The effects of proficiency on sub-lexical processing in bilingual visual word recognition

by

Anna Elisabeth Piasecki

A thesis submitted to Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy

Victoria University of Wellington
2012

Für Mutti
Dla Mamy - To Mum

Abstract

Systematic psycholinguistic research has considered the nature of the coexistence of two (or more) languages in the cognitive system of a fluent bilingual speaker. There is increasing consensus that when a bilingual is presented with a visual stimulus in one language, both of their languages are initially activated (*non-selective access*; e.g. Dijkstra & van Heuven 2002a). However, more recent research shows that certain factors may constrain (or eliminate) the activation of a task-irrelevant language (Duyck, van Assche, Drieghe, & Hartsuiker 2007; Elston-Güttler, Gunter, & Kotz 2005). The objective of the research in this thesis was to investigate how cross-linguistic activation is modulated by specific characteristics of a bilingual's languages. This exploration was mainly limited to an under-investigated area, namely early sub-lexical word processing.

The first of two studies focussed on word processing in the presence or absence of critical sub-lexical information. Specifically, I investigated whether onset capitals – a prominent marker indicating nouns in German – acted as a language-specific cue, and the extent to which this cue constrains competitive, lexical interaction between the bilingual's languages (e.g. *Hose-hose*, the first being a German word meaning 'trousers' in English). This study also considered the extent to which the use of such information is affected by priming for a specific language from a preceding context sentence.

The second study arose from a claim that readers employ distinct sub-lexical reading strategies, depending on the extent of spelling-to-sound (in)consistency in their language (e.g. Ziegler, Perry, Jacobs, & Braun 2001). Employing a bilingual population whose two languages were clearly distinguished in terms of such consistency, I explored the reading strategy used by bilingual participants reading in each language. A key issue is competitive activation between sub-lexical orthographic and phonological representations across languages.

Each study was conducted with two groups of bilingual speakers, English-German and German-English. Individuals varied in their L2 proficiency, allowing a test of whether sub-lexical processing changed as a consequence of increasing proficiency.

The main results from study one demonstrate that bilingual speakers are dependent upon sub-lexical, language-specific information. However, this is influenced by L2 proficiency, with a stronger effect for lower proficiency bilinguals. In addition, lower proficiency bilinguals were more dependent on sub-lexical cues when primed by a

sentence in L2. In contrast, bilingual speakers performing in their L1 used these cues largely under very specific circumstances, i.e. when they did not know an item.

The central finding of study two is that competition between sub-lexical orthographic and phonological representations across languages largely depends on the amount of spelling-to-sound (in)consistency in the bilinguals' *more dominant* language. This is reflected in (1) slower identification of orthographically similar cognates which map onto different phonological representations across two languages, and (2) slower identification of cognates which do not share the same orthographic form across languages but have a common phonological representation. In addition, increasing L2 proficiency is reflected in attenuation of certain effects as processing becomes more automatic, and the development of a common reading strategy accommodating reading in either language.

A major contribution of the research conducted is what findings from both studies reveal about how the bilingual lexicon develops as proficiency increases. Furthermore, the findings contribute to our understanding of the organisation of the bilingual mental lexicon and the processes of word identification, and impose constraints on possible cognitive architectures.

Acknowledgements

Undertaking this PhD was primarily a means of understanding my own language use and behaviour. Even though this journey started off as an individual project, it could not have been completed without the support of various people, who each facilitated my progress in their own different way. This acknowledgment section reflects only a fraction of my huge appreciation of the support I was given throughout this important chapter of my life.

First of all, I am indebted to my primary supervisor, Associate Professor Paul Warren, who was involved in every phase of this thesis. His constant excitement about my project, his mentoring and the speed at which he provided feedback on my writing have been absolutely crucial for the progress of my thesis and a source of inspiration to the way in which I approach conducting research. I will always remember those fearful moments, checking the (usually daunting) number of comments he left on my pieces of writing. In retrospect, however, I see them as having enormously helped me extend my point of view and critical thinking, and shaped my development as an academic.
I am further grateful to my secondary supervisor, Dr Irina Elgort, who agreed to join the project even though it was well underway. Her recent experience of completing a doctoral degree was invaluable for various aspects of my own journey, as she always stayed positive and energetic. Among other things, I aspire to her research and career ambitions as a woman in academia.
Other staff members at the School of Linguistics and Applied Language Studies (SLALS) at Victoria University of Wellington have also been hugely supportive during this stage of my life. This includes staff members who have offered their moral support, such as Dr Meredith Marra and Dr Averil Coxhead; others who have created interesting research opportunities in which I have been involved, including both my supervisors, Dr Stuart Webb, Dr Elaine Vine and the LWP; and finally, all other SLALS members as being part of the incredibly supportive School environment.

Clearly, this project could not have been conducted without the support of various people and institutions immediately prior to and during the data collection stages. In the order of these stages, I would like to acknowledge the support of Natalia Peters and Tanja Schubert-MacArthur, the staff members at VUW's German department (particularly that of Anke Szczepanski), the German Club members and the Goethe Institute in Wellington. In Germany, I am indebted to the support of my previous mentor

and colleague, Professor Dr Brigitte Halford, various other staff at the Albert-Ludwigs University of Freiburg, and my various English teacher friends. I am similarly grateful for the unlimited support provided by Dr Catherine Watson and the German Department members at the University of Auckland. Finally, thank you to all my participants for lending me 'their minds' in the name of science.
Thank you also for the financial support provided by VUW and Education New Zealand, without which I would not have been able to complete my studies.

A more personal thank you goes to my 'families' and friends. To my weekly phone calls and/or Skype sessions with my family in Germany (including Mutti, Oma, Robert, Sabine, Maya and Lana) and the many parcels full of my favourite treats. To my new and ever-changing family in the Salmon Mansion (including Ewa, Niamh, Armando, Alicja, Lorraine, Alejandro, Sam, Brent and Michel), who have let me be the way I needed to be and gave me the necessary strength, but also a reality-check. Special thanks to Ewa, for her creative and caring bulk of energy, and for reviving the Polish part of my soul; and to Niamh, for our lovely morning rides and for creating the many opportunities to play 'touch'. To my 'holiday'-family in Gisborne (including Gary, Lyn, Jannah, Greta and many others), who always make me feel welcome and part of their family. Thank you to all my friends, who have stayed my friends and disregarded the long periods of silence.
A last but most special thank you goes to Kieran, for his constant encouragement, his positive attitude, his humour and patience, and for introducing me to New Zealand sports.

Without all your support I would not be where I am today. Thank you.

Contents

List of Tables..... x

List of Figures..... xii

Chapter 1. INTRODUCTION 1

1.1 Aims of the research 1

1.2 Importance of the research 2

1.3 Background to the research 3

1.4 Organisation of the research 4

Chapter 2. CURRENT THEORIES OF VISUAL LEXICAL RETRIEVAL 7

2.1 Introduction 7

2.2 Part One: Monolingual Research 7

2.2.1 Sub-lexical information processing 8

2.2.2 Lexical level processes 15

2.3 Part Two: Bilingual Research 16

2.3.1 Bilingualism and L2 proficiency level 16

2.3.2 Modelling bilingual visual word recognition 19

Chapter 3. THE ROLE OF SUB-LEXICAL ORTHOGRAPHIC CUES IN BILINGUAL WORD RECOGNITION 29

3.1 Introduction 29

3.1.1 Bilingual lexical organisation – the current standpoint 30

3.1.2 Interlingual homographs 31

3.1.3 Language information 34

3.1.4 Proficiency 45

3.2 Experiment One: L2 Lexical Decision Task (English-German Bilinguals) 49

3.2.1 Methodology 49

3.2.2 Statistical analysis 61

3.2.3 Results and discussion 69

3.2.4 Summary 104

3.3 Experiment Two: L1 Lexical Decision Task (German-English Bilinguals) 105

3.3.1 Methodology 105

3.3.2 Statistical analysis 107

3.3.3 Results and discussion 107

3.3.4 Summary 136

3.4 General Discussion 137

Chapter 4. THE ACTIVATION OF SUB-LEXICAL PHONOLOGY AND ITS INTERACTION WITH/MODULATION OF ORTHOGRAPHIC PROCESSES 153
4.1 Introduction 153
4.1.1 From orthographic representations to phonological retrieval – 'the neglected role of phonology' 153
4.1.2 Proficiency 162
4.2 Experiment Three: Progressive Demasking in English-German Bilinguals 165
4.2.1 Methodology 165
4.2.2 Statistical analysis 176
4.2.3 Results and discussion 177
4.3 Experiment Four: Progressive Demasking in German-English Bilinguals 191
4.3.1 Methodology 191
4.3.2 Statistical analysis 192
4.3.3 Results and discussion 192
4.4 General Discussion 207
Chapter 5. CONCLUSION 215
REFERENCES 221
Appendix A. Models of bilingual word recognition 237
Appendix B. Language proficiency assessment material 239
Appendix C. Experimental material used in Study One 245
Appendix D. Experimental material used in Study Two 259
Appendix E. Ethics Approval 273

List of Tables

Table 1. An overview of the design, number of items and sentence frames used in Experiments One and Two. 49
Table 2. Lexical properties of stimulus material used in Experiments One and Two 52
Table 3. Mean RTs and percent error rates for stimuli in Experiment One 71
Table 4. Mean RTs and percent error rates for stimuli in Experiment Two 109
Table 5. Lexical properties of stimulus material used in Experiments Three and Four. 172
Table 6. Mean RTs and percent error rates for stimuli in Experiment Three. 180
Table 7. Mean RTs and percent error rates for cognates used in Experiment Three. 184
Table 8. Mean RTs and percent error rates for stimuli in Experiment Four. 194
Table 9. Mean RTs and percent error rates for cognates used in Experiment Four. 198

Table 10. Detailed lexical properties of IHGs used in Experiments One and Two..... 245
Table 11. Detailed lexical properties of controls used in Experiments One and Two. ... 246
Table 12. Detailed lexical properties of nonwords used in Experiments One and Two. 247
Table 13. Examples of stimulus material in different experimental conditions..... 249
Table 14. Mean RTs and percent error rates for stimuli in Experiment One..... 257
Table 15. Mean RTs and percent error rates for stimuli in Experiment Two..... 257
Table 16. Orthographic and phonological similarity measures for cognate pairs..... 260
Table 17. Detailed lexical properties of cognates and controls used in Experiments Three and Four..... 261

Regression models

Table E1. Response accuracy to interlingual homographs and controls..... 250
Table E2. Response accuracy to English-German homographs..... 250
Table E3. Response accuracy to German controls..... 251
Table E4. Response accuracy to nonsense items (both English and German). 251
Table E5. Response latencies to interlingual homographs and controls..... 252
Table E6. Response latencies to English-German homographs..... 252
Table E7. Response latencies to German controls..... 253
Table E8. Response latencies to nonsense items (both English and German). 253
Table E9. Response durations to cognates and controls..... 269
Table E10. Response accuracy to cognates and controls..... 269
Table E11. Main response durations to cognates. 269
Table E12. Response accuracy to cognates..... 270
Table G1. Proportion of incorrect responses to IHGs and controls. 254
Table G2. Proportion of incorrect responses to German-English homographs..... 254
Table G3. Proportion of incorrect responses to German control words. 254
Table G4. Proportion of incorrect responses to nonwords..... 254
Table G5. Mean durations of responses to IHGs and German control words..... 255
Table G6. Mean durations of responses to German-English homographs..... 255
Table G7. Mean durations of responses to German control words..... 255
Table G8. Mean durations of responses to nonwords. 256
Table G9. Response durations to cognates and controls..... 271
Table G10. Response accuracy to cognates and controls..... 271
Table G11. Main response durations to cognates..... 272
Table G12. Response accuracy to cognates and controls..... 272

List of Figures

Figure 1. Hypothesised stimulus-response bindings in a lexical decision task, in relation to the language of the sentence preceding a target. 39
Figure 2. Hypothesised stimulus-response bindings in a lexical decision task, in relation to the language of the sentence preceding a target and the presence of a language-specific cue. 44
Figure 3. Schematic of an experimental trial in Experiments One and Two. 56
Figure 4. Proportion of incorrect responses to words by L2 proficiency. 72
Figure 5. Proportion of incorrect responses to IHGs and controls by L2 proficiency. 72
Figure 6. Proportion of incorrect word responses by L2 proficiency and sentence language. 73
Figure 7. Proportion of incorrect word responses, showing an interaction of L2 proficiency, stimulus format and sentence language. 74
Figure 8. Proportion of incorrect word responses by sentence language and stimulus format. 75
Figure 9. Proportion of incorrect word responses frequency. 75
Figure 10. Proportion of incorrect word responses by length. 75
Figure 11. Proportion of incorrect word responses by experimental session. 76
Figure 12. Proportion of incorrect IHG responses by L2 proficiency. 77
Figure 13. Proportion of incorrect IHG responses, showing an interaction of L2 proficiency, stimulus format and sentence language. 77
Figure 14. Proportion of incorrect IHG responses by sentence language. 78
Figure 15. Proportion of incorrect responses to IHGs, in relation to stimulus format and sentence language. 78
Figure 16. Proportion of incorrect IHG responses by frequency. 79
Figure 17. Proportion of incorrect IHG responses by experimental session. 79
Figure 18. Proportion of incorrect IHG responses by syntactic category. 79
Figure 19. Proportion of incorrect IHG responses by L2 proficiency. 80
Figure 20. Proportion of incorrect responses to controls by sentence language and L2 proficiency. 80
Figure 21. Proportion of incorrect responses to controls, showing an interaction of L2 proficiency, stimulus format and sentence language. 81
Figure 22. Proportion of incorrect responses to controls by frequency. 82
Figure 23. Proportion of incorrect responses to controls by length. 82

Figure 24. Proportion of incorrect responses to controls by L1 neighbourhood. 82
Figure 25. Proportion of incorrect nonword rejections by L2 proficiency. 84
Figure 26. Proportion of incorrect nonword rejections by sentence language. 84
Figure 27. Proportion of incorrect nonword rejections by L2 neighbourhood. 85
Figure 28. Proportion of incorrect nonword rejections by L1 neighbourhood. 85
Figure 29. Proportion of incorrect nonword rejections by L2 bigram frequency. 85
Figure 30. Mean word RTs by L2 proficiency and stimulus format. 90
Figure 31. Mean word RTs by L2 proficiency and sentence language. 90
Figure 32. Mean word RTs, showing an interaction of L2 proficiency, stimulus format and sentence language. 90
Figure 33. Mean RTs for IHGs and controls. 91
Figure 34. Mean RTs for IHGs and controls by sentence language. 91
Figure 35. Mean word RTs by frequency. 92
Figure 36. Mean word RTs by length. 92
Figure 37. Mean word RTs by experimental session. 93
Figure 38. Mean RTs for IHGs by L2 proficiency and stimulus format. 94
Figure 39. Mean RTs for IHGs by L2 proficiency and sentence language. 94
Figure 40. Mean RTs for IHGs by sentence language and stimulus format. 94
Figure 41. Mean RTs for controls, showing an interaction of L2 proficiency, stimulus format and sentence language.. 96
Figure 42. Mean RTs for nonwords by L2 proficiency. 97
Figure 43. Mean RTs for nonwords by L2 proficiency and sentence language. 97
Figure 44. Mean RTs for nonwords by sentence language. 97
Figure 45. Mean RTs for nonwords by sentence language and stimulus format. 97
Figure 46. Mean RTs for nonwords by bigram frequency. 98
Figure 47. Mean RTs for nonwords by length. 98
Figure 48. Mean RTs for nonwords by L2 neighbourhood density 99
Figure 49. Mean RTs for nonwords by experimental session. 99
Figure 50. Proportion of incorrect word responses by sentence language. 110
Figure 51. Proportion of incorrect word responses by L2 proficiency. 110
Figure 52. Proportion of incorrect word responses by length. 111
Figure 53. Proportion of incorrect word responses by experimental session. 112
Figure 54. Proportion of incorrect word responses by syntactic category. 112
Figure 55. Proportion of incorrect IHG responses by sentence language 113

Figure 56. Proportion of incorrect IHG responses by L2 proficiency 113
Figure 57. Proportion of incorrect responses to L1 controls by sentence language. 114
Figure 58. Proportion of incorrect responses to L1 controls by L2 proficiency. 114
Figure 59. Proportion of incorrect nonword responses by stimulus format. 116
Figure 60. Proportion of incorrect nonword responses, showing an interaction of L2 proficiency and stimulus format. 116
Figure 61. Proportion of incorrect nonword responses in the first experimental session, by sentence language and L2 proficiency. 117
Figure 62. Proportion of incorrect nonword responses by L1 neighbourhood 118
Figure 63. Proportion of incorrect nonword responses by L2 neighbourhood 118
Figure 64. Proportion of incorrect nonword responses by length 119
Figure 65. Mean RTs for IHGs and controls. 123
Figure 66. Mean RTs for IHGs and controls by sentence language. 124
Figure 67. Mean word RTs by L2 proficiency and sentence language. 124
Figure 68. Mean RTs for IHGs by sentence language. 126
Figure 69. Mean RTs for IHGs, showing an interaction of L2 proficiency and stimulus format. 126
Figure 70. MeanRTs for IHGs by frequency. 126
Figure 71. Mean RTs for IHGs by experimental session. 126
Figure 72. Mean RTs for L1 controls by sentence language. 127
Figure 73. Mean RTs for L1 controls by L2 proficiency and sentence language. 127
Figure 74. Mean RTs for L1 controls by frequency. 128
Figure 75. Mean RTs for L1 controls by syntactic category 128
Figure 76. Mean RTs for nonwords by stimulus format 129
Figure 77. Mean RTs for nonwords by sentence language. 129
Figure 78. Mean RTs for nonwords, showing an interaction of sentence language and stimulus format. 129
Figure 79. Mean RTs for nonwords by length. 130
Figure 80. Mean RTs for nonwords by experimental session. 130
Figure 81. Mean RTs for nonwords by L1 neighbourhood density. 130
Figure 82. Schematic of an experimental trial in Experiments Three and Four 174
Figure 83. Mean RTs for cognates and controls. 181
Figure 84. Mean RTs for cognates and controls by L2 proficiency. 181
Figure 85. Mean RTs for cognates and controls by frequency. 181

Figure 86. Mean word RTs by language 182
Figure 87. Mean word RTs by language and L2 proficiency..... 182
Figure 88. Mean cognate RTs, showing an interaction of orthographic similarity, phonological similarity and language..... 184
Figure 89. Mean RTs for +O cognates by phonological similarity and language. 186
Figure 90. Mean cognate RTs by language and L2 proficiency. 187
Figure 91. Mean RTs for cognates and controls. 195
Figure 92. Mean RTs for cognates and controls by frequency.. 195
Figure 93. Mean word RTs by language 196
Figure 94. Mean word RTs by language and L2 proficiency..... 196
Figure 95. Mean word RTs by frequency. 196
Figure 96. Mean word RTs by length 196
Figure 97. Mean word RTs by syntactic category. 197
Figure 98. Mean word RTs by L1 neighbourhood density. 197
Figure 99. Mean RTs for cognates, showing an interaction of orthographic and phonological similarity. 199
Figure 100. Mean RTs for -O cognates by phonological similarity..... 200
Figure 101. Mean RTs for -O cognates by language and L2 proficiency. 200
Figure 102. Mean RTs for cognates by language and L2 proficiency. 201
Figure 103. Mean RTs for cognates, showing an interaction of orthographic similarity, L2 proficiency and language..... 201
Figure 104. Mean RTs for cognates in the L2 task, showing an interaction of orthographic similarity and frequency. 202
Figure 105. Bilingual Interactive Activation Model 237
Figure 106. The Semantic, Orthographic, Phonological Interactive Activation Model 237
Figure 107 The Inhibitory Control Model. 237
Figure 108. The Bilingual Interactive Activation Plus (BIA+) model 238
Figure 109. Mean RTs for the four O-P cognate conditions (as a means of a visual comparison with Schwartz and colleagues, 2007). 270
Figure 110. Mean RTs for the four O-P cognate conditions (as a means of a visual comparison with Schwartz and colleagues, 2007)..... 272

Abbreviation key

BIA = Bilingual Interactive Activation model (Dijkstra & van Heuven 1998)

BIA+ = Bilingual Interactive Activation model Plus (Dijkstra & van Heuven 2002a)

GS = Graphic Similarity

HPD = Highest Posterior Density

IC = Inhibitory Control model (Green 1998a)

IHG = interlingual homograph (a word that is spelled similarly across two languages, but does not share its meaning)

L1 = first language

L2 = second language

LDT = lexical decision task

lmer = linear mixed-effects (modelling) for R

MCMC = Markov Chain Monte Carlo (sampling)

MROM = Multiple Read-Out Model (Grainger & Jacobs 1996)

NW = nonword

OS = Orthographic Similarity

ODH = Orthographic Depth Hypothesis (Katz & Frost 1992)

PDM = Progressive Demasking task

PS = Phonological Similarity

RHM = Revised Hierarchical Model (Kroll & Stewart 1994)

RT = response time

Chapter 1
INTRODUCTION

1.1 Aims of the research

A considerable amount of psycholinguistic research has been devoted to determining the nature of the coexistence of two (or more) languages in the cognitive system of a fluent bilingual speaker. One aspect of this bilingual system that has been extensively explored concerns the activation status of a bilingual's two languages when s/he is presented with a visual stimulus. Resulting from this research, there is now a growing consensus that the bilingual system is fundamentally non-selective, i.e. that both of a bilingual's two languages are activated in parallel (e.g. Dijkstra & van Heuven 1998, 2002a).

Most evidence in support of this non-selective view has come from studies manipulating some higher level aspects of their stimulus material. These include properties that can be argued to affect the overall organisation of the lexicon such as stimulus frequency or neighbourhood density, and/or task demands. However, considerably less attention has been paid to the processes taking place at the very early stages of bilingual visual processing, namely at the sub-lexical level. Similarly, because the focus of most of these studies has been on bilingual *visual* word recognition, researchers have largely neglected the status of cross-linguistic activation of phonological representations and the role such activation might play in the processing of a visual stimulus. More recent research has shifted the focus from providing further evidence for the non-selective access view to the identification of factors that may constrain (if not eliminate) the activation of one of the bilinguals' two languages. To this point, it has been suggested that sentence context may to some extent limit the activation of a language that is not relevant to the task at hand (Duyck et al. 2007; Schwartz & Kroll 2006). Finally, while acknowledging the importance of research testing the impact of varying proficiency on the established processes underlying bilingual visual word recognition, most research to date has typically only included highly fluent bilingual speakers.

In order to address these aspects of bilingual visual word recognition that have received little attention to date, this thesis singles out four major issues for exploration. The first two issues focus on the *orthographic* aspect of sub-lexical processing, and the last two shift their primary focus to the *phonological* aspect of sub-lexical bilingual processing. Accordingly, the research first establishes how activation of the bilingual's two

languages can be modified in the presence or absence of critical language-specific information. Secondly, the research investigates to what extent additional language information (in the form of a task-(ir)relevant sentence language preceding a stimulus) can enhance or reduce the impact of this language-specific cue on bilingual lexical activation. The third aim of this research concerns the amount of sub-lexical *phonological* activation that occurs across languages when bilinguals read silently, and how this activation spreads back to the sub-lexical orthographic level. Finally, the research aims to clarify whether the interactive activation dynamics between the sub-lexical orthographic and phonological levels depend on the amount of print-to-sound transparency of a bilingual's two languages. The common thread in the studies is the effect of increasing proficiency on the processes involved during these earliest stages of visual word recognition in bilingual speakers.

The research looks mainly to the BIA+ (Bilingual Interactive Activation, Dijkstra & van Heuven 2002a), the most prominent model of bilingual visual word recognition, for theoretical support. Related theoretical accounts (i.e. the BIA model, Dijkstra & van Heuven 1998; and the Inhibitory Control model, Green 1998a) are also taken into consideration. The research aims to test how well these current models can capture the complex architecture of the bilingual mind and the processes underlying visual word recognition. The research also aims to provide information on factors influencing bilingual visual processing (including sub-lexical language-specific cues, proficiency, and the transparency of two languages in their print-to-sound mapping), with a view to guiding future experimental design and research directions.

1.2 Importance of the research

In the introduction to his chapter on visual word recognition, Balota noted that "[the] word is as central to psycholinguists as the cell is to biologists" (1994:303). Andrews (2006) suggests three main reasons for this. Firstly, interest in visual word recognition arose because of the obvious fact that the ability to recognise words is the baseline for literacy. Second, experimental paradigms designed to investigate word recognition processes provide a vehicle for exploring other cognitive processes, such as memory structures and psychopathological disorders (e.g. aphasia). Finally, visual word recognition research offers crucial insights into general issues of pattern recognition and memory retrieval.

Apart from these more general aspects of the importance of visual word recognition research, the present thesis identifies three other aspects. Firstly, given that previous research has provided little empirical data relating to the effects of proficiency, the present research addresses this issue in each of the studies presented in the experimental chapters below. The findings reported below suggest that proficiency effects may be primarily reflected in the different use of the resources available for the execution of a task. Another important aspect of the research in this thesis is that it analyses the extent to which sub-lexical language-specific cues can influence the activation of a particular language (i.e. reduce non-selective access). The findings reported below suggest that such cues are particularly effective when processing is less automatic. Finally, the research also provides perspectives from both the L1 and the L2 of (unbalanced) bilingual speakers, as well as from two languages that have different degrees of transparency of print-to-sound mappings. These investigations underpin the importance of conducting studies with complementary bilingual groups (i.e. conducting a task in both L1 and L2 and reversing this L1/L2 pattern across two different groups of bilinguals), as well as the need to consider which languages are involved. This is because different processing strategies may be employed in each of the two languages, depending on the degree of similarity of key aspects of the languages concerned.

