F

12. Many people like to play games……………….......................... T F

13. There are some days when the sun is green……………………. T F

14. Ants are small……………………………………………….. T F

15. Some farmers grow corn………………………….................... T F

16. A puppy grows into a cat………………………………………. T F

17. A phone book has many numbers…….…………………….. T F

18. The letter C is the last letter of the alphabet………………….... T F

19. The moon is in the sky……………………………………….. T F

20. A spoon can be used for eating……………………………..... T F

21. People may listen to music on a radio…………………………. T F

22. A roof is at the top of a house………………………………….. T F

23. Elephants are small animals……………………………………. T F

24. A jackrabbit has two ears………………………………………. T F

25. A boy may wear a shirt…………………….......................... T F

26. Many plants have green leaves………………………………… T F

27. People like to eat rice with a pen………………………………. T F

28. Games can be played with a deck of cards…………………….. T F

29. June is the month after March…………………………………. T F

30. A key may open the lock on a door……………………………. T F

31. People can light a candle with a match……............................... T F

32. Some bikes have two wheels....……………………………….... T F

33. An airplane has wings……………………………..................... T F

34.	Swimming pools are always filled with balloons………………	**T**	**F**
35.	A cup may be full………………………………………………	**T**	**F**
36.	People place stamps on letters before they mail them………...	**T**	**F**
37.	W is a letter of the alphabet…………………………………….	**T**	**F**
38.	A glass may break if it is dropped on the floor…………………	**T**	**F**
39.	The weather in summer is always snowy……………………...	**T**	**F**
40.	A box may be made of wood…………………………….........	**T**	**F**
41.	A baby may want a bottle………………………………………	**T**	**F**
42.	A child may hide inside a cup……..……………………………	**T**	**F**
43.	A cow makes honey from flowers….......................................	**T**	**F**
44.	It may be hot inside an oven……………………………………	**T**	**F**
45.	April is the first day of the week………………...…………..	**T**	**F**
46.	All spiders have only two legs…………..................................	**T**	**F**
47.	Many cats and dogs wear long pants…………………………...	**T**	**F**
48.	Oranges can be used to make juice for breakfast……………...	**T**	**F**
49.	A picture can be hung on a wall………………………………	**T**	**F**
50.	Most people smile when they are sad………………………….	**T**	**F**
51.	An ocean has plenty of water………………………………...	**T**	**F**
52.	Cattle often go to school in a bus……………………………….	**T**	**F**
53.	Some students write stories when they are in school…………..	**T**	**F**
54.	Many plants grow in gardens…...	**T**	**F**
55.	A sink can hold water……………………...………………...	**T**	**F**
56.	Horses tend to live under water………………………….......	**T**	**F**

57.	All girls have blue eyes and brown hair………………………..	**T**	**F**
58.	Some people wear coats in winter……………………………..	**T**	**F**
59.	A dictionary has many words………………………………….	**T**	**F**
60.	Children are all different ages……………………..................	**T**	**F**
61.	Most people fill their pillows with rocks before sleeping….......	**T**	**F**
62.	Most snakes fly through trees………………………………..	**T**	**F**
63.	A rake is needed to make your bed…………………………..	**T**	**F**
64.	A car is usually much bigger than a bus………………………..	**T**	**F**
65.	Different types of animals may be found at a city zoo………....	**T**	**F**
66.	Some people like to fish on lakes………………………………	**T**	**F**
67.	The letter A is the last letter of the alphabet……………………	**T**	**F**
68.	An alarm clock may wake you up in the morning……………....	**T**	**F**
69.	A bag filled with bricks would be very light…………………...	**T**	**F**
70.	People can earn money by working……..……………………...	**T**	**F**
71.	A broken pen may leak ink……………………………………..	**T**	**F**
72.	A lion usually eats paper when he is hungry…………………...	**T**	**F**
73.	People park their cars on top of their chimneys………………...	**T**	**F**
74.	Flies are bigger than horses…………………………………….	**T**	**F**
75.	Children and adults are all the same height and weight………..	**T**	**F**
76.	Both coffee and tea can be served in the morning......................	**T**	**F**
77.	Many people grow thick leaves on their heads…………………	**T**	**F**
78.	Some families have several children…………………………...	**T**	**F**
79.	A bird growls like a dog………………………………………..	**T**	**F**