1.3 Background to the research

The current research is based on the premise that bilingual lexical access is non-selective in nature. While this position may feel counterintuitive to someone who is a fluent speaker of two languages, researchers have collected a wide range of evidence in support of this view, using different stimuli and tasks (see Chapter 3, for an extensive overview of some relevant research). Consequently, a question that arises is whether the activation of a particular language can be limited to some extent, as long as specific language cues are provided. Views on this matter are quite divided: while earlier research suggests that certain language information may eliminate the activation of a task-irrelevant language (e.g. Elston-Güttler, Gunter, et al. 2005), other researchers argue that this may only be possible at some later processing stages (e.g. Libben & Titone 2009), and yet others doubt whether it may be possible at all (cf. Kroll, Bobb, & Wodniecka 2006). Notably, research that argues that lexical activation is limited (in some way) to the language that is relevant to the task has come predominantly from studies providing language information in the form of a semantically constraining

context sentence. Theoretically, however, language information inherent to the visual stimulus itself (e.g. in form of a diacritical marker) may have a similar language-constraining impact (cf. Dijkstra 2005). Study One in this thesis (see Chapter 3) is the first study to clarify this issue.

As noted above, most bilingual word recognition research has been conducted with visual stimulus material, and relatively little attention has been assigned to understanding the role of phonology during visual processing. Considering the previous account of bilingual lexical access (i.e. the non-selective view), it is reasonable to assume that phonological representations across the two languages of a bilingual speaker should become activated; provided that phonology gets activated at all. Support for phonological activation in visual processing has come from various studies (e.g. van Wijnendaele & Brysbaert 2002), but it is less clear what effects phonological activation may have on visual word recognition (e.g. Dijkstra, Grainger, & van Heuven 1999; Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen 2010). From a theoretical point of view, one way of clarifying this issue is to achieve a better understanding of the precise interactions between the orthographic and phonological levels (cf. Stone, Vanhoy, & van Orden 1997). Study Two (Chapter 4) represents the first attempt at such a clarification in an exclusively visual word recognition task.

1.4 Organisation of the research

Although the research focus of the thesis is on the processes involved in *bilingual* visual word identification, it is important to include a brief overview of current theories of lexical recognition in monolinguals, because this is where the relevant cognitive research has initially been carried out. The results from the extensive research into the architecture of the mental lexicon and the processes underlying lexical recognition in a monolingual reader provide a background for research on bilingual lexical processing.

The next chapter focuses on the key elements of the available monolingual research, in order to give the reader a brief account of the most relevant work and to clarify some of the assumptions that have led to the corresponding bilingual investigations, as discussed in the following chapters. Accordingly, Chapter 2 comprises two parts. The first part focuses exclusively on monolingual research and summarises the current standpoint of research investigating sub-lexical processing at both the orthographic and phonological levels. The second part of Chapter 2 discusses the question of proficiency in current

bilingual research and introduces the theoretical framework relevant to this research, namely the most prominent models of bilingual visual word recognition.

Chapter 3 focuses on the extent to which bilingual speakers make use of sub-lexical language-specific information to determine the language membership of an item, and consequently the relative activation of a particular language. The chapter begins with an overview of relevant bilingual research, including the formulation of specific research questions for the study presented in the chapter. Two experiments are reported in the context of this study. Each experiment provides a methodology and a statistical analysis section, which are followed by a presentation of results and an extensive discussion of both the error rates and response time analyses. The experiments then provide a brief summary of the main results. The chapter concludes with a general discussion section, which brings together the results of both experiments and explores the findings within the chosen theoretical frameworks.

Chapter 4 explores the extent to which activated phonological information transmits signals both back to the orthographic level and across languages. This chapter also begins with an overview of relevant bilingual research, at the end of which specific research questions are formulated. These questions are addressed in a study consisting of two experiments. Prior to the presentation of these experiments, their common methodology and statistical analysis procedures are presented. Results for each experiment are then presented in turn, together with summary discussions. The chapter concludes with a general discussion comparing the findings of both experiments and considering their implications in the context of current theoretical models of bilingual word recognition.

The final chapter of this thesis (Chapter 5) provides a discussion of the major findings from both studies, along with a consideration of potential limitations of these studies and future research directions.

Chapter 2

CURRENT THEORIES OF VISUAL LEXICAL RETRIEVAL

2.1 Introduction

This chapter comprises two parts. The first part contains a literature review, which focuses mainly on the current theories of lexical recognition in the monolingual area, limiting its scope to the processes identified at the sub-lexical level. An equivalent bilingual account of these processes is provided in the corresponding chapters. The second part of this chapter provides a description of two key issues in bilingual research relevant to the studies undertaken in this thesis, and introduces the current and most prominent models of bilingual visual word recognition. The chapter concludes with an outlook to the two experimental studies.

2.2 Part One: Monolingual Research

Consider the following group of words: *integral alerting, altering, relating triangle*. What these words have in common, is that they all share the same set of letters, which differ only in regard to their position in the word. Written language contains a remarkable number of such anagrams. The obvious question that arises here is how are readers able to distinguish between a group of similar-looking words with relative ease? A priori, it could be assumed that information about letter position must be both extracted from the stimulus and, simultaneously, form part of the stored representation in order to avoid confusion about which word is currently read. Given this assumption, a number of fundamental questions can be asked. For instance, how are words stored, processed and retrieved from the mental lexicon? What variables are crucial in letter and lexical recognition? Is the phonological representation of a word accessed during silent reading? Are word-like nonwords processed in the same manner as real words? These and other questions will be the main focus of the following discussion, providing a theoretical framework on which bilingual visual word identification research has been based. As mentioned earlier, the discussion will be largely limited to the sub-lexical processing stages, given the overall goal of the thesis.

2.2.1 Sub-lexical information processing

2.2.1.1 Orthography

It may seem surprising, but research investigating the initial, sub-lexical stages of visual word processing is relatively sparse (cf. Grainger 2008; Lupker 2005). Two papers by Grainger and Dijkstra (1996) and Grainger (2008) summarise what, according to the authors, has been the most revealing and influential research from more than a century, providing insights into our current understanding of orthographic coding (i.e. the kind of information retrieved from a visual input) and other processes involved during silent reading. Some of the early word recognition research centred around the question of the size, type and form of information stored in memory (see, for instance, Cattell 1886; Reicher 1969; Wheeler 1970, for some of the pioneering works). Based on this research it has been suggested that, since words are made up of letters, a letter should be the primary unit of analysis, followed by the word as a whole.[1] This intuitive notion has in fact been applied by the majority of computational implementations (see, for instance, section 2.3.2 below) which include the single letter as an elementary mediating unit. In empirical research, much evidence for the importance of individual letters as a primary orthographic unit (or code) has been obtained from experiments which included measurements of the probability of an orthographic structure, i.e. a given letter in a given position in a word of a particular length, also referred to as *spatial redundancy* or *position-specific letter frequency* (Grainger & Dijkstra 1996:142). This research demonstrated a performance advantage which correlated with the position-specific frequency of a letter (e.g. Mason 1975; McClelland 1976; McClelland & Johnston 1977; Rey, Ziegler, & Jacobs 2000), and the findings were interpreted as pointing to the central role of letters or graphemes as functional orthographic units.

As an alternative to this arguably very rigid position-specific coding scheme for orthographic input, researchers have subsequently suggested a more flexible orthographic decoding mechanism. For example, perceptual identification experiments using masked priming and conducted by Humphreys, Evett, & Quinlan (1990)

[1] Researchers have also suggested that larger sub-lexical units (i.e. larger that the single letter but smaller than the entire word) may be the critical mediators between first visual contact and access to the actual word representation. A number of proposals have been made, including letter clusters (but see Plaut, McClelland, Seidenberg, & Patterson 1996, for contradictory evidence; i.e. bigrams or trigrams, e.g. Seidenberg 1987; Seidenberg & McClelland 1989), orthographic syllables (e.g. Taft 1979), phonologically defined syllables (e.g. Carreiras, Alvarez, & Devega 1993; Conrad, Grainger, & Jacobs 2007), or even morphemes (e.g. Rastle, Davis, & New 2004; Taft & Forster 1976); however, none of them could be unambiguously defined as the functional sub-lexical unit.

demonstrated that short letter sequences can prime longer words, as long as the relative position of some letters (or letter sequences) in the prime reflects their position in the target word (e.g. botk-BLACK; or, oitk/ohik-WHITE where both items prime the target, so it is the position of letters relative to one another and not just relative to the target word that is important). In addition, there is persistent evidence (e.g. Forster, Davis, Schoknecht, & Carter 1987; Perea & Lupker 2003) for significant positive priming with transposed letter primes (e.g. gadren-GARDEN; jugde-COURT).[2] Taken together, these results have led to the general consensus that in the earliest stages of visual processing

> individual letters are the key elements for orthographic processing and that it is the mechanism used to code for the positions of these letters that critically determines the nature of the orthographic code. […] The relative success of [a] letter position coding scheme [however] hinges on the flexibility that it assigns to the process of coding for letter position. (Grainger 2008:22–23; see also Davis 1999; Dehaene, Cohen, Sigman, & Vinckier 2005; Grainger & van Heuven 2003; Whitney 2001)

What is still under debate is the precise letter position coding scheme that is in place during the sub-lexical stages of visual word recognition, and how the scheme might be learned (e.g. Dehaene et al. 2005; Forster, Witzel, & Qiao 2010; Goswami & Ziegler 2006; see also Grainger 2008, for a discussion of recent proposals).

One final research direction with respect to sub-lexical orthographic coding involves the mapping of feature information onto letter representations (e.g. Chauncey, Holcomb, & Grainger 2008; Lété & Pynte 2003; Perea & Rosa 2002). While some early (anglophone) studies have pointed to the advantages of lower case in processing – resulting from the presence of ascenders and descenders that are not found in CAPITALS (Smith 1969; Tinker 1963) – other research has found no clear differences between reading a lower case as opposed to an upper case text (e.g. Paap, Newsome, & Noel 1984; see also Legge & Bigelow 2011, for a recent overview of relevant research). Consequently, most theorists and computer modellers have embraced the idea that input detectors are largely case, font, and size insensitive; hence, the primary research focus has been directed to later processing stages. However, recent research, employing more sensitive research methods and tools (e.g. ERP, fMRI and EEG), suggests that the first processing stages are *not* invariant to the size and font (or shape) of the input, but rather that any

[2] Similar findings have also been obtained using other experimental paradigms, such as the LDT (e.g. Andrews 1996; Holmes & Ng 1993), where nonwords which have been created by reversing one letter (e.g. JUGDE) are more difficult to reject than those which have undergone a one-letter exchange (e.g. JUDPE), or even completely unrelated control nonwords (e.g. SLINT).

script deviations are resolved very early on (e.g. Chauncey et al. 2008; Grainger, Kiyonaga, & Holcomb 2006; Grill-Spector et al. 1999; Holcomb & Grainger 2006; Petit, Midgley, Holcomb, & Grainger 2006; Sawamura, Georgieva, Vogels, Vanduffel, & Orban 2005).

2.2.1.2 Phonology

Investigations concerning (sub-lexical) phonology have largely concerned two interrelated issues, namely whether phonological information is automatically retrieved upon the presentation of a visual stimulus (in silent word recognition) and if so, the extent to which it interacts with lower-level (orthographic) information. These issues are set out in the next two sections.

2.2.1.2.1 The phonological involvement debate

One of the longstanding debates in visual word recognition research has been whether retrieval of lexical information (or skilled reading) necessarily requires the involvement of phonology (see Frost 2005; van Orden & Kloos 2005, for recent and extensive discussions of this and related issues). Two major theoretical positions have developed: the *dual-route theory* and the *single-route strong phonological theory*. The proponents of the first theory have collected evidence suggesting that the meaning of a written item (word) *can* be recovered without any involvement with phonology (e.g. Baron & Strawson 1976; Besner & Smith 1992; M. Coltheart 1980; M. Coltheart, Rastle, Perry, Langdon, & Ziegler 2001; Forster 1976; Morton & Patterson 1980; Paap & Noel 1991; Zorzi, Houghton, & Butterworth 1998). In contrast, their opponents (i.e. those in favour of the strong phonological theory) argue that meaning cannot be recovered from print if the written stimulus is not primarily recoded in a phonological form (e.g. Frost 1998; Lukatela & Turvey 1994a, 1994b; van Orden, Pennington, & Stone 1990).

The majority of research conducted on monolingual visual word identification processes has come from Anglophone research communities and institutions; for various reasons.[3] One reason may be because English sets itself apart from other (at least Germanic)

[3] More recently scientists have repeatedly pointed out the outlier position of English, questioning the foundations of current knowledge of skilled reading, reading development, and dyslexia (e.g. Share 2008; Vaessen et al. 2010; Ziegler, Bertrand, et al. 2010). Share (2008:584), for instance, argued recently that "the extreme ambiguity of English spelling-sound correspondence has confined reading science to an insular, anglocentric research agenda addressing theoretical and applied issues with limited relevance for a universal science of reading". Other researchers have cast doubt on the relevance and function of a (*fallible*) phonological pathway during orthographic processing, particularly with respect to English (see Seidenberg 1985:3, for a detailed argumentation). Clearly an important issue, this is central to one of the studies in the present thesis (Chapter 4), where a discussion of relevant aspects is also further elaborated.

languages, by being a highly non-transparent language (see Chapter 4, section 4.1.1.2, for a more elaborate discussion of this issue and its significance), and hence it offers an ideal opportunity for the investigation of whether and how phonology may be recovered from the visual input. For instance, English contains a large number of words with irregular spelling-to-sound correspondences, such as for the vowels in PINT, HAVE, and HEAD. Typically we find that skilled readers have no problems reading such irregular words aloud (but see section 2.2.1.2.2 below, for complementary experimental evidence and discussion). This observation might suggest that a skilled reader uses a whole word as input, disregarding the letter-sound relationship within the item. In other words, one common view is that a written stimulus can be recognised as a familiar sequence of letters if it corresponds with an orthographic representation stored in memory, i.e. the orthographic lexicon (see Rapp, Folk, & Tainturier 2001, for a review). The orthographic representation subsequently activates the corresponding phonological representation; or, alternatively, once the meaning of a lexical item is retrieved from its orthographic representation, the item's semantics may then be used to restore its phonological representation from a long-term memory store.[4] In either instance, phonology is "addressed" (Zorzi et al. 1998:1132). This is the main argument adopted by the proponents of the *dual-route theory*.

Extensive research has also shown that when presented with nonwords especially constructed to highlight irregularities (for instance, KINT, MAVE and FEAD), skilled readers will perform in a way that reflects the regular pronunciation of these letters, and not that of the individual exception words (e.g. Glushko 1979). This phenomenon has been taken up by the dual route theorists and explained by suggesting an alternative process, which omits semantic mediation altogether: the indirect route. The indirect route has been argued to rely on a sub-lexical mechanism to convert an orthographic representation into a phonologically plausible sequence of phonemes; that is, phonology is "assembled" (Zorzi et al. 1998:1132). This sub-lexical mechanism contains a set of rules which reflect the most frequent relationship between a letter and its sound (in a specific language). The mechanism is most widely known as the (sub-lexical) grapheme-

[4] Note that although early involvement of the semantic lexicon can only in rare cases be fully ruled out, the former of the two routes (i.e. bypassing semantics) is depicted as the more 'direct' and faster route, and is preferably utilised to explain word naming in skilled readers (e.g. Zorzi et al. 1998). More evidence for the existence of a non-semantic route comes from observations with impaired participants, who are able to accurately read aloud irregular words, but are unable to comprehend these words (see Rapp et al. 2001, for an extensive review).

to-phoneme conversion (GPC) system (for reviews, see Besner & Smith 1992; Carr & Pollatsek 1985; Paap & Noel 1991; Patterson & Coltheart 1987). To illustrate, common or highly regular patterns (e.g. ‹ea› → [i]) are represented more strongly than irregular patterns (e.g. ‹ea› → [ɛ]). Subsequently, a nonsense item such as FEAD is likely to be more easily retrieved as [fid] rather than [fɛd] since the former pronunciation of the ‹ea› cluster in English is overall more likely than the latter one.[5]

The advocates of the dual-route theory never adequately specified which route will be utilised under what circumstances. One view is that normal readers are able to use both processes; the two mechanisms operate in parallel, and the faster process wins, yielding recognition (e.g. Paap & Noel 1991; Seidenberg 1985). A different account, however, suggests that *normal* reading development (at least in English) can be seen as a progression from an early reliance on the slower indirect process, which involves phonological coding, to more automatic visual processing, and perhaps a greater reliance on the direct route later on (cf. Frost 2005; Pennington, Lefly, van Orden, Bookman, & Smith 1987; Ziegler & Goswami 2006).

Notably, recent computational simulations and comparisons to actual experimental data could not fully accommodate the predictions made by the dual-route theory (e.g. Pritchard, Coltheart, Palethorpe, & Castles 2012; Ziegler & Goswami 2006), and therefore an alternative, the single-route mandatory phonological recoding, needs to be considered.

In contrast to the dual route theory, *mandatory phonological recoding* theorists assume a secondary access to the semantic system, bypassing the orthographic lexicon. Consequently, visual input is at first converted into a phonological representation, by means of sub-lexical conversion rules. This representation then provides access to the semantic system. Only later, if at all, is contact sought with the orthographic system (e.g. Frost 1998, 2005:276–277; van Orden 1987). The strongest argument in support of this view is the fact that human natural languages are phonological in nature, and the huge majority of words are learned first in their spoken form. In the process of learning to read, visual stimuli and symbols are associated with the phonological form of words in order to make meaning; hence, the core lexical representations of words must be

[5] Clearly, this matter is somewhat more complex than presented here. For instance, depending on whether you look at types or tokens, different frequency patterns may arise for such spelling-to-sound correspondences; this, in turn affects the preferred (or the speed of) pronunciation. However, given that this is not a primary issue in the current work, this will not be explored further.

necessarily phonologically defined (e.g. Rubenstein, Lewis, & Rubenstein 1971). Finding experimental evidence supporting this strong phonological recoding theory has proven more difficult (if not impossible) and controversial. A common method to investigate phonological recoding experimentally has been by employing homophones (words that share the same pronunciation but differ in spelling; e.g. SAIL/SALE or HEEL/HEAL) or pseudohomophones (nonwords that are typically pronounced like existing words; for instance, BRANE or BLOO) across a large variety of experimental tasks (e.g. Daneman & Reingold 1993; Daneman & Stainton 1991; Ferrand & Grainger 1992; Jared, Levy, & Rayner 1999; van Orden et al. 1990; Perfetti, Bell, & Delaney 1988). In this context, one of the earlier studies conducted by van Orden (1987) demonstrated that in a semantic categorisation task homophonous targets received more false positive responses (e.g. *is ROWS a flower*?) than their orthographic controls (e.g. *is ROBS a flower*?). While this finding was taken to suggest that automatic phonological activation plays a major role in the comprehension of written words, subsequent studies contested heavily the argument that orthography is not involved in the word recognition process. On the contrary, further investigations showed that the inhibitory homophone effect was only limited to processing of the lower-frequency member of the homophone pair (e.g. V. Coltheart, Patterson, & Leahy 1994; Jared & Seidenberg 1991), and that it was also obtained when pseudohomophones were included in the experimental list instead of nonwords (Pexman, Lupker, & Jared 2001). The latter finding is particularly interesting because the homophone effect was found irrespective of the frequency of words. As argued by Lupker,

> [t]his should never happen if the homophone effect were due to selecting the higher-frequency member of the pair first in lexical search because that event should not be altered by changing the type of nonword being used. In contrast, this result is quite consistent with the claim that these effects are feedback effects. (2005:52)

In sum, although most researchers agree that phonological codes are (co-) activated during language processing, one might be tempted to conclude that there has been a failure to reach a consensus to date on what role exactly phonology plays in visual word recognition. As pointed out by van Orden and Kloos in a recent review,

> so far no one has discovered a generally robust phonology effect in skilled reading. [...] And yet, although no particular phonology effect can be found to familiar words in all contexts, a phonology effect of some kind can be found in most contexts. (2005:67)

2.2.1.2.2 The feedforward and feedback effect

Closely related to the homophone effect mentioned above is the *feedforward consistency effect*. This effect is based on the occurrence of spelling-to-sound irregularities, where one specific orthographic code maps onto multiple phonological codes (this phenomenon is particularly common in English and French). As with the homophone effect, a common finding here is that given certain experimental settings and constraints (including task requirements, the respective frequency of a homophone member, etc., Ziegler, Montant, & Jacobs 1997), performance on *inconsistent* words is worse compared with performance on *consistent* words (e.g. Andrews 1982; Schwartz, Kroll, & Diaz 2007:107; Ziegler et al. 1997:533–534). To illustrate, inconsistent words such as DOUGH (with up to five competing pronunciation patterns) or WIND generate longer response latencies and higher error rates in the naming task than consistent words such as LUCK or RENT (cf. Jacobs, Rey, Ziegler, & Grainger 1998:152–153). The theoretical explanation of this effect is that spelling-to-sound inconsistency gives rise to (or feeds forward) the activation of multiple pronunciation patterns, which compete for selection, and the need to resolve this inconsistency is what slows the recognition process (cf. Ziegler, Petrova, & Ferrand 2008:643).

This (in-)consistency account has been perhaps most difficult to replicate in the lexical decision task (LDT), due to the purely orthographic nature of the task and the disputed degree of phonological involvement (see section 2.2.1.2.1 above; cf. Ziegler et al. 2008:643). One of the most influential contributions to shed further light on this issue, however, was a study conducted by Stone and colleagues (1997). The authors' main argument challenged the more traditional, unidirectional bottom-up approach of information processing (cf. Forster 1976; Paap, Newsome, McDonald, & Schvaneveldt 1982; see also Norris, McQueen, & Cutler 2000, for later argumentation from speech recognition research), to argue for the existence of an interactive flow of activation between the different processing levels and codes (an assumption which has been adopted in virtually all computational models of visual perception). Their argument, that the flow of activation between the orthographic and phonological codes is inherently bidirectional, implies that there must be evidence not only for feedforward (spelling-to-sound), but also for feedback (sound-to-spelling) consistency effects. Stone and colleagues provided evidence for their hypothesis by means of a more rigid control of their experimental items compared to other studies (see stimulus list used by Andrews 1982), categorising the stimuli based on four conditions: bidirectionally consistent (e.g.

DUCK; there is only one way to spell and pronounce the word's body) versus bidirectionally inconsistent (e.g. WORM; alternative pronunciation and spelling of the body, as in DORM and FIRM, respectively), feedforward inconsistent (e.g. MOTH; body has an alternative pronunciation, as in BOTH) and feedback inconsistent (e.g. HURL; body can be spelled differently, as in GIRL). The researchers obtained clear results showing that performance is inhibited (i.e. delayed responses and increased error rates) when words are in some way inconsistent; that is, only bidirectionally consistent words indicated an unaffected performance.

Ever since its first detection, the feedback effect has been replicated in numerous studies (for a recent overview of relevant studies, see Ziegler et al. 2008) and extended to include and reinterpret the research conducted on homophones (cf. Lupker 2005; Schwartz et al. 2007). Researchers even pointed out that "[an] interesting aspect of feedback inconsistency is that it should be an important variable for cross-linguistic research" (Jacobs et al. 1998:154), an aspect which forms the basis of the second study (Chapter 4) reported in this research.

2.2.2 Lexical level processes

At the lexical level, a number of factors have been identified as playing a crucial role in the speed of processing and successful activation of an item. These factors include item frequency and the number and frequency of neighbouring items (e.g. Andrews 1997). For some of these factors, extensive psycholinguistic research has observed recurring behavioural response patterns and identified them as typical experimental effects. Accordingly, experimental stimulus selection must take as many of these factors into consideration as possible.

While all studies conducted in the present thesis controlled for the factors identified as important in previous monolingual and bilingual visual word recognition literature, these higher-level processing factors were not the main focus of investigation and will not be discussed further. For comprehensive overviews of some of the long-established effects, however, see for instance Grainger and Dijkstra (1996), Lupker (2005), and Balota, Yap and Cortese (2006).

2.3 Part Two: Bilingual Research

A priori it can be expected that the aspects of monolingual visual word recognition identified in the previous section will similarly affect investigations conducted with bilingual participants. While there have been extensive investigations of factors influencing bilingual word recognition at the lexical level, less attention has been devoted to lower-level processes. Consequently, a major aim of this thesis is to provide more insight into this area of research, focusing on sub-lexical processes at the orthographic (Study One, Chapter 3) and phonological level (Study Two, Chapter 4). Given that a detailed discussion of previous relevant work will be provided for each study, these issues will not be further outlined here. Instead, another key aspect explored in this thesis will be introduced in the next section, namely the bilinguals' proficiency in their L2.

Finally, it could be argued that while research in the area of monolingual word recognition has had a long tradition of observation, experimentation, and hypothesis testing, a prominent feature of research conducted in the bilingual domain is the advancement of computational modelling verified using behavioural data. For this reason, the concluding section of this chapter will briefly discuss three prominent models of bilingual visual word recognition, which will be considered in the light of the findings of the studies conducted in this thesis.

2.3.1 Bilingualism and L2 proficiency level

Bilingual research practices

Adopting the distinction of Mitchell & Myles (2004:23), second language learning differs from bilingualism to the point that it refers to "learners who embark on the learning of an additional language, at least some years after they have started to acquire their first language", rather than simultaneous exposure to more than one language in the first years of life, as is usually the case for a bilingual speaker.

As a common practice and despite this distinction, bilingual word recognition research adopted the term *bilingual(s)* to address their participant population, even though virtually all subjects are (relatively) high proficiency second language (L2) learners. Moreover, the term bilingual is sometimes extended or interchangeably used to refer to multilinguals, i.e. those speaking and/or reading more than two languages (cf. de Angelis 2007, chapter 1; Grosjean 1992).

Depending on the concern of the investigation, the research to date has distinguished between *coordinate* and *compound* bilinguals, *early* and *late* bilinguals, and *balanced* and *unbalanced* bilinguals. The first distinction is usually applied when the experimental focus is on the architecture of bilingual conceptual information, i.e. shared set of concepts for compound bilinguals' two languages versus two distinct sets of concepts for coordinate bilinguals (e.g. de Groot 1992; Kroll & Tokowicz 2005). The second pair of terms (early vs. late bilinguals) has been particularly applied in research employing neuroimaging studies, examining the differences between participant groups with respect to their age of acquisition (AoA) of an L2 (cf. Kroll & Tokowicz 2005). Finally, the focus of other researchers has mainly been on balanced bilinguals (e.g. Thomas & van Heuven 2005). Van Wijnendaele & Brysbaert (2002) have recently elaborated the distinction between *balanced* and *unbalanced* bilinguals and the research connected with it. The authors report that early research on bilingualism was predominantly interested in the former group of participants (balanced bilinguals), i.e. those who grew up bilingually from birth. This direction of research, however, was soon abandoned since it became clear that completely balanced bilinguals are rare and, thus, such research has limited practical value. (For instance, it has been argued that the typical diglossic bilingual situation leads to lack of "balance".) In other words, given that the vast majority of the population who knows and/or speaks more than one language are unbalanced bilinguals it seemed more reasonable to focus on this particular participant group (van Wijnendaele & Brysbaert 2002:618).

As argued most strongly by Grosjean (1997, 1998, 2001) research with unbalanced bilinguals should optimally take into consideration different proficiency levels among the participants as this might have a major impact on the participants' mental processes. A similar argument was adopted by van Hell & Dijkstra, who argued that the relative fluency in their two languages

> will affect the bilingual's sensitivity to L1 interference when he/she is processing in L2, and the sensitivity to L2 interference when processing in L1. The underlying rationale here is that less activation is needed to recognize words that are used relatively frequently, as are words in a language in which the speaker is relatively proficient. (2002:782)

It is surprising then that despite the repeated acknowledgment of the importance of bilingual participants' proficiency level (e.g. Dijkstra 2005; Lemhöfer et al. 2008; Schwartz & Kroll 2006), bilingual research to date has largely omitted this issue.

In the few studies that have attempted to include proficiency level of their participants as one of the influential factors we find a large variety of proficiency measures employed to determine this variable. For instance, while some researchers used the number of instruction years as a main determiner of the L2 proficiency level (e.g. Duyck, Diependaele, Drieghe, & Brysbaert 2004; van Hell & Dijkstra 2002), others asked their participants to self-assess their language skills (e.g. Elston-Güttler, Paulmann, & Kotz 2005; Haigh & Jared 2007), or completed an on-line vocabulary test (Lemhöfer, Dijkstra, & Michel 2004). A further complicating factor is that participants assigned to different proficiency levels may differ in their age, so that the claim could be made that differences between the groups can to some extent be due to cognitive differences, familiarity with experimental tasks, or greater experience in general (Schulpen 2003). Regardless of the measure of proficiency (but perhaps due to the limitations of the measuring techniques available to the researchers), the common practice adopted in these studies has so far been an allocation of bilinguals into separate proficiency groups, rather than to investigate more gradual changes that come with increasing L2 proficiency (but see Chambers & Cooke 2009, for an exception). This aspect is addressed in both studies discussed in this thesis, providing new insights into the effects of proficiency on bilingual visual word recognition and the development of the mental lexicon(s).

Bilingual processing and the effects of automatisation

An important issue related to increasing proficiency is the level of automaticity of cognitive processing.[6] Automatic processing has been extensively explored in the monolingual domain, where an automatic process has been defined as being effortless, unconscious, and involuntary, as well as making few demands on cognitive resources (see Tzeglov 1999, for a comprehensive review of conducted work and a discussion of relevant criteria). Among others, two classified empirical phenomena associated with automatic processing include the observation that automatic processing is fast, while (the opposite) controlled processing is slow, and that "extended consistent training is required in order to develop automatic processing" (Schneider & Chein 2003:6).

[6] Another key aspect of increasing L2 proficiency on bilingual visual processing is the amount of cognitive control. Given that the effects of cognitive control are not investigated in this thesis, this aspect will not be further explored. For relevant research and overview of this topic, however, the reader is referred to the work by Schulpen (2003, chapters 3-7) and Tzeglov and Kadosh (2009).

These two phenomena can be directly related to bilingual processing and the effects associated with increasing proficiency in an L2. More specifically, on the assumption that proficient bilingual speakers have collected more experience in their L2 (hence, an 'extended consistent training'), these speakers' performance is characterised by faster word recognition, which seems also less affected by their L1 compared to less proficient L2 speakers (e.g. Schulpen 2003; Tzelgov, Henik, & Leiser 1990).
In the bilingual domain, a number of investigations have focussed on finding a suitable measure of automaticity in L1/L2 visual word processing (e.g. N. S. Segalowitz & Segalowitz 1993; S. J. Segalowitz, Segalowitz, & Wood 1998; but see Schulpen 2003, for evidence of the limitations of a proposed measure). This is not the research direction undertaken in this thesis. Instead, the thesis (and more explicitly explorations in Study One, Chapter 3) focuses on an investigation of how the 'demands made on cognitive resources' may change with increasing levels of automaticity.

2.3.2 Modelling bilingual visual word recognition

Over the past few years, computational modelling has become a key approach to testing certain hypotheses about processes underlying visual word recognition. As argued recently by Ziegler and colleagues "computational modelling is particularly useful when the system to be investigated is too complex, too interactive, or too difficult to deal with directly" (2010:642). However, the researchers point out that every model is to some extent false, being typically incomplete or oversimplified (see also Dijkstra & De Smedt 1996, for an extensive overview and arguments in favour of computational modelling).
Within the bilingual domain, the most prominent model in the past decade to capture the processes underlying visual word recognition has been the BIA+ model (bilingual interactive activation; Dijkstra & van Heuven 2002a). The BIA+ is a successor of the BIA model (Dijkstra & van Heuven 1998; Grainger 1993; Grainger & Dijkstra 1992; van Heuven, Dijkstra, & Grainger 1998), having adopted several features from two other models of visual word recognition, namely the IC (inhibitory control; Green 1998a, 1998b) and the SOPHIA (semantic, orthographic, and phonological interactive activation; van Heuven & Dijkstra 2001) models. The following sections will briefly introduce each model – according to their order of appearance – and consider how they can account for the parallel activation of different languages (i.e. the non-selective access view), the interactions taking place within and between the sub-lexical level(s) (e.g.

between orthography and phonology), and how the language membership of a stimulus is determined.

BIA

The BIA model is a localist connectionist model, in that it is based on the principles of interactive activation and competition between lexical representations. As an extension of the monolingual interactive activation (IA) model developed by McClelland and Rumelhart (1981), the BIA model assumes that a bilingual possesses an integrated L1/L2 lexicon in which lexical representations from both languages are accessed in parallel, in a language non-selective way.

The model consists of four levels of representational units (or 'nodes'), proceeding from the bottom-level layer of orthographic features, to letters, orthographic forms of whole words, and finally a top-level layer of language nodes (see Appendix A, Figure 105). Whereas the bilingual's two languages are modelled to share the orthographic feature and letter nodes, the word nodes are organised in language subsets, but are fully connected to one another. For a bilingual, the layer of language nodes comprises only two nodes, one for each language. Connections and interactions between neighbouring (higher and lower) levels are established by means of excitatory and inhibitory signalling pathways. The signalling pathways are not limited to neighbouring levels, but also exist within levels. Note that the signal within a particular level can only be inhibitory and acts as a suppressor of competitor items at that level. This mechanism is commonly referred to as "lateral inhibition" and functions equally as an inhibitor of word nodes from different languages (e.g. Thomas & van Heuven 2005).

For instance, when a stimulus is perceived visually, the flow of activation is mainly bottom-up, i.e. from the feature level to the letter units in the correct position of a word (e.g. 'b' in the onset position), and ultimately to the word level. Once the signal reaches the word level, words from both languages that have the letter at the correct position (e.g. 'book' and 'bunt', meaning *colourful* in German) are activated. At this word level, activated words inhibit each other, irrespective of the language they belong to. (So, for instance, 'book' and 'bunt' would start inhibiting each other.) At the same time, a top-down signal reinforcement from the word-level to the letter-level is put in place, as well as further bottom-up activation of the corresponding language node(s). That is, an English word (e.g. 'book') will activate the English language node, which will start inhibiting any German words. In contrast, a German word will have the opposite effect,

namely to activate the German language node, which will inhibit all English words. As described by the modellers "[t]he language nodes collect activation from words in the language they represent and inhibit active words of the other language. The activation of the language nodes reflects the amount of activity in each lexicon" (Dijkstra & van Heuven 2002a:177). The mechanism involved here is identified as interactive processing (Grainger 2008). This mechanism ensures that highly activated representations persistently and finally fully inhibit the activation of less activated units. With time, one word unit will become the most active, and lexical selection will take place when this word unit reaches a certain activation threshold for recognition. Different words are assumed to have different resting level activation, which is highly dependent on the word's frequency relative to other words.

While the model was successful in replicating a large number of effects identified in behavioural data (see Dijkstra & van Heuven 2002a:178–180, for a full account of successful simulations), it failed to account for other empirical findings, such as cross-linguistic phonological effects (p. 181). This is not surprising, given that the model does not include any phonological level(s) and/or representations. This obvious shortcoming was later addressed in the model's successor, the BIA+ (see below).

The most important feature of the BIA model for the purposes of the current study (particularly Study One presented in Chapter 3) is its incorporation of language nodes. Dijkstra and van Heuven (2002a:177) discuss four main functions of the language nodes, which can be summarised as follows. First, language nodes serve as "language tags" (or language labels), indicating the language membership of each item. Second, collecting the activation of all active words in a particular language, they reflect the "global lexical activation" levels associated with that language at one specific point in time. Third, language nodes may function as "language filters" or a language selection mechanism, given that they may differ in relative activation levels across languages. And finally, language nodes could in principle collect context information from sources outside the word identification system (e.g. participants' expectations regarding the task), and use this information in a top-down inhibitory way. Hence, this would lead to interactions between context and lexical identification processes.

IC

A similar, yet different, approach with regards to language membership has been adopted in the Inhibitory Control (IC) model, which was published in the same year as

the BIA model. Notably, the IC model has been more frequently applied to bilingual language production and translation issues than to visual word recognition processes (cf. Schulpen 2003:52); hence, the following discussion will be limited to the most important aspects of the model for our current purposes.

The IC is constructed from seven elements (see Appendix A, Figure 107). Its main element is the supervisory attentional system (SAS), which is the model's overarching control mechanism. This SAS mechanism explicitly regulates language behaviour by either retrieving or setting up task schemas. The task schemas are "mental devices or networks that individuals may construct or adapt on the spot in order to achieve [or execute] a specific task" (Green 1998a:69), and their distinct feature in the IC is that different task schemas may inhibit each other. Accordingly, the IC assumes that assigning a task to a participant leads to the activation of a particular task schema, which in turn will change the activation and subsequent selection of lemmas in the bilingual lexico-semantic system (see below). A lemma (in the IC) stores information about the word's form, meaning and syntactic properties, and contains its specific language tag, indicating the word's language membership.

The other key elements in the IC include a "conceptualizer", which is a non-linguistic system that builds conceptual representations from information in long-term memory, and which is driven by the "goal", i.e. an intention to produce a word in a particular language. Finally, the model also incorporates a "bilingual lexico-semantic system", which stores all of a bilingual's "lemmas".

Perhaps the most important aspect of the IC for the present purposes is that it proposes a similar top-down inhibitory control and modulation of lexical activation in a particular language to that of the previously discussed BIA; implicitly via the language tags. In the IC, this modulation can be done in two ways. The first occurs externally, or pro-actively, via an interaction between the lemmas (tagged by a language membership tag) in the lexico-semantic system and the task schemas. That is, bilinguals adopt a particular task schema based on a specific signal from the bilingual lexico-semantic system. For instance, when a task schema is set to "press yes in the presence of items tagged as English words", items that do not contain this language tag will induce a "no" response. Adaptation of this task schema arises when bilinguals encounter conflicting signals, e.g. with an item tagged for two languages, where one signal indicates a "yes" response and the other signal indicates a "no" response. In this case, bilinguals are assumed to shift their response criterion in the task schema, for instance, "by reducing the amount by

which activation in the 'yes' response units must exceed that in the 'no' response units" (von Studnitz & Green 2002a:17). The second way to modulate lexical activation in a particular language is accomplished internally, or reactively. That is, participants alter the global activation level of units representing a particular language (i.e. either in the target or in the non-target language; von Studnitz & Green 2002a:2). As a consequence, a word unit among the highly activated units in one particular language can more easily reach the set activation threshold for recognition. In addition, these high global activation levels also have a reverberant effect on subsequent word trials, subduing the activation levels of the non-target language units.

Regardless of the question of the locus of control (or modulation of lexical activation), Schulpen identified two major problematic issues with the IC, namely "there is neither a complete account of how the recognition of items within the lexico-semantic system takes place, nor an account of how the task demands in lexical decision would vary […] under the influence of stimulus lists with different compositions" (2003:182). This leads us to an aspect that the BIA has similarly omitted, namely a more precise description of the interactions at the sub-lexical level, including some phonological representations.

SOPHIA and *BIA+*

Due to their close proximity and the same developers, the last two models can be discussed together. In the semantic, orthographic, and phonological interactive activation (SOPHIA) model van Heuven and Dijkstra (2001) have addressed the major problematic issues identified in the previous two models, by providing a detailed account of the orthographic, phonological and semantic interactions. Accordingly, the architectural structure of SOPHIA includes three sub-lexical orthographic levels of units (i.e. letters, letter clusters smaller than the syllable unit, and syllables) and a fourth unit, the word level (see Appendix A, Figure 106). For the first time, phonology is represented in an analogous way, in four corresponding levels of nodes. The processing assumptions are largely consistent with the preceding BIA model. More specifically, nodes at one particular level (e.g. orthographic clusters) can activate and inhibit units in neighbouring levels via excitatory and inhibitory connections, respectively (i.e. orthographic letters and orthographic syllables). In addition, representations at a particular orthographic or phonological level mutually inhibit each other via lateral inhibition. In contrast, the signalling pathway between an orthographic level and its analogous phonological level are of an excitatory nature only (i.e. they do not inhibit

each other). For instance, if the visual stimulus 'boy' has been presented, it will activate its orthographic word node for 'boy' and the corresponding phonological node [bɔɪ]. Another modification performed in SOPHIA is its inclusion of a semantic system which can directly interact with remaining structural levels in the model. Finally, and most importantly, another major alteration to the SOPHIA model relates to the language nodes. The reader will recall that in the BIA word nodes were connected to their corresponding language node via excitatory pathways, and that in turn a language node exerted inhibitory influence on all the word nodes from the opposing language. The latter inhibitory connections have been removed from SOPHIA, given that "evidence has accrued suggesting that the [proposed] functions of the language nodes must be assigned to different levels of processing and that they may operate in a different way" (Dijkstra & van Heuven 2002a:186). Because these top-down connections served several crucial functions (including a language membership determining mechanism; see above), an alternative mechanism needed to be put in place; one that can be seen in the BIA+ model.

The BIA+ model distinguishes between two systems (see Appendix A, Figure 108): the word identification system – consisting of a variety of linguistic representations, equivalent to those constructed in SOPHIA – and the task/decision system, which incorporates non-linguistic task schemata specifications (e.g. participants' expectations, task demands, etc.). In the BIA+ the task/decision system continuously reads out the activation present in the word identification system. At the appropriate point in the schema, it weighs different kinds of activation input from the identification system to arrive at a response which conforms to the task at hand (Dijkstra & van Heuven 2002a:190–191). Similar to the BIA, a response can be made as soon as the activation of a particular lexical item surpasses its recognition threshold. Notably, the task/decision system is assumed to respond flexibly to a number of variables prior to and/or during an experiment, which may lead to dynamic adaptations of the system (cf. 'external control' in Green 1998a). (Note that a full account of relating stimulus to response in the BIA+ is given in the corresponding chapters.)

The distinction between a word identification system and a task/decision system made in the BIA+ reduces considerably the functions of language nodes. In fact, while the activation of language nodes is now solely based on the activation within the word identification system, it is assumed that the similarity of input to the internal lexical representation and not its language membership determines the relative activation of a

lexical unit. Accordingly, the BIA+ postulates that language information becomes available rather late and provides only limited feedback from language nodes to the lexical level; hence language membership can only have limited influence on word recognition (Dijkstra & van Heuven 2002a:186).

Model comparison

Out of the three most prominent theoretical viewpoints discussed above (BIA, IC, and BIA+), only the BIA+ has an explicitly implemented array of (sub-lexical) phonological representations, the manifestation of which forms the basis for the investigation conducted in Chapter 4 (Study Two). As a consequence, only the BIA+ will be considered for theoretical support in that study, in an attempt to interpret the results. While a detailed account of the processes involved at the sub-lexical phonological and orthographic level is given in the sections above, a further assessment in context of the findings presented in this thesis is provided in the corresponding chapters.

With respect to the way language membership can be determined and how it influences the relative activation of one of a bilingual's two languages (this aspect is the main focus of Study One, Chapter 3), all three models present a somewhat different account.

First, in the BIA (Dijkstra & van Heuven 1998) the state of language activation is indirectly reflected in the resting level activation of the different words (i.e. at the word form level). While this resting level activation is thought to vary depending on a participant's L1/L2 proficiency, relative language activation is also assumed to be affected by top-down inhibitory signals from the language nodes (e.g. suppressing the activation of the non-target language). These language nodes are considered to be a key language determination mechanism. The BIA model also assumes that the identification level and the decision level interact, to arrive at a common outcome (cf. Schulpen 2003).

Second, the IC (Green 1998a) makes some similar assumptions as the BIA model. For instance, early language determination is signalled by means of language tags, which are attached to each lemma in the bilingual lexico-semantic system. Information from this lexico-semantic system then feeds into a specified language task schema, which is responsible for the relative activation of a particular language. Similar to the BIA, Green (1998a) also suggests a top-down inhibitory regulation of lexical activation (at the lemma level), and an obligatory interaction between the identification and the decision levels.

Finally, the BIA+ (Dijkstra & van Heuven 2002a) shares certain features with the other two models, but sets itself apart in certain other assumptions. Language activation, for instance, is again implicitly determined by the resting level activation of words, which is predefined by a bilingual's L1/L2 proficiency. Although the model incorporates language node representations, these language nodes are no longer assumed to act as a primary language determination mechanism (for more detail, see the description above). Instead, language membership information is assumed to be retrieved "via the item's lexeme (orthographic or phonological) or lemma (more abstract syntactic/semantic) representation" (Dijkstra & van Heuven 2002a:186). Notably, even though the current structure of the model does not seem to allow an information flow from the language nodes back to word level representations, it has been argued that a feedback parameter may be in place, but that it must be set rather low (ibid.). A similar (exclusively) bottom-up information flow is assumed between the word identification component of the model and the task/decision system; which is in contrast with the interactive dynamics advocated in the other two models. Despite the similarity of the task/decision system to the language task schemas in the IC model, it is clear that its executive functions with respect to relative language activation are somewhat more restricted relative to those described for the task schemas.

In sum, with a distinct language determining mechanism set out in the first two models (i.e. the BIA and the IC), the relative activation of a particular language is assumed to progress in an interactive way (i.e. bottom-up and top-down). In contrast, the BIA+ has no explicit language filtering mechanism, and the relative activation of a particular language is assumed to be restricted to linguistic context effects only (e.g. lexical factors and syntactic or semantic aspects of sentence context). These aspects will be important for the interpretation of the impact of sub-lexical language-specific information on bilingual visual recognition, which forms the basis of Study One, and they are further discussed in the corresponding chapter.

Despite the advantages of computational modelling, it is clear that any modelling implementations and theoretical assumptions require validation through comparison with behavioural studies. It is important then to provide suitable empirical explorations and match them against the assumptions made in available computational models. This thesis provides such an investigation, focussing on some of the earliest processes during bilingual visual word recognition. The first of the following two studies focuses on sub-lexical processing at the orthographic level, and the second presents an investigation of the sub-lexical interaction between orthography and phonology across languages. Each study assesses in turn how well the results fit current theories of bilingual visual word recognition.

Chapter 3
THE ROLE OF SUB-LEXICAL ORTHOGRAPHIC CUES IN BILINGUAL WORD RECOGNITION

3.1 Introduction

Bilingual speakers may agree that sometimes the way our mind works as it attempts to deal with two languages is Strange. As a reader you might be wondering why I have capitalised the word *strange*. However, German bilinguals may have had a different reaction. In fact, the introductory sentence and rationale to my first study are based on a personal anecdote: a few years back I received a text message written all in German, except the word 'Strange' in the middle of it. It took me several minutes before realising that the word I tried to make sense of in German was actually the English word *strange*. Being slightly shocked with my failure to decipher a highly common English word, I showed it to two other German-English bilingual friends. Interestingly enough, they too spent a little while wondering what that word meant in German before noticing its English meaning! The word 'Strange' does not exist in German but it is very similar in form to the German word *Strang* (meaning "strand" in English), and the presence of an orthographic marker (i.e. a capital first letter) can encourage German speakers to think that it is a word in German.

Several questions arise when you consider the scenario presented above: can the visual form of a word activate a particular language and influence the initial word recognition stages? Could the language of the context in which the word is embedded reduce or amplify such an effect, if it existed? And in what way would less proficient bilinguals make use of language-specific information?

To address these questions, two experiments were designed for this study, exploring three related theoretical issues: (1) How does the bilingual word recognition system function in language-specific situations as opposed to a less defined language context? (2) What role exactly does language information (e.g. in the form of a language-specific cue, or a particular grammatical constraint) play in the relative activation of the two languages, for both L1 and L2 speakers? (3) How do recognition and susceptibility to language information develop with proficiency in the second language (L2)?

The following sections will begin with an overview of current theories about how a bilingual accesses lexical information and more generally about the organisation of a

bilingual lexicon. These sections will be followed by the first and second experiments, respectively, including the methodology, results and discussion of each part of the study. The chapter will conclude with a general discussion, relating the empirical data to current models of bilingual visual processing.

3.1.1 Bilingual lexical organisation – the current standpoint

A central issue in bilingual visual word recognition research has been the distinction between language-dependent and language-independent lexical storage. That is, some researchers have argued for the co-existence of two separate (non-integrated) lexicons – one for each language – while others have argued for the existence of a single, integrated lexicon for both languages. Similarly, the question of lexical access has evolved around whether it is language non-selective, or whether access of lexical information can be selectively limited to one of the bilinguals' two languages (for reviews, see Desmet & Duyck 2007; Dijkstra 2005).

While the first issue (integrated vs. non-integrated lexicon) is a structural one, the second (i.e. access to this/these lexicon/s) is a process issue (see van Heuven et al. 1998). Evidence for the first issue is provided (only) by neighbourhood and morphological family size studies (e.g. Dijkstra, del Prado Martín, Schulpen, Schreuder, & Baayen 2005). Evidence for the second issue is, for instance, provided by research involving false friends and cognates. Given that the research in the present thesis is mainly concerned with the processing aspect, the first issue will not be further discussed here.

Accordingly, research focussing on the processing aspect in bilingual visual word recognition has presented evidence that indicates strong cross-language connections at different levels: at the sub-lexical level (e.g. Brysbaert, van Dyck, & van de Poel 1999), at the lexical level (e.g. von Studnitz & Green 2002a), and/or at the conceptual level (e.g. Kroll & Stewart 1994). Given those strong interfaces between languages, several interrelated questions have been addressed. Firstly, can a bilingual ever function in the L1 or L2 without constant susceptibility to influence of one language on the other? Secondly, how well or poorly can a bilingual activate only the appropriate language at the appropriate time and to the appropriate extent?[7] Thirdly, when bilinguals are

[7] Occasionally, this discussion also appears under the term *language* or *cognitive control*, *cross-language lexical competition*, *bilingual language conflict*, and others.

presented with word-like stimuli, how do they know what language an input item belongs to? It has been suggested that this kind of information must be stored in the bilingual's mental lexicon for each word. Some researchers talk of a *language node* (Dijkstra & van Heuven 1998, 2002a), others of a *language tag* (Green 1998a, 1998b). As pointed out by Dijkstra (2005:186), each word may hold its own separate language tag/node; alternatively, all words of one language may be linked to a single language tag/node - more explicit information on the nature of such tags or nodes is still lacking. And finally, if a bilingual's two languages share the same orthography or script (e.g. both roman script), which lexical candidate is activated (i.e. from L1, from L2, or from both languages) when a letter string is presented? Such issues have been typically discussed under the heading of *language selective* versus *non-selective lexical access* (see de Groot 2011, for the most recent account of this debate).