80. A piano has keys that are usually painted red and green…......... T F

81. Many people carry money in wallets or purses………………... T F

82. A dentist will help you with problems with your feet................ T F

83. A giraffe has a very short neck……………….……………...... T F

84. A plumber may fix a leak……………………………………… T F

85. Some people like to go skiing on the weekend………………… T F

86. Pilots fly airplanes……………………...................................... T F

87. A carpet belongs on the ceiling……………………................... T F

88. A suitcase can be used to hold an elephant…………………...... T F

89. Candy is always bitter to taste……………………..................... T F

90. A fan may produce a cool breeze……………………………… T F

91. An adult may purchase a home that is for sale……................... T F

92. Dinosaurs may be found roaming in most national parks……... T F

93. People put saddles on cats so they can ride them…................... T F

94. A scientist may work in a laboratory ……................................ T F

95. A child may enjoy an entertaining puppet show……………….. T F

96. You may see an acrobat walk on a tightrope at the circus……... T F

97. Horses often sleep in garages………………………………….. T F

98. Many types of reference books are found in public libraries….. T F

Answer key:

1. T
2. F
3. F
4. T
5. T
6. F
7. F
8. T
9. T
10. T
11. F
12. T
13. F
14. T
15. T
16. F
17. T
18. F
19. T
20. T
21. T
22. T
23. F
24. T
25. T
26. T
27. F
28. T
29. F
30. T
31. T
32. T
33. T
34. F
35. T
36. T
37. T
38. T
39. F

40. T
41. T
42. F
43. F
44. T
45. F
46. F
47. F
48. T
49. T
50. F
51. T
52. F
53. T
54. T
55. T
56. F
57. F
58. T
59. T
60. T
61. F
62. F
63. F
64. F
65. T
66. T
67. F
68. T
69. F
70. T
71. T
72. F
73. F
74. F
75. F
76. T
77. F
78. T
79. F

80. F
81. T
82. F
83. F
84. T
85. T
86. T
87. F
88. F
89. F
90. T
91. T
92. F
93. F
94. T
95. T
96. T
97. F
98. T

Appendix 3b: Italian Reading-fluency test (adapted by Kasparian for this dissertation)

Prova di lettura

Esempi:

A.	Una tigre è un animale………………………………………..….	**V**	**F**
B.	Lo squalo vive sulla terra……………………………………..….	**V**	**F**

Esercizi:

A.	Una pera è blu……….……………………………………….….	**V**	**F**
B.	Un uomo ha due braccia..……………………………………….	**V**	**F**
C.	La luna è nel mare……………………………………………….	**V**	**F**
D.	Il fuoco è freddo…………………………………………………	**V**	**F**

Test:

1.	Si può mangiare una pera……………………...…………….	**V**	**F**
2.	Un toro può volare……...…………………………………...….	**V**	**F**
3.	I gatti hanno cinque zampe………...………………………...…	**V**	**F**
4.	Un berretto si mette in testa..…………………………………...…	**V**	**F**
5.	Un libro ha molte pagine……………………………………….	**V**	**F**
6.	Uno squalo ha due braccia e due gambe………………………..	**V**	**F**
7.	Il numero ‘4’ è una lettera………………………………………	**V**	**F**
8.	Un volante è rotondo…………………………………………...	**V**	**F**
9.	Un gallo può deporre uova…………………………..................	**V**	**F**
10.	Una persona vede con il naso………………………..................	**V**	**F**
11.	Un’aquila vola nel cielo……………………………..................	**V**	**F**

12.	A molte persone piace leggere………………………………….	**V**	**F**
13.	Ci sono alcuni giorni in cui il sole è verde……………………..	**V**	**F**
14.	Le zanzare sono piccole………………………….................	**V**	**F**
15.	Alcuni agricoltori coltivano fragole…………………………….	**V**	**F**
16.	Un cucciolo cresce in un pesce………….……………………	**V**	**F**
17.	Un orologio ha molti numeri………………………………...	**V**	**F**
18.	La lettera 'A' è l'ultima lettera dell' alfabeto…………………..	**V**	**F**
19.	La stella polare è nel cielo……………………………………...	**V**	**F**
20.	Un coltello può essere utilizzato per bere………………………	**V**	**F**
21.	La gente può ascoltare la musica in macchina………………….	**V**	**F**
22.	La cima è la parte superiore di una montagna………………….	**V**	**F**
23.	Gli elefanti sono animali di grandi dimensioni…………………	**V**	**F**
24.	Una scimmia ha due orecchie…………………….................	**V**	**F**
25.	Un ragazzo può indossare una cravatta…………………………	**V**	**F**
26.	Molti alberi hanno foglie verdi…………………………………	**V**	**F**
27.	Il cibo si mangia con una pala………………………………….	**V**	**F**
28.	Maggio è il mese che viene dopo marzo……………………..	**V**	**F**
29.	Una chiave si usa per accendere la luce in una stanza…………	**V**	**F**
30.	Si può spegnere una candela con un accendino…….................	**V**	**F**
31.	Di solito le biciclette hanno due pedali....……………………	**V**	**F**
32.	Una barca ha delle ali………………………………..................	**V**	**F**
33.	Le piscine sono sempre piene di animali……………………….	**V**	**F**
34.	La luna può essere piena………………………………………	**V**	**F**