3.1.2 Interlingual homographs

To address the question of language selectivity, researchers have typically employed experimental material which can induce subconscious activation of a task-irrelevant language. One such set of materials is interlingual homographs (IHGs). IHGs are words that share the same basic orthography[8] but carry a different meaning across two languages. In addition, these words may differ in pronunciation and syntactic category across the two languages. For instance, the German-English homograph GIFT[9] (meaning "poison" in German) shares pronunciation and syntactic category across the two languages, whereas the homograph KIND differs from its English correspondent in (almost) all respects (German meaning, pronunciation and syntactic category are "child", [kɪnt], and *noun*, respectively).

If bilingual language processing is *selective*, then the fact that IHGs belong to two languages and the fact that they have different meanings in those two languages should not affect participants' responses, and the stimuli should be recognised with the same

[8] Given that even languages using the same script (e.g. roman) differ to some extent in their writing systems, it may happen that an IHG carries a diacritical marker in one language whereas its equivalent in the other language bears no such feature. For instance, a number of IHGs have their first letter capitalised in German but not in English. The term 'basic orthography' seeks to disregard such minor (but not trivial, as will be shown in the present study) differences.

[9] Words presented in CAPITALS represent German-English homographs. Given that IHGs that are nouns in German take an onset capital letter, this representation format has been chosen to avoid a potential language bias towards a particular meaning of the word. The reader is also informed that henceforth 'onset' capital letter refers to the *initial letter* in the word only.

ease as matched control words within the target language. If however lexical access is *non-selective*, recognition of IHGs should suffer due to a "stimulus-based language conflict" (van Heuven, Schriefers, Dijkstra, & Hagoort 2008:2707), i.e. based on interference from the non-target language.

To test these hypotheses, various experimental techniques have been employed with a large variety of tasks (for a recent overview, see Degani & Tokowicz 2010). Two of the earliest behavioural studies using IHGs put forward evidence for either hypothesis. The earlier of the two, conducted by Beauvillain and Grainger (1987), used English-French stimuli in a primed lexical decision task. On some critical trials French advanced learners of English were instructed to read a homographic prime (e.g. *four*, meaning 'oven' in French) in either their L1 or L2 and then decide whether the following target item was an English word. Although in their first experiment the reading of an item in a particular language was ensured by explicit instruction (i.e. stating that the prime is to be read in French, whereas the target word was an English item), their second experiment seems to have lacked such information; hence, it is not entirely clear what language the primes were actually read in. The authors observed that when the target item was related to the English meaning of the prime (e.g. *five*, after the prime *four*) there was a strong priming effect (showing that the English meaning of the prime word had been processed). A further observation was that this effect was irrespective of the prime-language (i.e. whether the prime was assumed to have been read in English or in French). However it depended highly on the relative frequency of the prime in a particular language (i.e. "four" is a relatively more frequent word in English than "oven" is in French). Beauvillain and Grainger (1987) interpreted this result in favour of the non-selective view (see also van Heuven et al. 1998, for a similar result based on neighbourhood density).

The second early behavioural study was conducted by Gerard and Scarborough (1989), using Spanish and English lexical decision tasks with Spanish-English bilinguals. Again, the frequency of usage of the included IHGs was manipulated across the two languages. Contrary to the results of the study conducted by Beauvillain and Grainger (1987), this study showed that bilingual responses were neither affected by the frequency of an item in the non-target language, nor by the language of the task. The authors argued that "a bilingual's lexical memory is organized into separate lexicons, one for each language, and that a bilingual can selectively access a particular lexicon" (Gerard & Scarborough

1989:312; but see de Groot, Delmaar, & Lupker 2000, for a reappraisal of the results, and in favour of the non-selective view).

These mixed results continued: after more than two decades of research on interlingual homographs, the span of the collected findings ranges from inhibitory effects of one language on the other (e.g. Dijkstra et al. 1999, experiment 2; van Heuven et al. 2008; Macizo, Bajo, & Cruz Martín 2010; Schulpen 2003, chapter 5; von Studnitz & Green 2002a, experiment 1 and 2), to null-effects (e.g. Dijkstra, van Jaarsveld, & ten Brinke 1998, experiment 1; Schulpen 2003, parts of chapter 4), to facilitation effects (e.g. de Bruijn, Dijkstra, Chwilla, & Schriefers 2001; Dijkstra et al. 1999, experiment 2; Lemhöfer & Dijkstra 2004, experiment 1; Schulpen 2003, chapter 6).

Two questions immediately come to mind: what is the final verdict on bilingual lexical organisation and the access of lexical information? And, why do we observe such seemingly contrasting results? First and foremost, despite their mixed results, the studies above have pointed collectively to the finding that processing in one language is to some degree affected by the other language (see also de Bruijn et al. 2001; Caramazza & Brones 1979; Dijkstra, Timmermans, & Schriefers 2000; Jared & Szucs 2002; the mentioned null-effects have been typically reinterpreted to comply with this overall picture). This finding is consistent with the view of an integrated lexicon and that bilingual lexical access is fundamentally non-selective (van Heuven et al. 2008). Accordingly, counter-evidence suggesting that individuals are able to selectively access one language rather than another (e.g. Rodriguez-Fornells, Rotte, Heinze, Nösselt, & Münte 2002; Scarborough, Gerard, & Cortese 1984; Soares & Grosjean 1984) "has either been reinterpreted as consistent with a non-selective account or it has been shown to be artefactual" (von Studnitz & Green 2002a:1).

With respect to the second question, some researchers have argued that *task*-driven (i.e. external) factors, such as list composition and task demands, have contributed to the conflicting outcomes. Other researchers have also identified *stimulus*-driven (i.e. internal) features as having a major impact on visual processing, including the degree of phonological overlap and the relative frequencies of an item across the two languages, as well as issues related to syntactic category membership (e.g. Baten, Hofman, & Loeys 2011; Dijkstra et al. 1999, 2005).

Even though evidence for a bilingual language system that is non-selective in nature seems more convincing, there are instances in which the effects of a non-target language

on target language processing can be difficult to detect (e.g. Elston-Güttler, Gunter, et al. 2005; von Studnitz & Green 2002a). The following section(s) will discuss those cases in which the relative activation of lexical representations across a bilingual's two languages seems affected, and will consider the possible sources for such high-impact interactions.

3.1.3 Language information

An issue which has interested researchers for some time has been whether information about the language that is being read or heard can be used to speed up the processing of presented words. If language-specific information does affect the selection process, then a further question concerns the processing level (sub-lexical, lexical, etc.) at which such information becomes available during the word recognition process. A measure of the availability of such information is the extent to which it facilitates (and/or inhibits) word recognition. In other words, assuming that such information is available early in the recognition process, it might help to speed up word recognition by excluding lexical candidates from the non-target language (Dijkstra 2005:186). The following sections will consider two sources of language information (i.e. the language of a sentence frame preceding the target and sub-lexical, language-specific cues) and discuss research that has investigated their relationship with lexical selection.

3.1.3.1 Top-down information/Sentence language

An early 'language context' study by De Bruijn and colleagues (2001) suggested that a language-specific reading of interlingual homographs (IHGs) is not affected by the language of a preceding single word prime stimulus. Interestingly, this particular study and much of the crucial research employing IHGs has been limited to tasks in which the word stimuli are presented in isolation. In response to this, a new research direction has emerged more recently, which focuses on bilingual lexical activation in sentence contexts. In addition to the consideration of ecological validity, i.e. creating task conditions that are more similar to everyday situations, the use of sentence contexts allows researchers to test under what circumstances and to what degree interference from the non-target language can still be observed given 'extensive' contextual information, i.e. going beyond the language context information provided by a list of individual word stimuli.

For instance, Elston-Güttler, Gunter, et al. (2005) conducted a behavioural study combined with measurement of event-related-potentials (ERP), to test whether an L2

(English) sentence context could reduce the activation of a task-irrelevant (L1=German) meaning of German-English homographs. Participants read a visually-presented sentence, then saw a prime, and subsequently carried out a lexical decision task on a single target item presented in place of the prime. On critical trials, the sentences ended in an interlingual homograph (in italicised small capitals in the example below) and the target item for lexical decision (in small capitals in the example)[10] was related in meaning to the non-target, L1 meaning of the homograph:

> The woman gave her friend an expensive *GIFT* – POISON (= meaning of German word *GIFT*).

In addition, the researchers investigated the impact of changing the preceding global activation level of a particular language by requiring participants to view a (silent) movie narrated either in the language directly relevant to the task (i.e. L2), or in a language incompatible with the task demands (i.e. L1). A major finding of the study was that there was no measurable interference from the L1 on the L2 when the L1 stayed relatively "inactive" (i.e. when it was not deliberately activated) prior to the experimental session. This means that participants were able to "zoom into" their L2 and stay in that language setting for the duration of the experiment (Elston-Güttler, Gunter, et al. 2005:66). Similarly, although bilingual participants with increased L1 activation levels (i.e. those who saw a German movie prior to the experiment) showed some L1 interference, this was restricted to the initial stages of the experiment (i.e. they also "zoomed into" their L2 in the course of the experiment).

Similar results were obtained by Schwartz and Kroll (2006) in a study which manipulated the amount of information provided by the sentence context.[11] In their study, Spanish-English homographs were embedded in L2 sentences that either *were* or *were not* biased towards the L2 meaning of the homograph (high constraint versus low constraint sentences[12], respectively). Highly proficient (Spanish L1 – English L2)

---

[10] Both prime and target words were presented in all-capital letters in the study to avoid a language bias (see previous comment).

[11] Note that Elston-Güttler, Gunter, et al. (2005) do not report an explicit measure of a semantic or syntactic bias resulting from their sentence frames towards the prime and/or target word meaning. In fact, the information the authors provide is slightly confusing because although they claim that the sentence frames employed in their experiment were relatively open ended (i.e. without a reference towards the target word meaning, p. 62), they also argue that "the sentences themselves supported the L2 meaning only" (p. 68). Hence, it seems the sentence material was *semantically* constraining after all.

[12] The researchers are not explicit with regards to the constraint type, i.e. whether they were semantic or syntactic constraints. However, given their description of the experimental material and example sentences, my interpretation is that a reference to a *semantic* constraint is more likely.

bilingual speakers were shown sentence frames, one word at a time, and had to name out loud a target item that appeared in red. The researchers obtained no significant context-related differences in naming times or errors for the interlingual homographs, which they interpreted as indicating only limited activation of the non-target language.[13] Interestingly, the same task conducted with lower proficiency (as indicated mainly by self-assessed proficiency ratings) L2 speakers returned a different pattern of results. Similar to the highly proficient group of speakers, the naming speed for the cross-linguistic material was no different to that for matched controls, and also did not differ across the two context types (high vs. low constraint sentences). However, in contrast, less proficient bilinguals produced significantly more errors when naming homographs than when naming controls.[14] This effect was again irrespective of the sentential constraint. Interpreting these results, the authors suggest that the proficiency of a bilingual speaker in their L2 is a key factor in target meaning activation: while more proficient bilinguals are better able to activate the "functionally" more dominant L2 meanings, less proficient participants may automatically revert to the activation of the "effectively" more dominant L1and hence display stronger inhibitory effects in target meaning selection (Schwartz & Kroll 2006:209). What is at the core of this interpretation is the implicit notion that an increase in L2 proficiency represents greater language-selective access.

Perhaps a problematic aspect of Schwartz and Kroll's explanation is that it explores merely the assumption that naming errors are due to the presence of lexical competitors. Although this may be true, the researchers did not explicitly analyse the types of error in order to identify whether they were due to lexical competitors rather than something else. For instance, as pointed out by other bilingual research (see discussion in section 3.1.2) it is important to assess the neighbourhood densities, frequencies and syntactic categories for items *across* the two languages in question, as they all affect the selection process. No such measurements were reported in Schwartz and Kroll's (2006) study and so it is not clear whether the researchers controlled for any

[13] The reader is reminded that in accordance with the *non-selective access* view, both of a bilingual's languages are assumed to be activated at the same time. Accordingly, Schwartz and Kroll's (2006) interpretation can be understood if we assume minimal activation of the non-target language. This non-target language activation, however, is not strong enough to influence target-language processing and to emerge as measurable interference.

[14] The authors give no reference to their naming error criteria, i.e. whether responses were judged to be errors for just pronouncing non-homophonic homographs with the L2 pronunciation, or making any kind of naming error.

such variables. Further, the study gives no indication as to whether participants were tested for their familiarity with the presented IHGs. A high number of unknown items (particularly for less proficient bilinguals) would explain an increased error rate. Finally, the authors do not comment on the fact that in tasks which require explicit activation of a pre-lexical phonological code, less proficient L2 speakers may be more inclined to use their (more automatic) L1 decoding mechanisms and strategies; hence the observed effect may be due to processes that take place prior to lexical activation (for a more elaborate discussion of this topic and relevant literature, see Study Two, Chapter 4). Lastly, given the absence of a sentence type effect, the authors propose further that "a relatively low-constraint sentence is sufficient to more strongly activate the target meaning allowing it to compete early for selection" (Schwartz & Kroll 2006:209); or, to use Elston-Güttler, Gunter, et al.'s (2005) terminology, participants may use the most minimal information provided to "zoom into" the language of processing (i.e. a target language).

Based on this accumulating evidence, it could be argued that relative language-selective access can be achieved (i.e. by eliminating the effects of the non-target language on item processing) providing there is minimal language context (i.e. low constraint sentences) and cross-linguistic *orthographic* overlap (cf. also Duyck et al. 2007). Given such insights, the findings of Libben and Titone's (2009) partial replication of Schwartz and Kroll's (2006) study may seem surprising at first, since it provides evidence for interlingual homograph *interference* in L2 sentence contexts.[15] Libben and Titone argue that their more balanced stimulus set and their experimental technique (i.e. eye-movement measurements) may have been superior to other studies for the assessment of the earliest stages of interlingual homograph activation. While this may be true, contrasting their study with Schwartz and Kroll's (2006) design, the absence of a comparable homograph effect in Elston-Güttler, Gunter, et al.'s (2005) study seems striking. What I believe to be a much more crucial difference between the experimental setups of the studies discussed in this section of the thesis, is the inclusion (or absence) of a frequency measure of the critical stimulus set in *both* of a bilingual's languages (see, for instance, Altarriba, Kroll, Sholl, & Rayner 1996; Dijkstra et al. 1999, for previous accounts of the frequency effect found for items presented in isolation, as well as in

[15] A strikingly similar result has been reported for the reactions to interlingual near-*homophones*, in a study investigating lexical competition during L2 listening (Chambers & Cooke 2009).

sentence context).[16] In Libben and Titone's study this variable was controlled in an attempt to increase the probability of obtaining evidence for non-selective access. More specifically, the authors selected mainly items that had higher frequencies in the (task-irrelevant) L1, which can be expected to have raised the global activation levels and measurable competition of the non-target language (2009:382). Under these circumstances, language-selective access *was* possible, but only for high constraint sentences, and only for late-stage comprehension measures (e.g. go-past time and total reading time[17]). In other words, while a semantically constraining context may considerably lower the activation threshold for representations in the target language, certain attributes of the input (in this case the stimulus frequency in the non-target language) may play a more immediate role during the early processing stages, by creating initial interference (see Duffy, Kambe, & Rayner 2001, for relevant monolingual data).

Perhaps due to the limited number of currently available studies discussing sentence context effects, there are several limitations which have not been addressed to date. One such limitation is that the language of the sentence context has typically been kept stable, i.e. does not diverge from the language of the task. As noted by Dijkstra, however, if sentence context affects the speed of word recognition, then bilinguals might be slower to recognise a stimulus in a language that differs from the language of the context sentence (2005:186; for previous research supporting this conjecture in the absence of sentence context, see, for instance, von Studnitz & Green 2002a). To test this proposition, in the research reported in this thesis participants completed two sessions: in one session the language of the sentence frame was consistent with the language of the lexical decision task (A in Figure 1 below; hence, an accentuated 'yes' response); in the other session, however, the language of the sentence frame differed from that of the lexical task (B). Note that the term "sentence context" has been used fairly inconsistently in previous studies (e.g. Elston-Güttler, Gunter, et al. 2005; Libben & Titone 2009; Schwartz & Kroll 2006) – for example, there have been differences in the amount of semantic information provided by a sentence, but also the target has been presented

[16] Both Schwartz and Kroll (2006) and Elston-Güttler, Gunter, et al. (2005) did not explicitly control for word frequency of their critical stimuli across the two languages. Their frequency measures included only calculations of the target language meaning, to match against the control word set.

[17] *Go-past time* refers to "the accumulated time from when a reader first fixated on a region [e.g. a critical word] until their first fixation to the right of the region", whereas *total reading time* refers to "the summed time of all fixations on a region" (Levy, Bicknell, Slattery, & Rayner 2009:21089).

both in final position or embedded in a larger sentence. For the present study, I will use the term 'sentence frame'. This should not only strongly imply that target word presentation was sentence final, but it will also position the current research against other studies (see also the *Materials* sections in the experiments reported below).

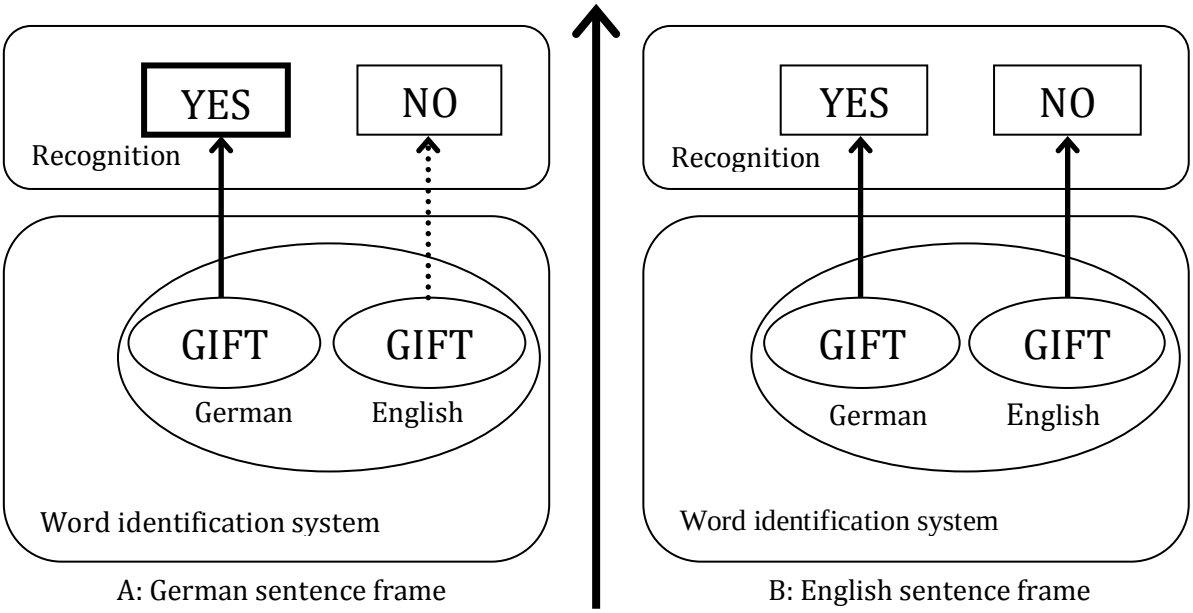


Figure 1. Stimulus-response bindings in a German lexical decision task (LDT), with the direction of a trial's sequential arrangement (black arrow in the middle). Simplified word and language membership representations are presented for the interlingual homograph GIFT. The panel on the left represents target presentation following a German sentence frame. With an expected reduced activation of the non-target language (English; indicated by the dotted arrow), the binding of the English reading of the IHG to the "no" response is weak and induces little interference. In contrast, a preceding English sentence frame leads to increased levels of non-target language activation (right panel). This in turn leads to a stronger activation of a "no" response and more detectable interference.

A related issue to be addressed in this study is that of language dominance. With all of the relevant (context) studies reported above focussing on bilingual processing in the L2, there is an obvious lack of studies addressing bilingual processing in the L1. Hence, the present study will compare the same task (i.e. German lexical decision task) being completed first by a bilingual English-German population in their L2 (see Experiment One), and then by a bilingual German-English population in their L1 (see Experiment Two).

3.1.3.2 Bottom-up information/Language-specific cues

A different type of language information is provided by language-specific orthography. A priori, we can assume that in the case of language pairs that have very distinct writing systems (e.g. Polish and Hebrew) the presence of a particular script provides sufficient

language information to affect language activation. However, to what extent does orthography play a role in language pairs that share the same (e.g. the roman) script? It is possible that in the absence of salient orthographic information and other stimulus-inherent cues (e.g. the relative frequency of an item), bilingual speakers simply turn to more subtle features, such as the relative occurrence of a particular letter sequence (e.g. bigram frequency) in either of the bilingual's languages. Another reliable source of language membership may be the more apparent (i.e. visibly noticeable) language-specific cues, such as diacritical markers in the form of accents in Spanish and French (cf. Mathey & Zagar 2000:200, for relevant monolingual data), or onset capitals for nouns in German. In principle, such language-specific cues may directly signal the language membership of an item[18] and affect lexical selection prior to actual word recognition. The question that remains to be answered is whether language cues are powerful enough to reduce lexical interference (i.e. the number of competitors) from the non-target language, or even to selectively access the target language only (Dijkstra 2003:20, 2005:187). This was the second major concern of the current study.

A number of researchers have addressed the question of language selection and cueing of language choice (e.g. Grainger & Beauvillain 1987; Orfanidou & Sumner 2005; Thomas & Allport 2000). However, the impact of sub-lexical information (i.e. in the form of language-specific cues) on such processes has been left relatively unexplored to date. Instead, the majority of the relevant research has focussed on the discussion of the *locus* of language switching or language choice (or, the "locus of control"), identifying two possible sources: an *internal* one (i.e. via the activation of units in the lexico-semantic system) or an *external* one (i.e. via a separate mechanism which coordinates responses to signals coming from the lexico-semantic system; see, for instance, Green 1998a; von Studnitz & Green 1997, 2002a). Because in most of these studies the question of language selection was addressed by means of the effects associated with language switching (and the associated switching costs), the stimulus lists in almost all studies were constructed by mixing items from both languages. In addition, a large number of previous studies have used experimental materials which were manipulated exclusively

[18] Clearly, depending on the diacritical marker, definite language membership information could also be subject to a number of other factors. If we take onset capitalisation, for instance, the question about which language is indicated by the diacritical marker (i.e. German or English) could additionally depend on a word's position in the sentence, whether it is a common noun or a proper noun, and so forth.

at the level of language-specific bigrams or trigrams (for an extensive review, see Meuter 2005; but also, see Kroll & Dijkstra 2002; Schwartz, Kroll, & Diaz 2000).

One study that sets itself apart to some degree (i.e. by not being primarily interested in the effects of language switching) was conducted by Vaid and Frenck-Mestre (2002). In this study, French-English bilingual speakers were presented with words that either contained or lacked bigrams that were specific to the speakers' L1 (French; e.g. VOEU vs. LOUP) or L2 (English; e.g. SNOW vs. DROP), respectively. Participants were asked to decide which of these two languages an item belonged to by pressing one of two response keys. Response latencies indicated faster processing of words that contained a language-specific bigram (i.e. orthographically marked words), particularly in the second language (English). The researchers interpreted this pattern of results as support for a perceptual search strategy (Vaid & Frenck-Mestre 2002:51–52). More specifically, they argued that orthographically marked words allowed participants to employ bottom-up cues by means of a redundancy check in their L2. That is, something that stood out as an L2 word (e.g. SNOW) was accepted as such, arguably even prior to its actual lexical identification. In contrast, responses to orthographically marked words in the L1 (e.g. VOEU) were considerably slower than those to orthographically marked L2 words, suggesting that for these L1 items an actual lexical identification may have taken place. In addition, these L1 words were recognised with a similar speed to words that did not contain language-specific bigrams in either language (i.e. orthographically unmarked words; LOUP and DROP), which was interpreted by the authors as further evidence that these items required an actual full identification. Finally, among the unmarked words the authors detected a tendency towards the 'classic' language-dominance effect, with items in the dominant language being recognised faster than items in the less dominant language.

While these results are in general agreement with the majority of the previous work (i.e. suggesting that bilinguals benefit from the availability of language-specific information), they do not clarify whether only the relevant lexicon is accessed in the process of identifying a stimulus. This is partly because the research design included stimulus material distinct to the bilinguals' L1 as well as L2, and because participants had to make a language (rather than a lexical) decision. As a result, reduction of lexical activation to a particular language does not seem possible. Appraisal of the role and impact of language-specific cues is further complicated by the fact that participants are likely to

have used different strategies to make their responses, i.e. *perceptual* versus a *lexical* search strategy (ibid.).

To address the limitations identified in the discussion of relevant work above, the design of the experiments reported in this thesis involved a number of important features. First and foremost, the present study limited its task requirements to a lexical decision in one language only (i.e. German). Similarly, the stimulus list contained less perceptibly mixed material: half of the words existed in one language only (German controls), and the other half were words which, even though they existed in both German and English (i.e. German-English interlingual homographs, IHGs), may have been read as German given the task setup and demands. There were no exclusively English words included. This stimulus list construction aimed to reduce the activation of the non-target language (i.e. English) for the two participating bilingual populations (i.e. English-German and German-English bilinguals, respectively).

Accordingly, the choice of language-specific cues for this study was directly linked to the language of the task (German) and the stimulus material (words, in a lexical decision task), i.e. accentuating the 'German-like' nature of a word in the form of onset capitals. For this reason, two identical sets of stimulus material were created – one set with all stimuli marked with the language-specific cue, and the other set containing stimuli whose presentation format was more ambiguous with regards to their language membership (i.e. German or English). This ambiguous format involved presentation of the stimuli for lexical decision all in lower case letters (see the methodology section for a more detailed description of the complete experimental material).[19]

To ensure that orthographic cues are employed in lexical access rather than their being a reflection of an orthography-based strategy (particularly for English-German bilinguals, completing the task in their L2), the current experimental design included *nonwords* which (depending on condition) carried a language-specific feature. The nonwords were not words in either language. Given such a setup, participants were not able to use onset capitalisation as a cue that a stimulus is a word in German (i.e. say "yes" whenever they

[19] As mentioned earlier, it is only nouns that carry an onset capital letter in German. Consequently, lower case is not exclusively a characteristic of English, but consistent with the orthographic patterns for many German words. Also, note that a third presentation format was considered for this study, namely all capital letters. Given its relatively infrequent appearance in writing, and also to avoid further complicating the already complex structure of the present study, this format was dropped. Future extensions of this study, however, may consider this format.

detect language-specific orthography). Instead, they will need to specifically access lexical information in order to make their response (cf. Thomas & Allport 2000:46).

As mentioned in the previous section (3.1.3.1), the visual presentation of a stimulus was preceded by an auditory sentence frame (prime). In one condition, the auditory sentence frame was congruent with the language-specific presentation format of the target item and the language of the lexical decision task (all German); in three other conditions, at least one of these three aspects was incongruent with the others (see the *Methodology* section for more detail). Assuming that bilingual speakers make use of language-specific cues, (relative) language selectivity[20] should be observed in a condition which fully supports the German reading of an item.
If 'language selectivity' is possible, then it should be most obvious in the responses to IHGs that carry a language-specific cue (onset capital), especially following a sentence frame (German sentence) which is congruent with both the language-specific cue and the language of the task (i.e. lexical decision in German; see A in Figure 2 below). In contrast, interference is expected in a condition which does not provide 'sufficient' language information (i.e. in the condition which provides a task-relevant sentence frame, but does not provide sub-lexical information; C), and even more so when the language of the preceding sentence frame differs from the language of the task (D). In this context, it is possible either that there will be an additive effect of these aspects of the task, or that one kind of language information is so powerful that changes in the other kind will have a negligible effect. Finally, it is important to note that the objective of the current study was not to investigate language switching costs. On the contrary, the main purpose of including a language-switching condition was to investigate how language-specific cues interact with other language information (i.e. a particular language of the sentence frame) and to what extent bilingual subjects can employ language-specific cues to resolve the issue of language membership ambiguity (B in Figure 2 below). This of course all presupposes that carrying out a lexical decision task in German required the resolution of such ambiguity (rather than just ignoring the English lexicon, as in a language-dependent view).

[20] The reader is reminded that given the large amount of research that provides evidence for the non-selective view of bilingual word processing, perhaps a better account of the processes involved might be reduced language interference.

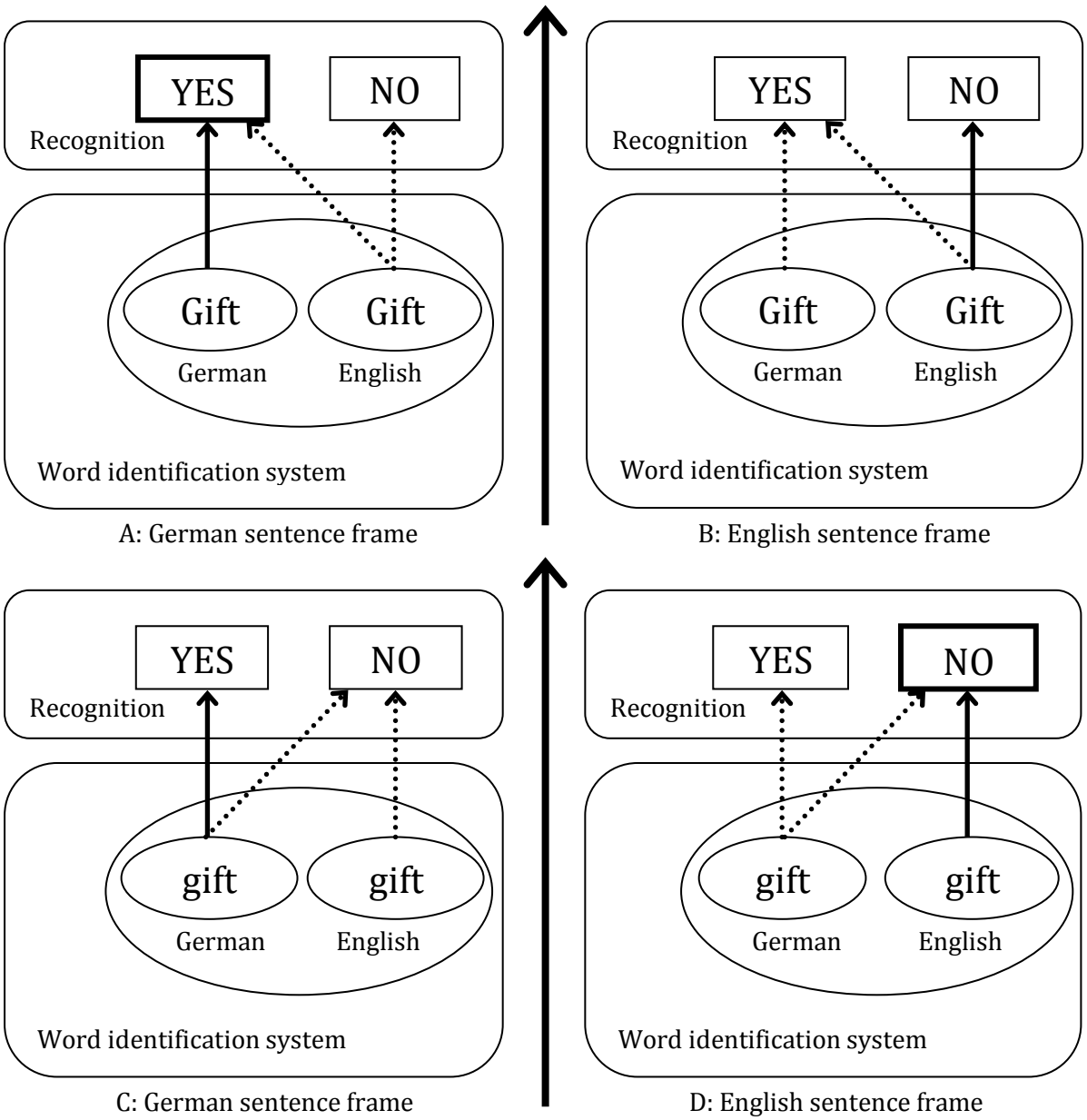


Figure 2.
Stimulus-response bindings in a German lexical decision task, with the direction of a trial's sequential arrangement (black arrows in the middle). Simplified word and language membership representations are presented for the interlingual homograph GIFT. Panels on the left (A and C) represent target presentation following a German sentence frame, which is expected to reduce activation of the non-target language (English). As a result, the binding of the English reading of the IHG to the "no" response is weak (dotted arrow) and induces little interference.[21] In contrast, a preceding English sentence frame leads to increased levels of non-target language activation (right panels; B and D). This in turn leads to a stronger "no" response binding (bold arrow) and more detectable interference. In addition, the availability of a language-specific cue (top pair of panels, A and B) reinforces the binding of a "yes" response; whereas the lack of it (bottom) strengthens the "no" response.

[21] Clearly, there is an alternative view of the scenarios presented in A and C, i.e. that the binding of the English word to the "no" response is just as strong, but that the activation flow along that binding is weaker because there is less activation of the English language node. Whichever may be the truer reflection of the investigated processes, this basic illustration was chosen over a complete one for the sake of simplicity.

3.1.4 Proficiency

A final variable which has been identified as an influential factor in bilingual visual word recognition but not widely investigated is a bilingual's proficiency in the less dominant L2 (e.g. Bijeljac-Babic, Biardeau, & Grainger 1997; Frenck-Mestre & Prince 1997; Kotz & Elston-Güttler 2004). Ever since the introduction of the RHM[22] (Kroll & Stewart 1994; see also Cheung & Chen 1998; Dufour & Kroll 1995, for a developmental interpretation of the RHM), substantial evidence has been brought forward to show that the links between formal representations in L2 and their connection to conceptual representations in L1 (via L1 equivalents) and/or L2 are formed differently depending on the relative proficiency of a bilingual speaker, with gradually stronger links with increasing L2 proficiency. When confronted with an L2 task, participants with varying proficiency in their L2 may therefore process information differently. While lower proficiency speakers are assumed to process an L2 item via their form equivalents in the L1 (resulting in major L1 interference), greater independence at the word form level for more proficient bilinguals is assumed to be reflected in less reliance on such links (and therefore reveal less L1 interference; e.g. Elston-Güttler, Paulmann, et al. 2005).

While some researchers have adopted the basic idea portrayed in the RHM to discuss a more general language dominance, or *automaticity*, issue (Jared & Kroll 2001; e.g. the common finding of increased cross-lingual competition when a task is performed in the less proficient L2; see, for instance, van Hell & Dijkstra 2002), others have attempted to apply it more specifically to an L2 speaker's proficiency level and to the question of *cognitive control* (e.g. Chambers & Cooke 2009; Elston-Güttler, Gunter, et al. 2005; Elston-Güttler, Paulmann, et al. 2005; Schulpen 2003; Schwartz & Kroll 2006).

As part of her doctoral dissertation, Schulpen investigated both cognitive control and automaticity in relation to the increasing proficiency level of her subjects[23], to shed light on the factors' apparently "paradoxical relationship with each other" (2003:66).[24] Her

[22] Although the Revised Hierarchical Model (RHM) has recently been criticised on a number of aspects (for the critical exchange, see Brysbaert & Duyck 2010; Kroll, van Hell, Tokowicz, & Green 2010), it is currently the only theoretical model that deals explicitly with the proficiency of a bilingual speaker.

[23] Proficiency in Schulpen's study was mainly addressed by recruiting subjects with differences in the amount and type of exposure to the L2 (English). The four groups of participants included students from two different high-school grades, university students and employees at a research institute. Although the results were not reported, Schulpen reports to have also collected proficiency scores based on a very basic (14 questions) language history questionnaire (2003).

[24] Schulpen notes that "increased cognitive control appears to imply a decrease in [automaticity], and vice versa. This is, of course, in apparent contradiction with the often made assumption that both control and automaticity increase with an increase in proficiency" (2003:66).

Dutch L1 and English L2 bilingual participants completed three different lexical decision tasks: an exclusive-English lexical decision task (LDT; requiring a 'yes' response to English items and a 'no' response to nonwords); a mixed-English LDT, which included Dutch distractors (requiring a 'yes' response to English items and a 'no' response to Dutch words and to stimuli that were not words in either language); and an inclusive-English/Dutch LDT (requiring a 'yes' response to English and Dutch items and a 'no' response to nonwords). Although Schulpen's data provided support for the idea of increasing automaticity with increasing proficiency level, as measured using Segalowitz & Segalowitz's (1993) *Coefficient of Variability* (CoV), other aspects of her study were left unexplained by this type of measurement (i.e. CoV). Similarly, the effects of cognitive control were only identifiable in the pure-language (i.e. exclusive-English) task, with more proficient participants being better able to suppress the non-target language (as observed in responses to IHGs). The author questions the validity of cognitive control as a prominent factor modulating the relative state of activation of a particular language, particularly under "non-linguistic circumstances" (i.e. effects that arise from instruction, participant expectations, et cetera; Schulpen 2003:180). Instead, she argues that in light of the prevailing absence of cognitive control in 'mixed' tasks (i.e. when items from both languages were present), the obtained results are due to post-lexical processes (i.e. at the task/decision level), as reflected in less effective "stimulus-response binding" for less proficient speakers (ibid.).

So what happens to *cognitive control* under 'linguistic circumstances', i.e. when additional (linguistic) information is provided? One recent study which has addressed this question to a certain extent was conducted by Chambers and Cooke (2009). Investigating the impact of sentence context on *spoken* language processing, the authors argued that it is the amount of information provided by a sentence that can reduce measurable interference from the non-target language (i.e. exert *control*), rather than a speaker's proficiency level. In their study, English native speakers with varying proficiency levels in French looked at a visual display depicting four objects while listening to sentences in their L2 (i.e. French), such as:

> Marie va décrire la poule (= Marie will describe the chicken).

On critical trials (such as the sentence presented above), the visual display included a picture of the final noun target (e.g. 'chicken') and that of an interlingual near-

homophone (e.g. 'pool') whose name in English has a similar pronunciation to the French target word ('poule'). The researchers measured listeners' eye movements during target noun playback. One observation resulting from this experiment was that there was temporary lexical competition for interlingual homophones (see also Vandeberg, Guadalupe, & Zwaan 2011). The same pattern was reported for lower as well as higher proficiency listeners, and was particularly robust in low constraint sentences (i.e. sentences where there is no clear bias towards either meaning of an interlingual homophone).

These findings are in line with some comparable *visual* bilingual research (e.g. van Hell & de Groot 2008; Schwartz & Kroll 2006, for responses to cognates), but not with other such research (e.g. Schwartz & Kroll 2006, for responses to interlingual homographs). Among the factors which may have led to differences across the available studies are limits on the number of participants, the nature of the experimental stimuli and the task itself. For instance, it is possible that a task in which the language material is presented only auditorily had a different influence on (increasingly highly proficient) bilingual lexical processing.[25]

Despite invaluable first insights from these studies, it is clear that more research is needed in order to arrive at some definitive answers. Hence, the present study aimed to shed further light on the issue of language membership determination and how it influences the relative activation of one of a bilingual's two languages. This was done by manipulating the available linguistic information in the form of top-down (language of the sentence frame) as well as bottom-up cues (orthographic language marker). In this context, the study tested what effects proficiency might have on the manifestation of facilitatory versus inhibitory dynamics in the bilingual system.

As mentioned above, all aspects of the study were addressed not only from the perspective of a bilingual speaker conducting a task in their L2, but also vice versa. This

[25] As pointed out by the authors, an important question here concerns the amount of lexical competition from the non-target language (L1) that can arise during spoken (L2) word recognition (Chambers & Cooke 2009:1038). Recent findings suggest not only that bilingual spoken word recognition is non-selective in nature (e.g. Schulpen, Dijkstra, Schriefers, & Hasper 2003), but also that auditory L2 processing may be most strongly affected at the "lexical level" (i.e. above the phonetic level but below the contextual level Broersma & Cutler 2008:31). It has been suggested that L2 listeners may exert more cognitive control the larger their L2 vocabulary (ibid.). If we relate this conjecture to Chambers and Cooke's study (2009), we notice that the authors employed merely a self-reported proficiency questionnaire to measure their participants' proficiency level. For this reason, the lack of a proficiency effect may be simply a result of insufficient language proficiency testing, rather than reflecting no differences in cognitive control.

was done to obtain a fuller picture of the processes underlying bilingual visual word recognition.

The remainder of this chapter is as follows: the first section (Section 3.2, Experiment One) will discuss the study involving English-German bilinguals, conducting the task in their L2. This section will then be followed by the same study (Section 3.3, Experiment Two) completed by German-English bilinguals, i.e. conducting the task in their L1. The chapter concludes with a general discussion (Section 3.4), relating the empirical data to current models of bilingual visual processing.

3.2 Experiment One: L2 Lexical Decision Task (English-German Bilinguals)

3.2.1 Methodology

3.2.1.1 Participants

Sixty-five native speakers of English took part in this experiment. Participants were mostly current and former students of Victoria University of Wellington**,** New Zealand, with varying knowledge of German. They all had normal or corrected-to-normal vision and no known hearing impairment. Participants were selected to represent three levels of proficiency, based in most cases on their course enrolments. To test their German knowledge individually and in order to acquire other relevant information, I asked each participant to fill out a language questionnaire and to complete a German language proficiency test at the end of the experiment (for detailed discussion of both assessment types, see section 3.2.1.4). Note that although course enrolment information was used to make sure that a range of speakers with varying L2 proficiency were represented in the study, the critical information for the analysis came from the proficiency scores collected from the German language proficiency test.

In accordance with human ethics requirements, all participants received information sheets and signed a written consent form. Participants received a gift voucher in recognition of their participation.

3.2.1.2 Materials

The experimental design required careful selection and preparation of all stimulus material involved. This included selecting critical target words (German-English interlingual homographs, referred to below as IHGs), as well as matched German control words (controls) and nonwords (NWs), and then designing sentence frames (i.e. language primes) into which these items could be placed (see Table 1). In a final preparation stage, two practice stimulus sets were designed, one for each session. Each preparatory stage will be described in detail as follows.

Table 1.

An overview of the design, number of items and sentence frames.

first experimental session			second experimental session			
words		nonwords	words		nonwords	
IHG	German controls		IHG	German controls		
39	39	78	same as previous session	39	78	Grand total (items): 273
sentence frames per session and item type						
39	39	78	39	39	78	Grand total (frames): 312

Selecting target words

Item construction was done in the following way. First, a preliminary list of interlingual homographs was created, which was partly based on Elston-Güttler, Gunter, et al.'s (2005) item list and partly extracted from an English learners' dictionary (*Cambridge international dictionary of English*, 1999). To ensure that lower proficiency learners of German would be familiar with the test items, this preliminary list was matched against the vocabulary list from two elementary German learners' course books (Albrecht et al. 2005; Perlmann-Balme & Kiefer 2004) provided by a German course instructor at Victoria University of Wellington. Meeting this criterion left 39 items, all of which had one meaning in English (cf. *tag* = "label") and a different meaning in German (cf. *Tag* = "day"). Appendix C contains a complete list of the 39 target words. The majority of the selected items were nouns in both languages. In a few cases, however, an IHG that was a noun in one language belonged to a different word class in the other, and in very few cases an item was a noun in neither language (commonly being an adjective, verb or adverb; usually varying between the two languages).[26]

Note that each critical IHG was presented twice in the course of the experiment, in different sessions (see further information on the experimental design below). To provide real word controls (which acted as non-IHG fillers at the same time), a pair of German control words (controls) was selected for each IHG. For instance, the English-German homograph DOSE (the German word means 'tin') was matched with the controls *Mehl* ('flour') and *Kamm* ('comb'), respectively. These controls were selected using the WordGen programme (Duyck, Desmet, Verbeke, & Brysbaert 2004), which uses the CELEX database (Baayen, Piepenbrock, & van Rijn 1993) as a resource. The set of control items was matched with the set of IHGs for number of letters, number of German neighbours, and German log frequency per million (see Table 2 and Appendix C). To match the critical stimulus set as closely as possible, the control words were

[26] The selection of IHGs in the present study included 29 items which could have been read as nouns in German and 10 which belonged to a different syntactic category. In addition, 12 IHGs shared a very similar pronunciation across the two languages and 27 IHGs were pronounced differently. Note that the categorisation of the similarity of pronunciation reflects an impressionistic analysis by the German-English bilingual author, rather than being an objective measure of similarity. Interestingly, because studies which controlled for phonological similarity in IHGs obtained highly contrasting results (e.g. Lemhöfer & Dijkstra 2004; von Studnitz & Green 2002a), Haigh and Jared have recently cast doubt on the relevance of such a measure in a lexical decision task (2007:625–626; see also Titone, Libben, Mercier, Whitford, & Pivneva 2011, for the most recent study showing no effects of phonological overlap). For this reason (as well as not being the primary subject of interest in this study), the relative similarity of pronunciation was not used in the selection of materials or later in the analysis.

mainly nouns, but also included verbs, adjectives and adverbs. Different members of each control pair appeared in the two experimental sessions.

Each of the two sessions therefore contained 78 real words (39 IHGs and 39 controls). To balance the number of real word ("yes") and nonword ("no") responses expected from participants (assuming they knew the German words), each session also contained 78 nonwords (NWs). A different set of nonwords was used in each session, meaning that 156 nonwords had to be created altogether. Once more, these NWs were created using WordGen (Duyck, Desmet, et al. 2004) and CELEX (Baayen et al. 1993), and were matched to the real word items (both, IHGs and controls) for number of letters, number of German neighbours, and German bigram frequency (see Table 2 and Appendix C). Care was also taken to ensure that all nonsense words obeyed German orthographic rules and were not existing English words. To illustrate, for the homograph HOSE (the German word means 'trousers') the German control word *Tanz* (meaning 'dance') was selected and the NWs *sarn* and *bekt* were created, matching the number of "yes" and "no" responses in the first session. In the second session, the IHG was matched with the German control word *Topf* (meaning 'pot') and the NWs *dand* and *lalb*. Ideally, all stimuli would also obey English orthographic rules. However, given that the present task was not language selection but lexical decision, I doubt that the effects of a perceptual search strategy (i.e. participants perform a redundancy check in their L1) observed by Vaid & Frenck-Mestre (2002) would be present here.

Table 2.

Mean letter length, count of German neighbours, and frequency of different target item types, with standard deviations in parentheses. *Note*. NB=neighbourhood; freq=frequency.

Target Items	Target letter length	Target *German* NB count	Target *English* NB count	Target *German* log [a] freq	Target *English* log freq	Target *German* bigram [b] frequency	Target *English* bigram frequency
IHGs (N=39)	4.44 (1.02)	4.26 (2.41)	9.03 (6.35)	1.47 (0.81)	1.46 (0.69)	14978 (10803)	8943 (5124)
Controls (N=78)	4.48 (1.04)	4.26 (2.38)	4.26 (4.25)	1.50 (0.72)	N/A	14037 (10247)	6727 (4368)
			3.44 (4.36)	1.50 (0.83)	N/A	12725 (10575)	6277 (3689)
Non-words (N=156)	4.48 (1.04)	4.26 (2.38)	4.31 (4.48) / 4.00 (5.04)	N/A	N/A	14735 [c] (10359) / 14821 [c] (10498)	6982 (4649) / 7247 (4048)
			4.36 (4.57) / 4.23 (5.23)			14410 [d] (10136) / 13017 [d] (9194)	7283 (4705) / 7351 (4320)

[a] Mean frequency per million of test and corresponding control targets, using the German log frequency in the CELEX database (Baayen et al. 1993). The matched value of log frequency for the English IHG entries was 1.46 (0.69).

[b] Mean frequency per million of test and corresponding control targets, using the German bigram frequency in the CELEX database (Baayen et al. 1993). English bigram frequency measures were substantially lower (usually smaller than half the German value). If anything, this should have underscored the German nature of the task. All measurements are listed in Appendix C

[c] Bigram frequency measure of NWs used in the two experimental sessions, and matched to the German bigram frequency of IHGs as closely as possible.

[d] Bigram frequency measure of NWs, matched to the German bigram frequency of controls used in the two separate experimental sessions, respectively.

Note 1. WordGen (Duyck, Desmet, et al. 2004) calculates word frequency using the logarithm (base 10) of the written frequency per million words in a given corpus. The main reason for this rescaling is that it corrects for the fact that the difference between two words occurring respectively 1 and 10 times per million is not the same as the difference between two words occurring respectively 101 and 110 times per million. In addition, this type of measure also corrects for the fact that the databases used for each language (i.e. German and English) contain a slightly different number of items. Finally, WordGen's (Duyck, Desmet, et al. 2004) word frequency is based on the combined measure of the different *lemma* frequencies. This means that the frequency of the word 'book' includes the frequency of all associated English word forms, i.e. the noun and the verb. In addition, lemma frequency calculations report merely the instances of base forms (e.g. 'house', not 'houses', and 'go', not 'goes') but include these morphologically related word forms in the same lemma entry (see Duyck, Desmet, et al. 2004, for elaboration).

Note 2. Bigrams are the adjacent letter pairs of an item (including its word edge). For instance, the word 'gift' consists of five bigrams: *#g*, *gi*, *if*, *ft* and *t#*. Duyck, Desmet, et al.'s WordGen (2004) tool computes bigram frequency by calculating the number of times a bigram appears in a respective database, however, irrespective of its position in items. To obtain the bigram frequency of a word or nonword, WordGen (Duyck, Desmet, et al. 2004) summates the respective position-nonspecific bigram frequencies of an item. The reason why a bigram frequency measurement is included is because it provides the only means by which the characteristics of existing words can be matched with nonwords.

Sentence frames

With stimulus selection completed, two sets of German sentence frames (incomplete fragments) were created for each IHG[27], as well as one for each control word and one for each NW. This gave a total of 312 German sentence frames. Each German sentence frame was then translated into English by the German-English bilingual investigator, resulting in 312 English sentence frames. (The English sentences were further checked by an English native speaker.) The semantic context provided by each sentence frame was relatively open with no obvious bias towards the target word meaning (as perceived by two native German speakers, including the investigator; again, no calculable measurements were taken). Hence, the visual stimulus was primarily an appropriate syntactic continuation of the auditory sentence fragment.[28] With respect to the critical IHG stimuli, this means that sentence frames were created with the primary aim of not allowing a bias towards either (English or German) meaning. If this goal could not be achieved (given rigid syntactic and semantic constraints across the two languages), then the sentence frames were biased as similarly as possible towards both meanings. For instance, the sentence fragment including its target word: *He thought of the cellar as the best hiding place for the gift*, was biased more strongly towards neither the English meaning of the target word ('present'), nor its German meaning ('poison').