35.	‘P’ è una lettera dell’alfabeto…………………………………..	**V**	**F**
36.	Una caraffa può rompersi se cade per terra…………………….	**V**	**F**
37.	Il clima in estate è sempre fresco……………………………….	**V**	**F**
38.	Una macchina è fatta sempre di legno…..……………………..	**V**	**F**
39.	Un cane può volere delle coccole……………………………...	**V**	**F**
40.	Un ragazzino può nascondersi sotto il letto…………………….	**V**	**F**
41.	Lo zucchero viene messo nel caffè dopo averlo bevuto……..…	**V**	**F**
42.	Dei trucchi magici possono essere realizzati con un mazzo di carte……………………………………...................................	**V**	**F**
43.	Le api producono il miele volando di fiore in fiore…………….	**V**	**F**
44.	Di solito è caldo all’interno di un frigo…………………...……	**V**	**F**
45.	Gennaio è il primo giorno della settimana…………..................	**V**	**F**
46.	Tutti i ragni hanno solo cinque zampe………………………….	**V**	**F**
47.	Alcuni gatti e cani hanno il pelo lungo…………..……………..	**V**	**F**
48.	Con il pompelmo si può fare la spremuta a colazione…………	**V**	**F**
49.	Un biscotto può essere appeso al muro…………………………	**V**	**F**
50.	Molte persone mangiano quando sono tristi……………………	**V**	**F**
51.	Un deserto è pieno di sabbia……………………………………	**V**	**F**
52.	I bovini vanno al lavoro in autobus…………………………….	**V**	**F**
53.	Alcuni studenti scrivono delle poesie quando sono a scuola…...	**V**	**F**
54.	Negli orti crescono molte verdure……………………...………	**V**	**F**
55.	Una lavatrice contiene acqua…………………………………...	**V**	**F**
56.	I leoni vivono sott’acqua……………………………………….	**V**	**F**

57.	Alcune ragazze hanno gli occhi azzurri e i capelli biondi……..	**V**	**F**
58.	Tutti portano costumi da bagno in autunno…………………….	**V**	**F**
59.	Un'enciclopedia contiene molte parole……………………...	**V**	**F**
60.	I bambini sono di tutte le età……………………………….......	**V**	**F**
61.	Molte persone riempiono i loro bagagli con delle rocce prima di viaggiare…………………………………………………..	**V**	**F**
62.	La maggior parte dei maiali vola tra gli alberi………………….	**V**	**F**
63.	Per preparare il cibo è necessario un pettine……………………	**V**	**F**
64.	Una macchina è di solito molto più grande di un treno………...	**V**	**F**
65.	In uno zoo si possono trovare diversi tipi di dizionari………….	**V**	**F**
66.	Ad alcune persone piace visitare i musei……………………….	**V**	**F**
67.	La lettera 'Z' è l'ultimo numero dell'alfabeto………………….	**V**	**F**
68.	Un uccello può svegliarti al mattino……………………………	**V**	**F**
69.	Un sacchetto rotto potrebbe perdere oggetti……………………	**V**	**F**
70.	La gente può guadagnare soldi lavorando……………………...	**V**	**F**
71.	Un sacchetto pieno di mattoni sarebbe molto pesante………….	**V**	**F**
72.	Le formiche sono più grandi dei gorilla…………….................	**V**	**F**
73.	I bambini sono tutti della stessa altezza e peso…………………	**V**	**F**
74.	Le persone di solito parcheggiano la loro macchina sui tetti…...	**V**	**F**
75.	Il caffè e il tè possono essere serviti nel pomeriggio..................	**V**	**F**
76.	Di solito il serpente mangia la pietra quando ha fame…………	**V**	**F**
77.	Le conchiglie crescono sugli alberi…………………………….	**V**	**F**
78.	Alcuni genitori hanno parecchi figli……………………………	**V**	**F**