Finally, all English sentence frames were recorded by a native speaker of English, and all German sentences were recorded by a native speaker of German. Two presentation lists were constructed, each containing all 312 target words or nonwords. In each presentation list half the sentence frames were in English, and half in German, rotated across lists so that if a target IHG, control or NW was preceded by an English sentence in one list then it was preceded by a German sentence in the other list. Within each presentation list, the targets with English and German sentence frames were presented

[27] The reader is reminded that the chosen critical items for this study (IHGs) were repeated in the second session of the experiment (i.e. approximately seven days after completion of the first experimental session). To reduce the amount of repetition across the two experimental sessions and limit potential recollection effects, two separate sentences were created for each IHG; one to appear in each session (hence, the difference between the number of constructed items [273] and the number of sentence frames [312; excluding their translations]; see also Table 1 above). Each sentence pair was assessed by two bilingual speakers (one of whom was the investigator herself), to match one another for difficulty and the extent to which they constrained the target word. The assessment was conducted by way of an individual impression (and instant adjustment) rather than a calculable measure.

[28] Despite all efforts to make sentence frames as neutral as possible, it could not be avoided that some targets had a stronger semantic and/or collocational link to the preceding words or content of the sentence frames. Again, impressionistic comparisons by two German-English bilinguals were taken as a measure of uniformity of the experimental stimulus set.

in separate sublists (i.e. a sublist of 156 English sentences and a sublist of 156 German sentences). These sublists were presented in separate experimental sessions one week apart. Half of the participants were exposed to the sublist with the English sentence frames in the first week, and to the sublist with the German sentence frames in the second week. This order of sublists was reversed for the other participants. This ensured that participants never heard the same sentence in both languages in a single session, or twice in the same language across the two sessions (see Table 11, Appendix C). The two sublists presented to any participant included the same set of 39 critical IHG words. Each sublist had a different set of 39 matching controls, and a different (but matching) set of 78 NWs. All sentence frames included were unique across the two sublists (except that for any one sentence frame there was a translation equivalent of that sentence frame in the other sublist). The 156 stimuli in each sublist were divided in six blocks, each containing 26 trials. Each block ended with a memory task (explained below) which was meant to ensure that subjects paid attention to the sentences and did not exclusively focus on the lexical decision task. The order of the six blocks was kept constant because of the limitations outlined in the following paragraph.

Stimulus order within the sessions was kept constant, so that effects of sequential order within a session (e.g. practice or fatigue effects) would be likely to affect each language condition equally. (Note though that the statistical approach taken in the analysis of the results includes sequential position as a factor, as an additional control for sequence effects.)

Participants recruited in the different classes were allocated randomly but evenly to one of two format conditions, which related to the presentation format of the visually-presented target word (hence, Format was a between-participant factor). The target was either entirely in lower case, or with the first letter capitalised (referred to as 'Title case' throughout all analyses). For example, after the spoken fragment *The woman listened to a radio show about the perfect* the visual target for the lexical decision task would be either *gift* (lower case) or *Gift* (Title case). The target format remained constant across both sessions for each participant. Table 11 in Appendix C illustrates the resulting conditions.

Practice session material

The experimental items were preceded by practice items. The construction of practice stimuli followed the same procedures as those described above; that is, employing the same word and nonword generating criteria and tools. Nine German words and nine nonwords were created which appeared nowhere in the experiment and had no orthographic or phonological resemblance with English words. Subsequently, eighteen low constraint English sentence fragments were constructed, with no bias towards the target item. The sentence fragments were then translated into German and the two versions were recorded by native speakers (i.e. together with the actual test material discussed above). Half of the practice stimuli were used at the beginning of the first experimental session, and the other half at the beginning of the second experimental session. The order of appearance of practice items was pseudo-randomised but constant across conditions and participants.

The practice session allowed participants to get familiar with the task. During the practice session participants received feedback on the accuracy and speed of their responses. This prepared all participants equally in terms of the responses required irrespective of the stimulus' presentation format. That is, in some cases participants were required to accept stimuli as correct German words, even though they lacked the capital initial letter in the case of German nouns (cf. lower case condition), and in other cases they had to say "yes" to words even though they carried a superfluous capital letter (cf. Title case condition). None of the participants commented on the presentation format prior to the experiment, but when explicitly asked about this after the second session, participants mentioned the practice session as a suitable clarification source.

3.2.1.3 Procedure

Participants were required to attend two (approximately) 30 minute sessions which were usually seven days apart**.** This time frame was chosen to minimise possible learning effects resulting from prior exposure to the interlingual homographs (see discussion above). Participants were seated in a spacious office, with standard lighting, approximately 60 cm away from a monitor (Dell UltraSharp 1908FP 19-inch Flat Panel Monitor). They were first asked to read all instructions for the experiment on the computer screen. During the experiment participants heard the auditory sentence primes presented in mono over closed-ear headphones (*Sony Professional*, dynamic stereo headphones MDR-7506). The main test followed a short practice session. At the

beginning of each trial, a blank screen was presented for 2500ms. The participants then heard a sentence fragment over the headphones. Immediately after the sentence fragment a fixation cross appeared on the empty screen. After 200ms, the fixation cross was replaced by the potential sentence-final word, i.e. the target, either in all lower case letters or with the first letter capitalised.

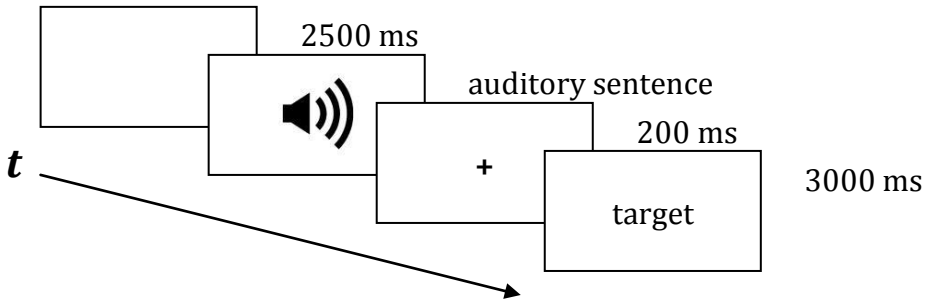


Figure 3. Schematic of an experimental trial in Study One (Experiments One and Two).

The participants' task was to decide as quickly and accurately as possible whether the word presented on screen was an existing German word, and to indicate this response by pressing one of two keys (labelled Yes and No) on a button box with millisecond timing accuracy. They were timed-out after 3000ms if they had made no response, and the next trial was started (see Figure 3 for illustration). The experiment was run in E-Prime (Schneider, Eschman, & Zuccolotto 2002a, 2002b) on a Windows personal computer (Dell OptiPlex 755 Desktop, Intel® vPro™ processor technology). Stimuli appeared in black Courier New (18 point font size) at the centre of the computer screen on a white background. Different response button configurations were selected depending on whether the participant was left- or right-handed, so that every participant used their dominant hand to indicate a 'Yes' response. Between trials, participants rested the index finger of each hand over the response buttons. Participants were tested individually. To keep the entire experiment as stable as possible, the same researcher (the author) conducted all sessions, and the procedure was exactly the same for all participants (except for the different conditions, of course). Communication between researcher and participant was limited to English.[29] The lexical

[29] Notably, participants may have noticed that the researcher was a German native speaker (or at least not a native speaker of New Zealand English). Such experimenter effects have been previously shown to affect participants' expectations regarding the task and their performance in the course of an experiment (Hay, Drager, & Warren 2009). This is not a major concern in the current setting, given that participants were performing a German lexical decision. So, if anything, conceivably elevated German activation levels should have facilitated the execution of the task.

decision task lasted no more than 25 minutes and was presented in six blocks, as described above. At the end of each block, a memory recall task was performed which included three sentences that were previously heard over the headphones in the preceding block and three sentences that were not heard anywhere during the experiment. Participants were presented with these sentences on screen, including their final word, and were asked to decide whether each combination of sentence and final word was included in the block they had just been exposed to. This was done to ensure that subjects paid attention to the sentences and did not exclusively focus on the lexical decision task.[30]

The participants involved in this experiment had varying knowledge of their second language (here, German). After the second experimental session, participants carried out a German proficiency test (Lemhöfer & Broersma 2011; Lemhöfer et al. 2004), and filled in a language history questionnaire (adapted from Li, Sepanski, & Zhao 2006), and were asked to give the English meanings of the German words represented by the IHGs in the experiment (e.g. for HOSE a correct response would be 'trousers' or 'pants'). Both assessment methods are described in further detail below. The entire experimental procedure, that is both sessions, was completed in approximately 60 minutes (roughly 30 minutes per session).

3.2.1.4 Language proficiency assessments

German proficiency test

To obtain an *objective* measure of the participants' second language proficiency level, I drew on an on-line proficiency test developed and employed by Lemhöfer and colleagues (2004). This test had been adapted from an unpublished English proficiency test by Meara and colleagues (Meara 1996). Participants responded to items that were presented on a computer screen for an unlimited amount of time, i.e. until the participants had made a response. The participants' task was to decide whether the stimuli were existing words in German, and to indicate this response by pressing one of two keys (labelled Yes and No) on a button box. Participants' responses were recorded for accuracy but not for response latencies. Hence, the test was a vocabulary test in the form of a non-speeded lexical decision task. The order of items was the same across all

[30] Given that this was the purpose of the recall task, data from that task were not included in any of the subsequent analyses.

participants. The proficiency test was run in E-Prime (Schneider et al. 2002a, 2002b) on a Windows computer, which had previously been used for the main experiment. Stimuli appeared all in upper case, black Courier New (18 point font size) at the centre of the computer screen on a white background. Different response button configurations were selected depending on whether the participant was left- or right-handed, so that every participant used their dominant hand to indicate a 'Yes' response. Between trials participants rested the index finger of each hand over the response buttons. In the following sections, the design of the test materials is discussed in more detail.

To create the German proficiency test, Lemhöfer and colleagues (2004) first had to develop a shortened version of Meara's English proficiency test (1996), which would later serve as a template for the German test. Accordingly, the English test template comprised 60 from Meara's (1996) original 240 English items. Following Meara's test design (1996), the authors selected both words and nonwords, which were distributed in an unequal manner; that is, two-thirds of the selected items were English words, and one-third were English nonwords. Consistent with Meara's original test (1996), the researchers chose not to use an equal number of words and nonwords because of the general difficulty of the test. Meara's (1996) test was developed for highly proficient language speakers, assuming that participants knew about 75% of the included words. Given that assumption, the proportion of known and unknown words in the test should have therefore been approximately the same (Lemhöfer et al. 2004; Meara 1996). As reported by Lemhöfer and colleagues (2004), the selected items for the English test template were between 4 and 12 letters long (mean stimulus length: 7.3) and the calculated frequency of the word stimuli ranged from 1 to 26 occurrences per million (mean word frequency: 6.4 o.p.m.). The latter calculation was based on information in the CELEX database (Baayen et al. 1993).

Next, Lemhöfer and her colleagues (2004) created a set of German items, which was designed to match the English items as closely as possible. As reported by the researchers, the main matching criteria for including items in the German test included item length, number of syllables, word frequency, and syntactic class. In addition, the researchers attempted to reflect the morphological characteristics of the English items in their German selection (Lemhöfer et al. 2004). For instance, adjectives which take the suffix '-ly' in English were matched with adjectives in German which take the suffix '-ig'

or '-lich'; English words which take the suffix '-ing' were matched with German words ending with '-ung', and so on (Lemhöfer et al. 2004). The nonword material subsequently constructed by Lemhöfer and colleagues (2004) had not only closely matching morphological structures, but were also highly word-like in German. All items selected by Lemhöfer and colleagues (2004) were employed in the current research and are listed in Appendix B.

As mentioned above, the proficiency test was simply a vocabulary test because it required participants to decide whether a presented string of letters formed a correct German word or not, by pressing one of two buttons. Unlike most vocabulary tests, participants could take as much time for their responses as they wished. In addition, participants were instructed to respond 'yes' to items only when they were certain that the presented item was a German word. In any other case participants were asked to press the 'no' button. The latter instructions in particular were crucial for calculating the test performance score.

In accordance with previous research, the test performance score can be calculated in two ways: as a percentage correct measure or as the measure Δ*M* (Lemhöfer et al. 2004; Meara 1996). As Lemhöfer et al. (2004) point out, the percentage correct measure needs to be corrected for the unequal number of words and nonwords by taking into account the mean percentage of correctly recognised words and correctly rejected nonwords. Δ*M*, employed in Meara's test (1996), reflects the proportion of words within the given frequency range known by a participant (Lemhöfer et al. 2004). The value of Δ*M* is expected to lie between 0 and 1 for a highly proficient speaker. Apart from the proportion of correct answers, the Δ*M* value is also based on the proportion of 'false alarms' (ibid.). A 'false alarm' is an incorrect acceptance of a nonword, reflecting participants' guessing (Lemhöfer et al. 2004). Too many 'false alarms' result in a negative ΔM and indicate that the proficiency level is very low, and not *precisely* measurable by the test. The precise formula for Δ*M* is shown in Appendix B.

One concern that I had in developing the experimental task was that the participant group in New Zealand may perform relatively poorly, given certain constraints, compared with the German learners of English tested in Experiment Two below. One such constraint is the immediate exposure to the second language (i.e. German), which is relatively limited due to the geographical location of New Zealand. Another constraint

may be the length of study (i.e. relatively few years of instruction), since German as a second language is not introduced as early as, for instance, English in other countries. Finally, a factor that could also have had an impact on participants' performance was their incentive to learn the L2. In other words, whereas there is pressure and strong desire to perform well in a language like English (i.e. in the international language), such factors do not necessarily play a role while learning German as an L2. This difference in incentives may have been reflected in participants' language level and performance. Taking the issues outlined above into consideration, I expected a lower level of performance from the English native speakers in the German L2 proficiency test than for the German L1 learners of English tested in Experiment Two, and for this to be reflected in a higher number of negative values for ΔM. Such a result, as suggested by Meara (1996), would indicate that proficiency is at a less precisely measurable level. This is not a major concern, mainly because the primary aim of this proficiency test was *not* to provide an exact measure of the participants' proficiency level but rather to indicate the range of performances at different learning stages. I believe that I have achieved this goal by employing the above mentioned tests. Further, I would like to point out the current lack of objective proficiency tests which are available for speakers of different language backgrounds. Lemhöfer et al.'s (2004) tests provided me with the best objective testing option. Finally, as well as assessing participants' L2 proficiency by employing an objective method, I also collected additional information from (most) participants related to their language background (see below).

Language history questionnaire

After completing the language proficiency test, participants completed an off-line, *subjective* paper and pencil language history questionnaire (adapted from Li et al. 2006). The questionnaire consisted of two parts. First, participants gave general information about their background and language history (for instance, age, sex, native language, other languages spoken). In the second section, participants reported more specifically on their language environment, usage and habits (for example, age of L2 acquisition, frequency and instances of language usage and language mixing). In this section participants also rated their proficiency in both German and English (and other known languages) along four dimensions (reading, writing, speaking, speech comprehension) on a scale from 1 to 6, with 6 indicating highest proficiency. Finally, participants also indicated any linguistic experiences, such as time spent in foreign countries, formal

language test scores (if available) and other relevant comments. The questionnaire was provided in English and is attached as Appendix B.

Even though two proficiency measures were collected from participants, unexpected experimental limitations (i.e. a considerable number of language history questionnaires were not returned to the researcher; see Experiment Two and Experiment Four, for more detail) allowed me to use only the former of the two, namely the on-line proficiency test. This was not a major concern, since the original purpose of the questionnaire was merely as a backup for inconclusive results. For this reason, the questionnaire data was not used in any of the subsequent statistical analyses.

3.2.2 Statistical analysis

3.2.2.1 The nature of experimental designs

Traditionally, researchers have opted for *factorial designs* with a limited number of treatments and items that had to be carefully selected and controlled for across a range of factors. In psycholinguistic research which relies heavily on response latencies as a primary source of information, some of the earlier studies focussed on a single variable (e.g. frequency of an item, length of an item, neighbourhood density) which was manipulated systematically across two conditions. As pointed out by Balling (2008; see also Baayen 2010), there are several problems with this design. First, we know from previous research that multiple variables affect lexical processing at the same time, so that it becomes challenging (if not impossible) to control for all these factors in one experimental design (cf. Cutler 1981). Second, many of the variables of interest are typically continuous in nature. This means that in order to measure the effect of a particular predictor, researchers typically dichotomise that variable (e.g. to test for a frequency effect, a researcher would divide the data set into items of high frequency on the one hand, and items of low frequency on the other hand). Dichotomisation of continuous variables, however, has been argued to lead to a severe loss of statistical power and to yield misleading results (e.g. Cohen 1983; MacCallum, Zhang, Preacher, & Rucker 2002). Finally, another problem which many researchers may face is a substantial loss of data and the lack of crucial observations due to an "aggressive screening for outliers" (Baayen & Milin 2010:8). The challenging loss of data is to some extent associated with the type of statistical analysis typically conducted in a factorial

design, namely the analysis of variance (ANOVA), in which data is usually averaged over items and/or subjects.

For the reasons outlined above (and others) an increasing number of researchers have recently started to use *regression designs* (e.g. Baayen, Davidson, & Bates 2008; Balota, Cortese, Sergent-Marshall, Spieler, & Yap 2004; Dijkstra et al. 2010; Jaeger 2008). A clear advantage of this type of statistical design is that it significantly reduces the sources of "statistical noise" (Balling 2008:179). As mentioned previously, one such source of noise lies within the characteristics of the experimental stimuli (e.g. in regression designs, the issue of dichotomisation can be avoided). Particularly in light of bilingual research, where stimuli have contrasting characteristics across two languages, this type of design and subsequent analysis seem beneficial.

Other factors which are impossible to address in factorial designs have to do with changes in the course of the experiment (e.g. Baayen & Milin 2010; Balling 2008). For instance, the speed, accuracy and type (word versus nonword) of a response have all been shown to have an impact on a subsequent response (e.g. Balling & Baayen 2008; de Vaan, Schreuder, & Baayen 2007; Kuperman, Bertram, & Baayen 2008). Other factors, such as the level of practice (or learning) and fatigue have similarly been identified as influential in the course of an experiment.

Finally, another source of statistical noise may be certain characteristics associated with the participants, such as age or handedness (e.g. Baayen, Feldman, & Schreuder 2006; Boulinguez, Barthélémy, & Debu 2000), and in bilingual research, the proficiency level in an L2 (e.g. Sunderman & Kroll 2006).

From this short description, it becomes clear that regression designs - which can accommodate the types of factor listed above - offer a more flexible and a more informative experimental approach for the majority of experimental setups (for a statistically driven discussion and suitable exceptions, see Baayen 2008, 2010). In the following section, I will briefly outline the regression type chosen for the present analysis, pointing to its advantages and outlining potential pitfalls.

3.2.2.2 Linear mixed-effects modelling[31]

There are a number of different statistical models that involve a regression design (see, for instance, Baayen 2008:269–278, for a comprehensive discussion and useful comparisons). An analysis technique which has recently been gaining wide recognition among scholars is linear mixed-effects modelling (e.g. Baayen et al. 2008; Jaeger 2008; Pinheiro & Bates 2000). In their seminal paper, Baayen and colleagues (2008) outline various advantages of mixed-effects modelling. Some of the key advantages will be discussed below, as they are directly relevant to the present study (for other advantages and a more detailed discussion, see Baayen et al. 2008; Jaeger 2008).

One of the key features of mixed-effects modelling is the inclusion of all individual data points. This means that, unlike in some uses of analysis of variance, the potentially harmful averaging[32] by participants and by items is avoided. Instead, both variables (i.e. 'participant' and 'item') are typically included in the (psycholinguistic) analysis as *random effects*. The inclusion as a random effect reflects the fact that, in principle, both these variables are not repeatable (i.e. there are no two items or subjects of the same kind), they do not have a fixed number of levels (i.e. the hypothetical number of items or participants is infinite)[33], and they are sampled randomly from the relevant populations. In contrast, predictors defined as *fixed effects* in the statistical model are either factors or co-variates. While co-variates are continuous (i.e. numerical) variables (e.g. length or frequency of an item), factors are categorical variables with a limited number of levels which exhaust the levels in the sampled population (e.g. session number or word class of an item).[34] This terminology is equivalent to the distinction between continuous and categorical independent variables in ANOVA (cf. Weiner, Freedheim, Schinka, & Velicer 2003:139). A distinctive characteristic of fixed effects is that they are repeatable (i.e. two

[31] The term includes analysis tools for both numerically distributed data (i.e. Linear Mixed-Effects Models; Baayen et al. 2008), as well as for binary response data (i.e. Generalized Linear Mixed Models or mixed logit models; Jaeger 2008).

[32] Potentially harmful averaging includes both the risks of averaging over small cell sizes and the impact of missing data which must either be removed from the analysis or replaced on the basis of some kind of statistical imputation. Such procedures are avoided using mixed-effects modelling.

[33] I acknowledge that this aspect is somewhat limited for language items. That is, for example, if you were running an experiment that looked at all existing words of English that rhyme with 'honey', then you would be looking at a finite set. Nevertheless, all relevant research using linear mixed-effects modelling as their statistical approach has included 'item' as a random effect. This limitation would apply of course equally to the MinF' approach based on ANOVAs using both participant and item averages.

[34] Other variables which can be included as fixed factors in a mixed-effects model are associated with the various longitudinal effects inherent to an experiment. This inclusion has been argued to be another key feature of mixed-effects modelling (see section above, as well as Baayen et al. 2008:399).

different items can have the same length, and three other items can have the same frequency and can all appear in the same session).

Across all regression models, a baseline value of the dependent variable (e.g. the latency or accuracy of a response) is reflected in the intercept of the regression function for the hypothetical scenario in which all independent variables have the value *zero* (with the exception of categorical variables, where the intercept is either at the level that is alphanumerically first or that has been explicitly set as the baseline; the intercept values for the corresponding regression models are reported in Appendices C and D). Including fixed effects in a regression model means specifying the main intercept according to the given factors or co-variates. The main intercept of the regression model can also be specified by *random intercepts*, which take each level of the random effect into account (e.g. if the variable 'participant' is included as a random factor, then random intercepts adjust the main intercept for the fact that some participants may be slower than others and will therefore exert lower individual intercepts). Likewise, the slope can be adjusted by the inclusion of *random slopes*. In fact, Baayen suggests that

> in general, predictors tied to subject (age, sex, handedness, education level, etc.) may require by-item random slopes, and predictors related to items (frequency, length, number of neighbors, etc.) may require by-subject random slopes. (2008:290–291)

The inclusion of random intercepts and slopes may be important to ensure that some of the obtained significant effects are not merely the result of individual differences, or that some fixed effects do not emerge as significant because participant (or item) differences were not accounted for (see Balling 2008:181–183, for an explicit example).

Researchers employing mixed-effects models point to certain areas where caution is required, namely in dealing with collinearity, outliers and model overfitting. Collinearity may be a concern in both an unbalanced and a balanced data set (Jaeger 2008:440), and is reflected in highly correlated variables. The problem with collinearity is that it makes it difficult to assess which variables are significant and which are not, which in turn makes the statistical model less stable. Among the tools which help prevent and/or remove collinearity are centering and standardizing of (highly) correlated variables, or residualisation procedures (see Balling 2008:184–186, for a detailed description of each of the methods).

With respect to outliers, it has long been suggested that the standard procedure of a priori screening and removal of outliers that are at a distance of (at least) two standard

deviations from the mean of the distribution may be in some cases too 'aggressive' and problematic (e.g. Luce 1986; Ratcliff 1993). In dealing with outliers in mixed-effects modelling, researchers have suggested minimal data trimming prior to analysis (based, for instance, on factors external to the experiment, i.e. too fast responses which follow an impulse rather than reflect cognitive processing), followed by model criticism (see Baayen & Milin 2010:7–12; 24, for an extensive discussion of this issue).[35]

Finally, researchers using mixed-modelling need to be careful to avoid model-overfitting. The danger of overfitting is clearly linked to the possibility of entering a wide range of factors into the statistical model. But how many variables are too many? One response to this question is that the factors in the model should ideally be *motivated*. This might be a priori motivation based on a) experimental hypotheses, b) previous findings, or c) common sense, or motivation that results from a step-wise model reduction approach. Similarly, a question which has been subject of much debate is whether non-significant variables should be left in the model or be simply left out. Whatever the best practice may be (see Balling 2008:186, for some recent suggestions), the step-wise approach seems widely accepted, at least in exploratory data analyses with no clearly set out hypotheses (Baayen 2008).

In the following two sections, I will briefly outline variable selection for the first study (i.e. including Experiments One and Two), as well as model specifications for the first of the two experiments.

3.2.2.3 Predictors

Primary experimental predictors

Four variables were of particular interest to the present study, namely Item Type (IHG versus German control versus nonword), Format (Title versus lower), Sentence Language (German versus English) and Proficiency (in the bilinguals' L2).

Properties of the items (secondary experimental predictors)

As secondary variables, the fitted model also included various item properties, namely Frequency (log and bigram), Length (in letters), the number of (German and English)

[35] Both minimal data trimming prior to analysis and model criticism entail an aspect of visual data inspection. That is, an element of data trimming looks at the overall fit of the data to a "normal distribution" under various transformations (log, inverse, etc.) and identifies outliers visually from the plot of that fit. Model criticism, on the other hand, involves an inspection of the residuals of a final model, and where necessary a subsequent removal of, for instance, "data points with absolute standardised residuals exceeding 2.5 standard deviations" (Baayen & Milin 2010:10).

neighbours, Word Class (noun versus other), as well as which experimental Session (one versus two) the items occurred in. While the former three factors (frequency, length and neighbourhood density) are commonly controlled for in experimental designs and included in analyses (for recent reviews and useful references, see Lemhöfer et al. 2008; New, Ferrand, Pallier, & Brysbaert 2006), the inclusion of the latter two factors (word class and session number) was due to task specifications. With regard to word class, the inclusion of this factor in our statistical model was motivated by the main predictor Format and the fact that all nouns receive an initial capital letter in German (the language of the lexical decision task and the participants' L2). Given that initial letter capitalisation may evoke certain a priori knowledge and processing mechanisms, but given also that not all items used in this study are nouns (see *Materials* section above), the aim was to take these potential effects into account and to ensure that our main effects of interest (i.e. our main predictors and their interactions) are not obscured by the syntactic category of an item. The rationale for including session number was the possibility of familiarisation with the task, an effect which has been previously reported in a number of psycholinguistic studies (cf. Elston-Güttler, Gunter, et al. 2005; von Studnitz & Green 2002a).

Longitudinal predictors

Trial Number and the response latency and accuracy to a Preceding Trial were also included in the statistical model, given that such factors have been recently identified as having a strong impact on language processing (e.g. Baayen & Milin 2010; Dijkstra et al. 2010; Kuperman et al. 2008).

Two random-effect factors

Participant and Item were the only two variables included in the model as random effects.

3.2.2.4 Model specifications

Prior to data analysis of Experiment One (i.e. data collected in New Zealand, from English-German bilinguals), three participants had to be excluded since they did not follow the given instructions and stopped paying attention to the task. In addition, one further participant had to be excluded due to an erroneous result in the on-line German proficiency test. (The test result returned 100% error rate on both existing words and nonwords, which most likely indicates equipment failure or a misinterpretation of

instructions.) Consequently, the initial data set comprised of 19032 observations. Closer data inspection resulted in further exclusions: first, responses that were faster than 300ms were removed, since they would not capture the required cognitive processing (cf. Dijkstra et al. 2010). Second, data from one last participant was excluded, who performed the experimental task with an extremely high error rate (greater than 50% on the entire data set, i.e. including IHGs, controls and nonwords). Finally, several experimental items were removed since they elicited high error rates.[36] The remaining data set contained 17431 observations.

Next, I inspected the distribution of response times (RTs) which revealed a marked non-normality. Following a comparison of three types of RT transforms (log transform, inverse transform and inverse square root transform), the response durations were subjected to an inverse transform (RT=-1000/RT), since it was most successful in attenuating the observed non-normality. The inverse transform includes a scaling factor of 1000 to avoid very small values for the dependent variable, and multiplied by -1 to ensure that positive analysis coefficients correspond to increases in response times (cf. Baayen & Milin 2010; Dijkstra et al. 2010).

Following model comparison, all analyses reported in this study include participants and items as random effects, with intercepts varying by participant and item (see discussions above). Random slopes for critical variables of interest were included where these provided a better statistical model. (Note that the model comparisons included tests for various random slopes.) Where necessary, predictor variables were either centered or residualised to remove collinearity between predictors.

For the continuous (ms, transformed) RT data, a minimally adequate *Linear Mixed-Effects Model* was fitted containing only correct responses and subjected to model criticism. Potentially harmful outliers were removed and the model was refitted. The measure of statistical significance of the fixed effects in the final model was based on Markov Chain Monte Carlo (MCMC) sampling (10,000 iterations). For ease of reading, the following presentation of significant results includes the mean and lower and upper limits of the 95% interval for the Higher Posterior Density, and the probability levels (*p*-

[36] Similar to Schulpen (2003), who also employed fairly low proficiency participants, the selected cut-off point was at an error rate larger than 30 percent, being a strong indicator that the items were not known by the participants. The group of eliminated items contained three IHGs, twelve German controls and two nonwords. Items in the matching presentation format were also excluded, even though they may not have produced the same number of errors (e.g. although 'gieben' produced a high number of errors, but 'Gieben' did not, both items were excluded from the analysis. Data loss due to item characteristics totalled no more than 6.4%.

values) generated by the MCMC sampling (full statistical models are listed in Appendix C).

For the binomial (incorrect/correct) error data, a *Generalised Linear Mixed Model* (henceforth, mixed logit) was fitted. In this analysis, statistical significance of the fixed effects is assessed using Wald's Z-test (see Jaeger 2008, for further detail and discussion of some problematic issues). For ease of reading and interpretation, the response time results are presented and discussed below based on plots of lmer model estimates (back-transformed for RTs) and standard errors in the relevant conditions.

3.2.3 Results and discussion

As outlined in the introduction, the aim of this study was to explore the impact of sub-lexical information (i.e. in the form of language-specific cues) on bilingual visual processing. The task was conducted in the second language (German) of unbalanced bilingual English-German speakers. The impact of such information was assessed with respect to the availability of a particular sentence language, as well as the proficiency level of the bilingual in their L2.

My main predictions are as follows: if language-specific cues can act as an "early language discrimination mechanism" (Lemhöfer & Dijkstra 2004:547) then responses to real words, and particularly the IHGs, should be facilitated when a target item has an onset capital letter. This is based on an assumption that the onset capital signals 'German' and provides early indication of the target language. Theoretically, this may either result in an increase in the activation of the German word, with no impact on the English IHG partner; or in a reduction of activation levels of the non-target language. The consequence of either scenario is that it should reduce lexical competition during the recognition process, which would be reflected in faster recognition of words marked with a language-specific cue. Such language information should be particularly beneficial when processing IHGs, because these words have an alternative reading in the non-target language (see discussion in the previous sections), which has been shown to compete during the selection process (e.g. van Heuven et al. 2008).
In contrast, rejections of nonwords with onset capitals should be inhibited, given their greater similarity to existing German words when carrying a language-specific cue.

These effects may be further defined by the language of the sentence frame that precedes the target. As discussed using Figure 2 above, traces of L1 interference should be considerably limited (if measurable at all) following an L2 sentence frame (cf. Elston-Güttler, Gunter, et al. 2005), and the reverse effect should be observed for L1 sentence frames. The latter is because changing from the language of processing to the language of the task may induce an additional processing cost on bilingual speakers (cf. Lemhöfer & Radach 2009).[37] Combining this conjecture with the presentation format of a target

[37] It is possible that this cost would diminish over the course of the experiment, as participants learn to separate their internal response to the prime sentence from their response to the lexical decision target. However, because the nature of the present experimental design was already very complex, it did not allow me to further test this conjecture (due to the lack of sufficient data cells). This aspect may need further clarification in future work.

item, both the facilitation effect for words and the inhibitory effect for nonwords may be manifest most strongly in a condition where the linguistic context meets the (L2) task demands in all aspects (i.e. L2 sentence frame, followed by the Title case presentation format). When the experimental setup, however, strongly supports the activation of the non-target language (i.e. L1 sentence frame, followed by an all lower case target item) facilitation should occur for responses to nonwords and inhibition should be more evident in responses to words. Perhaps the most interesting conditions are the 'middle grounds' (i.e. B and C in Figure 2). These conditions may tell us which source of information is more crucial immediately prior to the participants' decision making.

One last prediction concerns the L2 proficiency of the bilinguals involved in this experiment, given that in this experiment the task is conducted in the participants' less dominant L2. As shown in previous research on bilingual L2 processing, more proficient speakers may be able to process the L2 more automatically than their less proficient counterparts (cf. Schulpen 2003). As a result different resources may be employed to execute a task (cf. Elston-Güttler, Paulmann, et al. 2005), or the available resources may be employed in different ways (cf. Dijkstra 2005). What this means in the context of this experiment is that the predicted effects may be more pronounced for less proficient speakers, whereas they may be less clearly detectable with increasing proficiency in the L2. Alternatively, it could also be that increased exposure to the L2 may lead to, for instance, heightened expectation that an initial capital letter is indicative of a German word; hence, the predicted effects may be more pronounced for more proficient bilingual speakers.

Given these predictions, the main predictors included in the most basic mixed-effects regression model were Item Type (whose inclusion depends on the statistical analysis being carried out; see below), Format, Sentence Language and Proficiency. Subsequent models also included other factors, such as Frequency, Length, Word Class, the number of competing neighbours in both languages (i.e. German and English), Session number, Trial number and the response latency and accuracy to a Preceding Trial (for rationale, see previous discussion). Note that the final statistical model reported in each of the following sections (and throughout all analyses reported in Study One) included all statistically significant effects. The separation in to two analysis sections henceforth, i.e. main predictors and other variables, reflects a division between primary and secondary variables of interest only.

All analyses were first conducted on the word set as a whole (i.e. including both the German-English homographs and their German control words), and then separately on the critical stimuli (IHGs), and the controls. Responses to nonwords were analysed separately. The mean response latencies on correct trials and percent error rates are summarised in Table 3. The calculation of the mean values presented in this table followed the initial data trimming described in section 3.2.2.1 above, but preceded the statistical modelling reported below. Hence, they give a general pattern of response latencies and error rates.

Table 3.
Mean recognition latencies for correct responses (in milliseconds) and percent error rates (in parentheses) for IHGs, controls and nonwords in L1 (English) and L2 (German).[38]

Item Type	English (L1)		German (L2)	
	lower	Title	lower	Title
IHG	851 (18.64%)	811 (18.2%)	770 (16.86%)	848 (18.86%)
control	838 (14.81%)	871 (11.48%)	789 (16.47%)	906 (13.47%)
nonword	861 (14.88%)	903 (18.87%)	821 (13.66%)	933 (15.55%)

3.2.3.1 Error Analysis

3.2.3.1.1 Responses to IHGs and German control words

The following error analysis included both types of existing words, namely English-German homographs and their matched control words. A statistical (mixed logit) model was fitted to the data, following the procedure outlined above.

3.2.3.1.1.1 Main predictors

The analysis of error rates revealed Proficiency to be the only conventionally significant simple effect ($\beta = 1.137$, SE = 0.345, $z = 3.296$, $p < 0.001$, see Figure 4), with lower proficiency participants making more incorrect responses to existing German words than their more proficient counterparts (see Appendix C, Table E1). (Note that the values displayed in the figure(s) are predicted values from the model, with other factors set at their intercept value. This is why subsequent effects may differ from the values in Table 3 above). This effect was further defined by a marginally significant interaction

[38] In Appendix C (Table 14), I present the same data split into high and low proficiency groups (as determined by a median split value), as a better illustration of the relationship between RTs and accuracy for these bilingual speakers.

with Item Type (β = -0.228, SE = 0.119, *z* = -1.906, *p* = 0.057, Figure 5), showing that the same less proficient bilinguals struggled particularly with their responses to English-German homographs compared with German control words (for comparable results regarding participants' L2 proficiency, see Chambers & Cooke 2009; Schwartz & Kroll 2006).

The absence of (other) interactions involving Item Type reflects similar reports in previous research (see section 3.1.2 for an extensive literature review). This may not only reflect the fact that IHGs and controls were well matched in their characteristics, but also that the remaining variables of interest (i.e. Format and Sentence Language) had a similar impact on both item types. To investigate this issue further, separate analyses were conducted on IHGs and controls, respectively and the full reports can be seen in the sections 3.2.3.1.2 and 3.2.3.1.3 below.

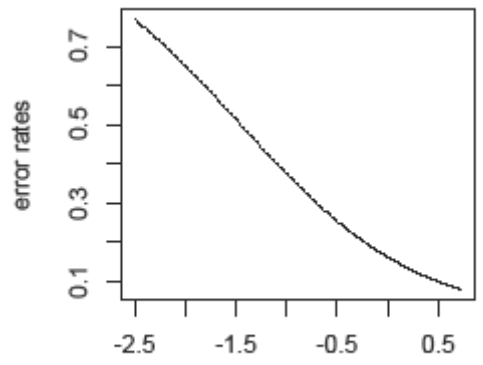


Figure 4. Proportion of incorrect responses (lmer *model estimates*) to words, with increasing L2 proficiency.

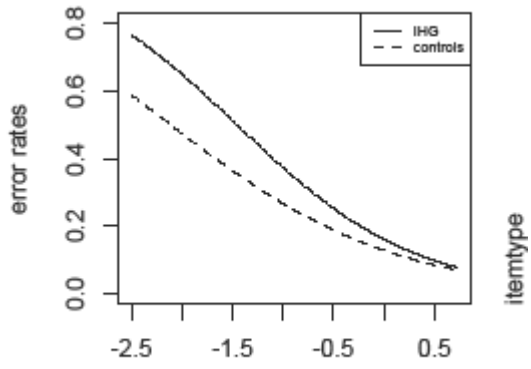


Figure 5. Proportion of incorrect responses (lmer *model estimates*) to IHGs and controls, in relation to the speakers' L2 proficiency.

Participants' L2 proficiency was also involved in a two-way interaction with Sentence Language (β = 0.536, SE = 0.189, *z* = 2.836, *p* < 0.001, see Figure 6), and a three-way interaction with Sentence Language and Format (β = -0.645, SE = 0.242, *z* = -2.668, *p* < 0.001, Figure 7).

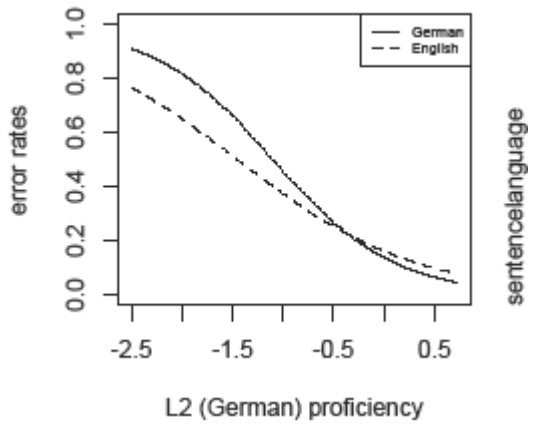


Figure 6. Proportion of incorrect responses (lmer *model estimates*) to words, after an L2 (German) or an L1 (English) context and with increasing L2 proficiency.

Several researchers have argued that compared with their more proficient counterparts, lower proficiency bilinguals make generally less efficient use of L2 contextual information available to them (e.g. Chambers & Cooke 2009; Elston-Güttler, Paulmann, et al. 2005; Gernsbacher 1993; Schwartz & Kroll 2006) and produce more errors, which is what the two-way interaction illustrated above shows us. Interestingly, these less proficient bilinguals seem more error prone when responding to the German target words after a German sentence. A possible explanation of this finding is that given various factors (e.g. poorer L2 vocabulary and/or grammatical knowledge), these participants have greater difficulty with the processing and interpretation of the auditory input in their less proficient language.[39] As a result, the decision making process of lower proficiency participants may be more reliant on resources other than meaning-based information, namely form-related information (e.g. Elston-Güttler, Gunter, et al. 2005; Schwartz & Kroll 2006).

[39] Note that the only way to test this conjecture would be to conduct a comprehension task with participants with varying L2 proficiency. Possible differences obtained in the memory recall task (performed between each of the six experimental blocks in a session) cannot be used as a reliable source of participants' comprehension level because this task was not explicitly based on comprehension and participants may have used various other strategies to execute it.

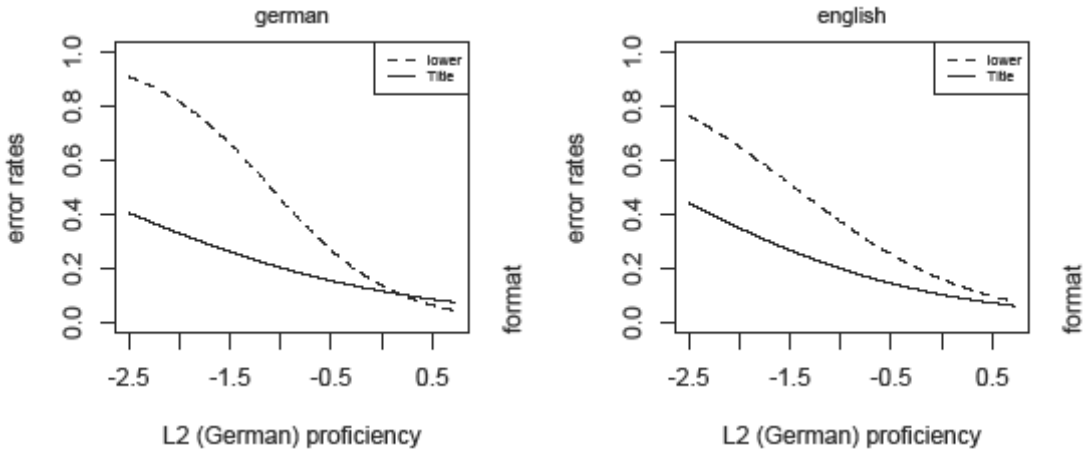


Figure 7. Proportion of incorrect responses (lmer *model estimates*) to words, by L2 proficiency, stimulus format (lower vs. onset capital) and preceding sentence language (German, in the left panel, and English, in the right panel).

We get a similar impression from the three-way interaction which shows that when items are presented with an onset capital letter (i.e. in the language-specific condition), lower proficiency participants' responses are generally made more accurately compared to the instances when no such information is available (i.e. all lower case presentation format). The strongest evidence for low proficiency participants' reliance on sub-lexical cues, however, is provided by participants' responses following a prime sentence in German. In particular, when there is no language-specific bottom-up information for German (i.e. when the target words are presented all in lower case), the participants' performance deteriorates dramatically. In addition, none of these effects show up in the results for more proficient L2 participants. Since this goes to some extent against what I would have predicted, this unexpected result will be looked at more closely in the discussion below (see sections 3.2.3.2 and 3.4).

Over the entire group of participants (i.e. irrespective of proficiency), the analysis reveals a significant interaction of Sentence Language and Format (β = -0.318, SE = 0.157, z = -2.030, p = 0.042, Figure 8). The results show that words are more reliably identified when presented with an onset capital (i.e. German bias), and that most errors are produced for the combination of an English (L1) sentence frame with the absence of language-specific bottom-up information supporting the L2 (i.e. lower case format).

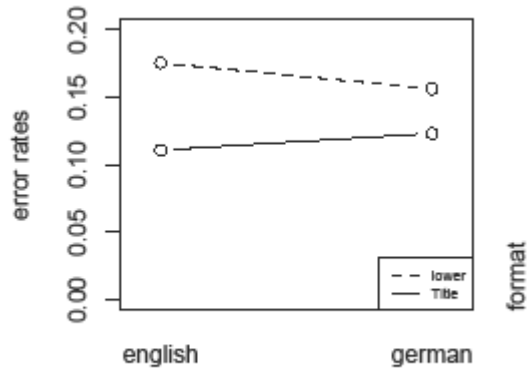


Figure 8. Proportion of incorrect responses (lmer *model estimates*) to words, in relation to the (prime) sentence language and the words' presentation format.[40]

3.2.3.1.1.2 Other variables

Among the secondary predictors which were found to be statistically significant in the fitted model were frequency (β = 0.801, SE = 0.116, z = 6.913, p < 0.001, Figure 9) and length (β = 0.393, SE = 0.083, z = 4.708, p < 0.001, Figure 10), with more frequent and longer words being identified more accurately (since frequency was residualised against length, this is a genuine length effect, not tainted by the relationship between length and frequency). Note that the negative values displayed in both figures reflect a consequence of residualisation.

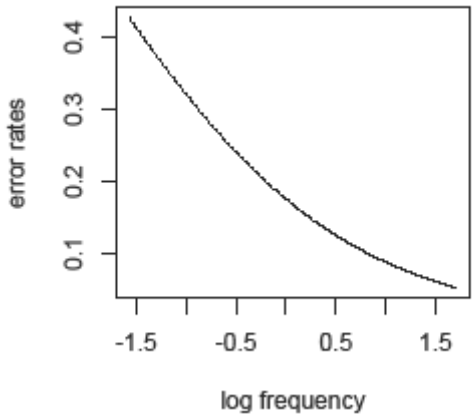


Figure 9. Proportion of incorrect responses (lmer *model estimates*) to words, with increasing word frequency.

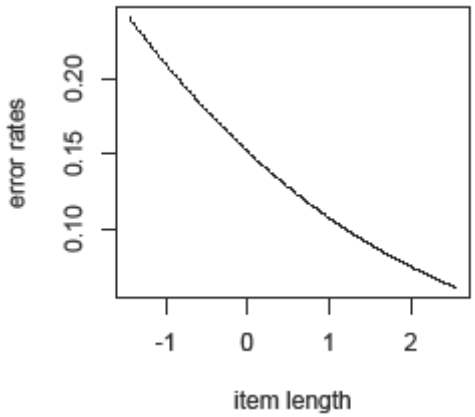


Figure 10. Proportion of incorrect responses (lmer *model estimates*) to words, with increasing word length.

[40] Note that even though the lines connecting the binary variables are meaningless, I opted for this type of presentation for ease of reading of the effect(s). This presentation type is adopted throughout the entire thesis.

We also observe a difference between the two experimental sessions (β = 0.360, SE = 0.082, *z* = 4.392, *p* < 0.001, Figure 11), which most likely reflects familiarisation with the task.

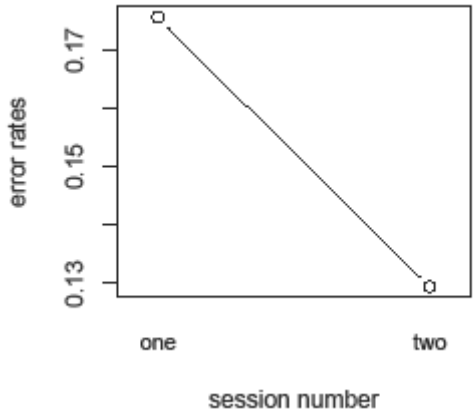


Figure 11. Proportion of incorrect responses (lmer *model estimates*) to words in the two experimental sessions.

3.2.3.1.2 IHG data

Given that the analysis above revealed differences in the accuracy of identifying IHGs as opposed to control words depending on the L2 proficiency of a speaker (see Figure 5), a separate analysis of responses to IHGs was carried out to determine the impact of the selected variables of interest (i.e. Format, Sentence Language and Proficiency). This analysis is identical to the previous analysis, except that it of course excludes Item Type as a predictor.

3.2.3.1.2.1 Main predictors

The analysis shows very similar statistical effects (see Table E2, Appendix C) to those obtained in the analysis reported above. Proficiency again shows a significant effect (β = 1.102, SE = 0.432, *z* = 2.549, *p* = 0.011, Figure 12), with more accurate responses to homographs among more advanced bilinguals. In addition, Proficiency interacts with Sentence Language (although only marginally so; β = 0.459, SE = 0.260, *z* = 1.768, *p* = 0.077) and with Format and Sentence Language (β = -0.681, SE = 0.324, *z* = -2.105, *p* = 0.035, Figure 13).

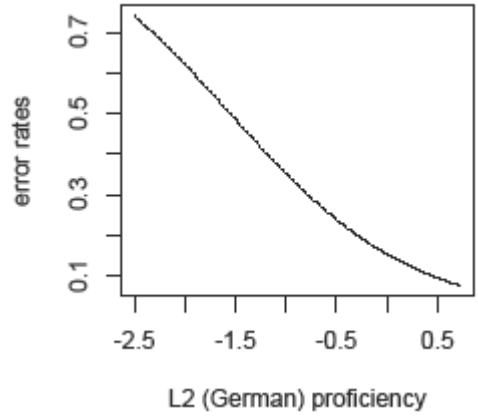


Figure 12. Proportion of incorrect responses (lmer *model estimates*) to English-German homographs, with increasing L2 proficiency.

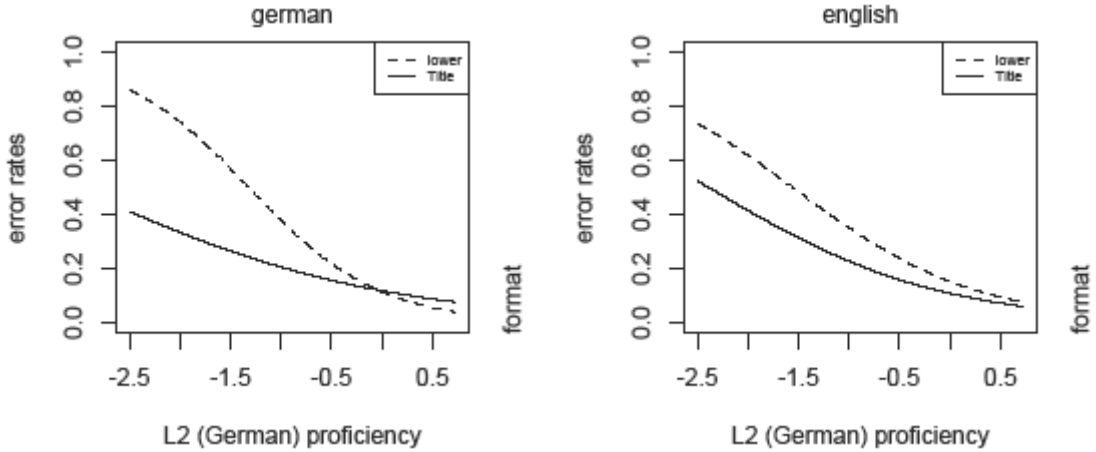


Figure 13. Proportion of incorrect responses (lmer *model estimates*) to IHGs, by L2 proficiency, stimulus format (lower vs. onset capital) and preceding sentence language (German, in the left panel, and English, in the right panel).

As can be seen from Figure 13 above, the three-way interaction is primarily carried by the difference between lower and Title case for less proficient bilinguals responding to homographs following a German sentence frame. Together with the simple effect of Proficiency, there are two major implications of the findings. First and foremost, form-related L1 lexical competitors seem to be activated more strongly for lower proficiency bilinguals than for their higher proficiency counterparts. As a result, the subsequent competition for selection causes greater cross-linguistic interference and produces more errors. Another implication of the findings is that lower proficiency speakers rely greatly on orthographic language cues for their lexical decision. The lack of such information clearly inhibits performance.