79. Una lepre abbaia come un cane……………………………….. **V** **F**

80. Molte persone portano libri o riviste da leggere in viaggio……. **V** **F**

81. Un avvocato risolve problemi con la schiena……….................. **V** **F**

82. La giraffa ha il collo molto lungo……………….…………... **V** **F**

83. Un idraulico può riparare una perdita d'acqua………………… **V** **F**

84. Ad alcune persone piace andare a ballare il fine settimana……. **V** **F**

85. I piloti guidano le barche a vela……………………................. **V** **F**

86. Un tappeto va messo sul divano……………………................. **V** **F**

87. Un'automobile può essere usata per trasportare un elefante…... **V** **F**

88. Una caramella è dolce al gusto………………………………… **V** **F**

89. Un calorifero può produrre una fresca brezza in estate………… **V** **F**

90. Un bambino può acquistare una casa in vendita…….. **V** **F**

91. I gamberetti si possono trovare nella maggior parte dei mercati pubblici……………………………………………………… **V** **F**

92. Le selle si mettono sulle lepri e sui cavalli…………................ **V** **F**

93. Un farmacista può lavorare in un laboratorio……….................. **V** **F**

94. Un adulto potrebbe apprezzare un'interessante rappresentazione teatrale…………………………………… **V** **F**

95. Al ristorante si vede sempre un cameriere ballare sui tavolini………………………………………………………. **V** **F**

96. Di solito i cavalli dormono nella stalla………………………… **V** **F**

97. Nelle biblioteche universitarie si trovano molti tipi di manuali di consultazione…………………………….. **V** **F**

Answer key:

1. V
2. F
3. F
4. V
5. V
6. F
7. F
8. V
9. F
10. F
11. V
12. V
13. F
14. V
15. V
16. F
17. V
18. F
19. V
20. F
21. V
22. V
23. V
24. V
25. V
26. V
27. F
28. F
29. F
30. F
31. V
32. F
33. F
34. V
35. V
36. V
37. F
38. F
39. V

40. V
41. F
42. V
43. V
44. F
45. F
46. F
47. V
48. V
49. F
50. V
51. V
52. F
53. V
54. V
55. V
56. F
57. V
58. F
59. V
60. V
61. F
62. F
63. F
64. F
65. F
66. V
67. F
68. V
69. V
70. V
71. V
72. F
73. F
74. F
75. V
76. F
77. F
78. V
79. F

80. V
81. F
82. V
83. V
84. V
85. F
86. F
87. F
88. V
89. F
90. F
91. V
92. F
93. V
94. V
95. F
96. V
97. V

Appendix 4a: English Letter-number sequencing task scoring (WAIS-IV; Wechsler, 1997)

Practice trials:

i. 6 – F
ii. 4 –G
iii. 3 – 5 – W
iv. 7 – L – T
v. 1 – A – J

Test trials:

1. 2 – L
2. 6 – P
3. 5 – B
4. 7 – F – L
5. 4 – D – R
6. 1 – 8 – H
7. 3 – 9 – A – T
8. 1 – 5 – J – V
9. 4 – 7 – L – N
10. 1 – 6 – 8 – D – G
11. 2 – 7 – C – K – S
12. 3 – 5 – N – P – Y
13. 2 – 4 – 7 – E - M – Q
14. 3 – 5 – 8 – F – H – W
15. 2 - 6 – 9 – A - G – S
16. 1 – 3 – 4 – b – C – R – Z
17. 2 – 5 – 7 – 9- J – T - X
18. 1 – 4 – 8 – E – D - H – R
19. 2 – 5 – 6 – 9 – A – H – N – S
20. 1 – 3 – 4 – 9 – B – D – K – R
21. 1 – 2 – 6 – 7 – F – M – T - Z

Appendix 4b: Italian Letter-number sequencing task scoring (Orsini & Pezzuti, 2013)

Practice trials:

vi. 1 – C
vii. 4 – A
viii. 1 – 2 - B
ix. 5 – A – D
x. 2 – 4 – B

Test trials:

22. 5 – E
23. 3 – A
24. 1 – C
25. 1 – 7 – G
26. 4 – 9 - H
27. 3 – 7 – Q
28. 1 – 5 – L - V
29. 4 – 7 – G – V
30. 6 – 9 – S – T
31. 1 – 6 – 8 – E – F
32. 2 – 4 – C – L - S
33. 3 – 5 – 6 – H – Q
34. 2 – 4 – 7 – M – P – R
35. 2 – 6 – 9 – D – N – S
36. 3 – 5 – 6 – F – H - U
37. 4 – 7 – 8 – F – T – R - V
38. 2 – 3 – 7- 9 – H – N - U
39. 1 – 4 – 8 – D – M – Q – R
40. 2 – 6 – 7 – 9 – A – N – P - S
41. 1 – 3 – 4 – 9 – D – P – R – U
42. 2 – 6 – 7 – 9 – A – F – M - T

In loving memory of Laurie Ann Stowe –
thank you for your teachings and dedication.