With regards to the impact of sentence language, the data suggest that hearing a sentence in the participants' L1 may contribute more strongly to cross-linguistic homograph competition (as also reflected in the simple effect of Sentence Language: β = 0.340, SE = 0.145, z = 2.349, p = 0.019, Figure 14), leading to an increased number of

errors for the L2 lexical decision task. As in the analysis of the combined IHG and control word set, this effect is further defined for these IHG tokens alone by an interaction of Format and Sentence Language (β = -0.436, SE = 0.213, z = -2.045, p = 0.041, Figure 15). When the language of the prime sentence does not support the L2 lexical decision response, the activation of the L1 representation of an interlingual homograph seems somewhat reduced by the availability of language-specific cues. Conceivably then, participants may experience less L1 interference.

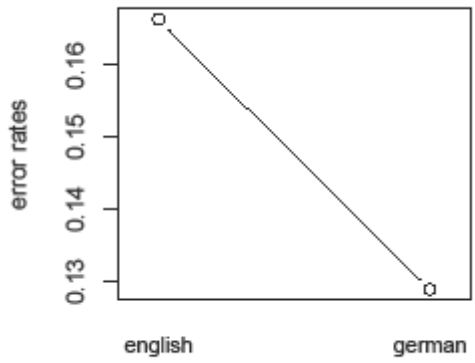


Figure 14. Proportion of incorrect responses (lmer *model estimates*) to English-German homographs, depending on the (prime) sentence language.

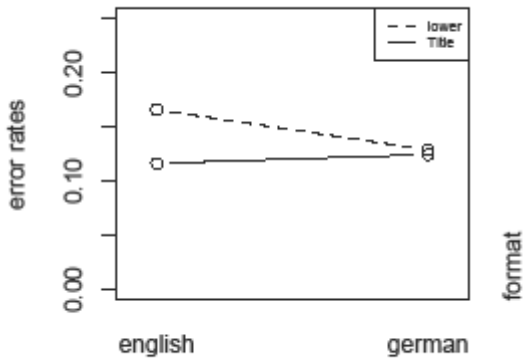


Figure 15. Proportion of incorrect responses (lmer *model estimates*) to English-German homographs, in relation to their presentation format and the preceding sentence language.

3.2.3.1.2.2 Other variables

In this subset of data, both the frequency of an IHG and the experimental session showed the same significant effects as in the main analysis (Frequency: β = 0.690, SE = 0.184, z = 3.742, p < 0.001, Figure 16; Session: β = 0.491, SE = 0.094, z = 5.198, p < 0.001, Figure 17). An additional factor which had an impact on participants' responses was the *English* word class of the homograph (β = 1.091, SE = 0.371, z = 2.944, p < 0.001, Figure 18), with less accurate responses to English nouns.

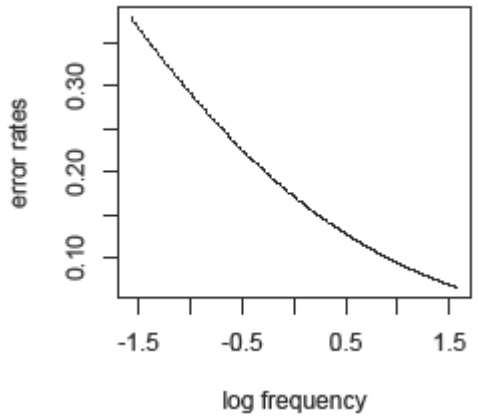


Figure 16. Proportion of incorrect responses (lmer *model estimates*) to English-German homographs, by increasing word frequency.

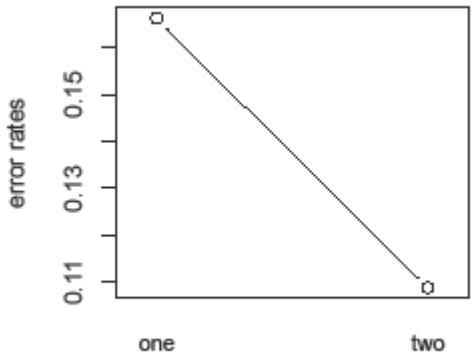


Figure 17. Proportion of incorrect responses (lmer *model estimates*) to English-German homographs, depending on experimental session.

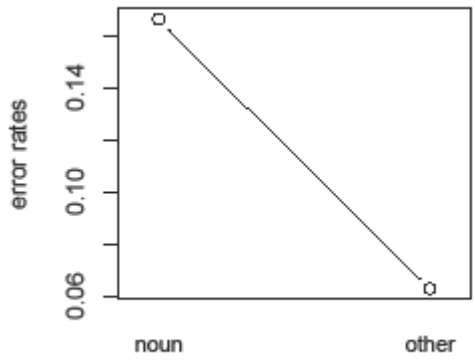


Figure 18. Proportion of incorrect responses (lmer *model estimates*) to English-German homographs, depending on the L1 word class of the stimulus.

Since this finding shows an effect of lexical properties of the L1 on responses to targets in the L2, it strongly implies non-selective access to L1 as well as to L2 representations. The question that arises is why did nouns produce more errors? A possible interpretation of this result may be related to a recent finding suggesting that responses to interlingual homographs are easier to make (and hence, more correct) when there is an agreement in the syntactic category across languages (Baten et al. 2011). Accordingly, there may have been few instances where IHGs classified as nouns in the participants' (more dominant) L1 did not share the same syntactic category in the L2 (German). This would have made the targets more difficult to process, resulting in more errors. I tested this conjecture by performing a separate analysis with a slightly reduced data set, i.e. after the exclusion of 5 IHGs that were nouns in English but belonged to a

different syntactic category in German.[41] Notably, the effect of English word class disappeared; hence, the present results are consistent with previous findings.[42]

3.2.3.1.3 German control word data

As a complement to the separate error rates analysis of IHGs, the following analysis focuses on responses to (German) controls only. A mixed logit model was fitted to the data, including the predictors discussed in section 3.2.3.1.2.

3.2.3.1.3.1 Main predictors

In this subset of data, including responses to German control words, only Proficiency shows as a conventionally significant simple effect (β = 0.877, SE = 0.377, z = 2.325, p = 0.020, Figure 19), with lower proficiency bilinguals displaying a larger error proportion (see Table E3, Appendix C). Proficiency is also involved in an interaction with Sentence Language (β = 0.645, SE = 0.278, z = 2.314, p = 0.021, Figure 20), and a marginal three-way interaction with Sentence Language and Format (β = -0.676, SE = 0.370, z = -1.826, p = 0.068, Figure 21). There was no interaction of Sentence Language with Format.

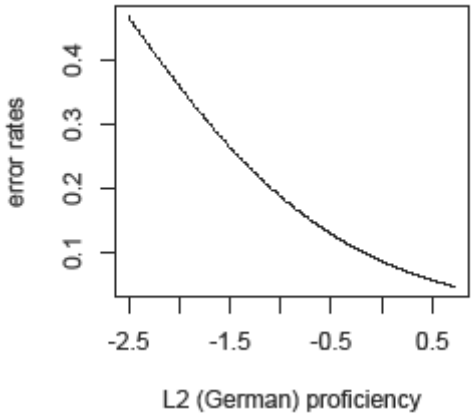


Figure 19. Proportion of incorrect responses (lmer *model estimates*) to German control words, with increasing L2 proficiency.

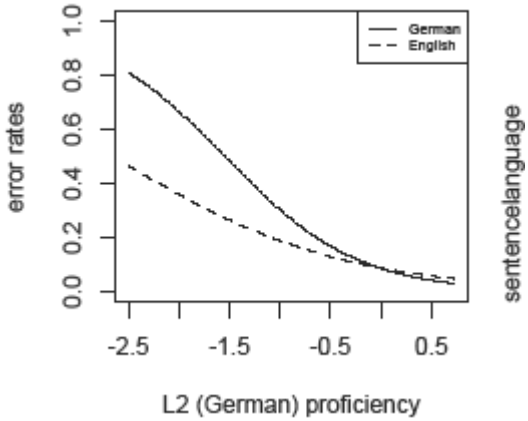


Figure 20. Proportion of incorrect responses (lmer *model estimates*) to German controls, in relation to the preceding sentence language (L1 or L2) and increasing L2 proficiency.

[41] Note that for the purposes of this study and ease of interpretation, interlingual homographs were classified as either 'noun' or 'other' in German and English, respectively. Based on this classification, 6 out of the 39 IHGs were not nouns in German, of which only one shared the same classification in English (i.e. the other 5 were nouns in English). The calculated ratio of category congruency for the entire IHG set was 0.72. In the slightly reduced set, however, the category congruency ratio increased to 0.82.

[42] A separate analysis also tested the possibility that this effect was due to the nature of the experiment, by inspecting whether (English) word class would enter an interaction with Proficiency or Format, respectively. Neither interaction was returned significant, which provides strong support for the previous discussion, i.e. the necessity of category congruency. As a note of caution, however, it needs to be stated that any reduction of the original dataset is going to make it more difficult to achieve significant effects (i.e. due to a reduction of power), and that I was aware of such.

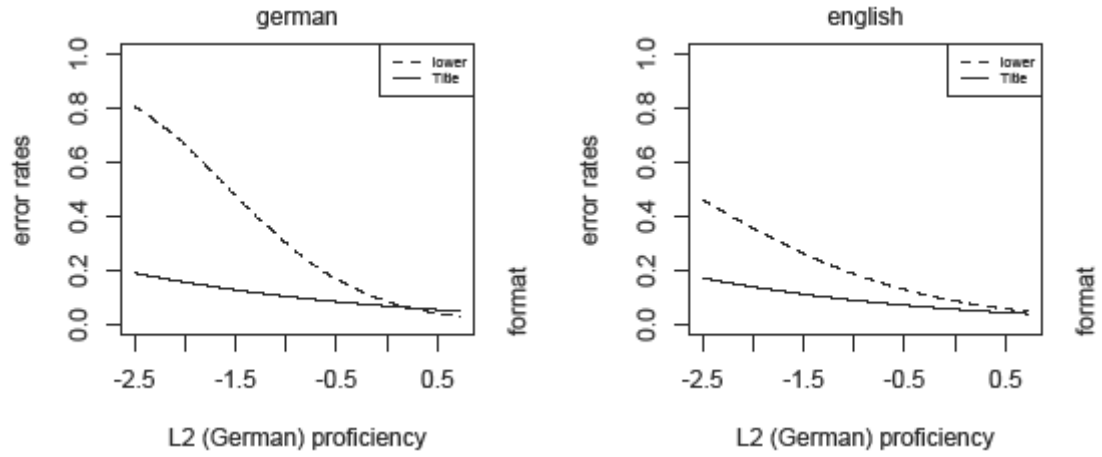


Figure 21. Proportion of incorrect responses (lmer *model estimates*) to German controls, by L2 proficiency, stimulus format (lower vs. onset capital) and preceding sentence language (German, in the left panel, and English, in the right panel).

This pattern of responses is very similar to those observed for the entire data set, with format having a major impact on less proficient bilinguals' response accuracy. In particular, given the absence of the previously reported interaction of Format and Sentence Language, these findings suggest that irrespective of the preceding sentence language, the presence or absence of language-specific cues is taken by these bilinguals as a primary resource in the decision making process.

These findings are in line with the argument set out by previous research, suggesting that less proficient bilinguals are particularly reliant on form-related resources, which may be a result of an inefficient use of other sources of information (e.g. language context; Elston-Güttler, Paulmann, et al. 2005; Gernsbacher 1993; Schwartz & Kroll 2006). The present data extend previous research by suggesting that the impact of form-related resources may come into play as early as at the sub-lexical level. As argued by Dijkstra (2005), this may be done in a top-down way by means of indicating the language of processing; however, this would be dependent on a foregoing bottom-up activation of a language determining mechanism (e.g. in the form of a language node). In contrast, more proficient bilinguals can not only make better use of the language of the sentence preceding the target, but also process both languages more automatically, with less reliance on *lower* level lexical information. It could be argued that these speakers are more effective in employing *higher* level lexical information, something that has been argued to be a major part of automatic language processing.

3.2.3.1.3.2 Other variables

Once more, we find significantly more correct responses to more frequent and longer words (Frequency: β = 0.748, SE = 0.144, z = 5.200, p < 0.001, Figure 22; Length: β =

0.304, SE = 0.125, z = 2.441, p = 0.015, Figure 23). In addition, the more English word neighbours a German control has, the more errors are made to the stimulus (β = -0.055, SE = 0.026, z = -2.157, p = 0.031, Figure 24). This is an interesting finding, and clearly in line with previous bilingual research, suggesting that bilingual language processing is largely non-selective in nature (e.g. van Heuven et al. 2008).

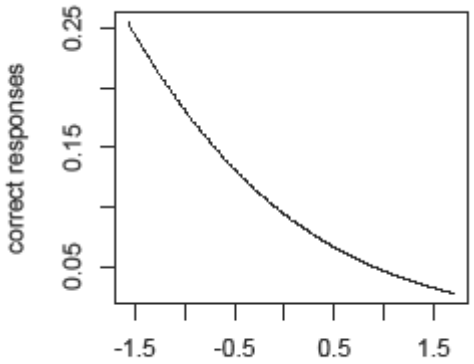


Figure 22. Proportion of incorrect responses (lmer *model estimates*) to German control words, with increasing word frequency.

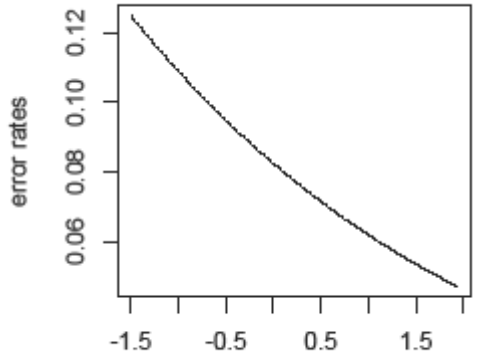


Figure 23. Proportion of incorrect responses (lmer *model estimates*) to German controls, with increasing word length.

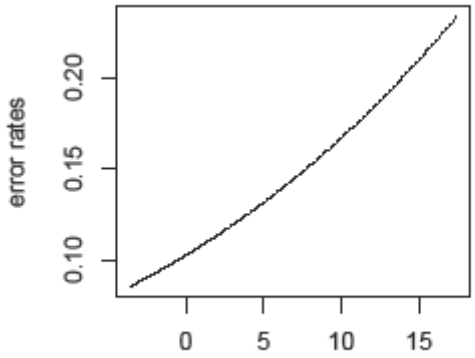


Figure 24. Proportion of incorrect responses (lmer *model estimates*) to German control words, with an increasing number of L1 neighbours.

3.2.3.1.4 Nonword data

The study of nonword processing is a relatively under-investigated area (particularly in the bilingual domain), despite a recent claim that it is

> a useful tool for the study of bilingual lexical organization, [since] it provides valuable extensions of current concepts and models of visual word recognition in bilinguals, in particular, with respect to the limits of non-selectivity during lexical access. (Lemhöfer & Radach 2009:46)

To understand potential implications of nonword processing, we first need to understand the underlying processing mechanisms. With respect to *monolingual* nonword processing, Grainger and Jacobs have suggested that in a lexical decision task (LDT) nonwords are rejected mainly on the basis of meeting a "temporal deadline" (1996:522). This means that if the search for a matching word entry in the lexicon remains unsuccessful within a certain period of time, a negative ("no") response is given. The authors argue that in the course of an experiment the deadline can be set to a longer deadline value if there is increased global activity in the orthographic lexicon (referred to as the σ value), i.e. if the stimulus set is more word-like (ibid.).
Applying this theory to a *bilingual* context, researchers have observed that nonword rejection times and error rates depend on the similarity of the nonwords to existing words in either one or the other language of a bilingual speaker (Lemhöfer & Dijkstra 2004; Lemhöfer & Radach 2009; see also van Heuven et al. 1998, for effects of neighbourhood density on nonword rejection).[43] The authors interpreted these findings in terms of an *early language discrimination mechanism*, which allows for nonwords (as well as words) to be distinguished as belonging to one or the other language before their actual recognition or rejection (Lemhöfer & Dijkstra 2004:547). Two questions which remain unresolved are: what is the nature of such a discrimination mechanism and at what level does such a distinction take place. Hence, the following analysis aims to clarify whether sub-lexical information (in the form of the language-specific cue carried by Title case in German) can provide such sufficient language discrimination information. The reader is reminded that apart from the possible influence of the lower or upper case of the initial letter, the nonwords were constructed to be as equally (dis)similar to words in both L1 and L2 in the present study as possible.

Similar to the previous analyses, a statistical mixed logit model was fitted to the data, in this case consisting of responses to the nonword tokens only. The same procedure was followed and the same (main and secondary) predictors were included in the fitted model, with the exception of Item Type.

[43] Even in research that included nonword types that were similar to existing words in both languages, nonword rejection times and error rates depended highly on the similarity of the nonwords to one language only (Lemhöfer & Radach 2009).

3.2.3.1.4.1 Main predictors

In responses to nonwords, Proficiency is once more revealed as a significant factor (β = 1.220, SE = 0.258, z = 4.724, p < 0.001, Figure 25), with lower proficiency L2 speakers being more inclined to accept something unknown as an existing German (L2) word. The only other significant simple effect found in these data is for Sentence Language (β = 0.291, SE = 0.071, z = 4.125, p < 0.001, Figure 26), showing that participants' responses are more accurate when a target is preceded by a task-relevant (L2) sentence frame (see Table E4, Appendix C). This finding may be indicative of a successful pre-activation of the task-relevant (L2) language, which results in reduced interference from the task-irrelevant L1 and produces fewer errors.

Unlike in the responses to real words, the two significant predictors did not enter a significant interaction with each other, or with Format. This implies that the orthographic cues employed in this experiment were not merely used as a strategic response facilitator (see the discussion section).

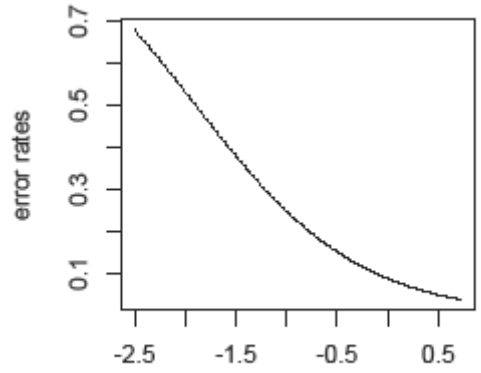


Figure 25. Proportion of incorrect rejections (lmer *model estimates*) for nonwords, with increasing L2 proficiency.

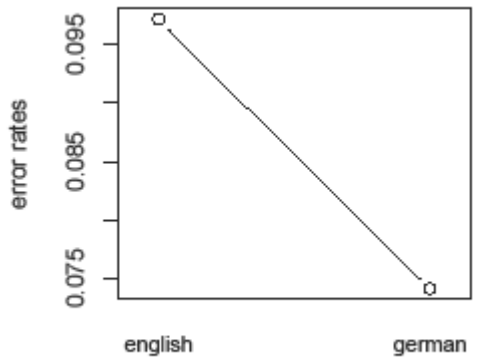


Figure 26. Proportion of incorrect rejections (lmer *model estimates*) for nonwords, as defined by the preceding sentence language.

3.2.3.1.4.2 Other variables

In line with expectations, a nonword that is more word-like in German (as indicated by a higher German bigram frequency) is more difficult to reject. This is revealed in the simple main effect of Bigram Frequency (β = 0.000, SE = 0.000, z = -3.699, p < 0.001, Figure 27). (The English [L1] bigram frequency count was not included in the statistical model, since the values were much lower than those calculated for the L2, and so it was not expected to have a major influence on processing. Lower bigram frequency count in the L1 was characteristic across all items, as can be seen in Table 2 above.). The fact that there is a higher number of competing German words in the stimulus' neighbourhood

makes a rejection even more difficult (β = -0.239, SE = 0.029, z = -8.130, p < 0.001, Figure 28). In contrast, an increasing number of English word neighbours has the reverse effect, i.e. facilitating correct rejections of unknown items (β = 0.080, SE = 0.016, z = 4.914, p < 0.001, Figure 29).[44] Both findings are in line with the proposal of a mechanism which is thought to be responsible for an early discrimination of input based on its characteristics, i.e. causing more difficulties with the rejection of stimuli that are more similar to the current language of processing, and vice versa (Lemhöfer & Dijkstra 2004; van Heuven et al. 1998).

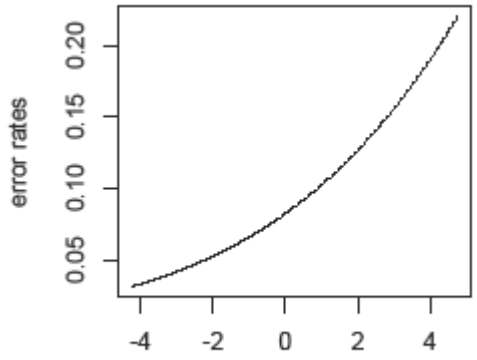


Figure 27. Proportion of incorrect rejections (lmer *model estimates*) for nonwords, against increasing size of the L2 neighbourhood.

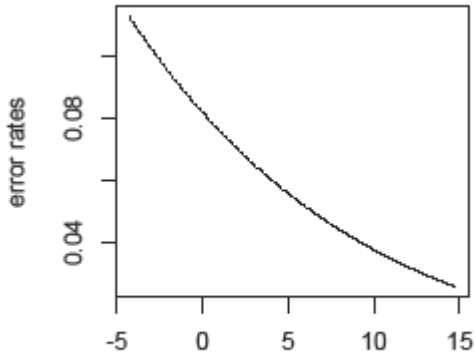


Figure 28. Proportion of incorrect rejections (lmer *model estimates*) for nonwords, against increasing size of the L1 neighbourhood.

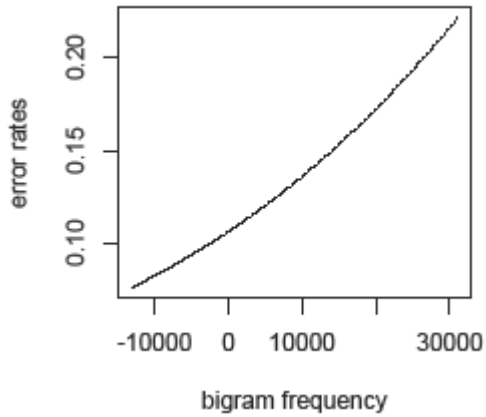


Figure 29. Proportion of incorrect rejections (lmer *model estimates*) for nonwords, against increasing L2 bigram frequency.

The analysis of the accuracy of nonword responses has two major implications. Most importantly, the analysis confirms the previous finding that "lexical access is to a certain

[44] Note that both factors were centred in the same way prior to analysis. The differences in the scales on the horizontal axes derive from the fact that although the items' mean neighbourhood density was relatively balanced, the standard deviation varied greatly between the two languages (see the methodology section). Notably, particularly shorter nonwords had a larger number of orthographic neighbours in English.

degree language-selective" (Lemhöfer & Radach 2009:46). However, although the present findings seem to support the idea that language discrimination takes place very early on (i.e. based on neighbourhood density), they do not indicate whether such processes can operate as early as at the sub-lexical level. An analysis of response latencies (see below) may shed further light on this issue.

3.2.3.2 Discussion of the Error Analysis

The present study investigated what impact language-specific orthographic information has on the recognition of interlingual homographs when processing in a bilingual's less dominant L2. The results reported above are in line with previous research (e.g. Schwartz & Kroll 2006), showing that for more proficient bilinguals there appears to be relatively minimal cross-linguistic activation. In contrast, form-related lexical competitors seem strongly activated for less proficient bilinguals, causing a large number of incorrect responses.

One central factor seems to play a crucial role in the decision making process for these lower proficiency bilinguals: The availability of a language-specific cue. The pattern which is consistently found in all three analyses of word data is that words carrying an onset capital letter are identified more correctly. (Notably, the lack of a format effect for nonwords shows that it is unlikely to be just a simple orthography-based response shift; see discussion below.) A possible explanation of this finding is related to the earlier notion that lower proficiency speakers process the L2 via their L2-L1 lexical links rather than accessing the relevant L2 concepts. In such cases, (L1) word form interference is expected to be much greater (e.g. Elston-Güttler, Paulmann, et al. 2005:1597), and as exemplified by the present study may only be reduced if language-specific information is available. According to Dijkstra, diacritical markers (here, onset capitals) qualify as adequate language navigators (or language-specific information). He suggests that "this kind of information is often used in a top-down way (i.e. the bilingual notices a particular [sub-lexical cue] that is unique to a language and then uses this information for his or her language decision)" (2005:187); particularly (if not exclusively) if these words a less known to the bilingual.

The magnitude of the impact of sub-lexical cues becomes visible when no such language information is provided (i.e. when target presentation is all in lower case). In such instances even a task-relevant sentence language (i.e. German) does not minimise the

seemingly detrimental effects. On the contrary, L2 word recognition seems particularly affected when stimulus presentation does not explicitly match the preceding German sentence, i.e. when stimuli do not carry an onset capital letter and hence do not accentuate their 'German-like' nature. A possible interpretation of this finding is that less proficient bilinguals try to compensate for their L2 sentence processing difficulties by over-relying on format as their 'language navigator', or rather the lack of it. Alternatively, it could also be that an L2 sentence frame leads to an expectation of the onset capital, which is not met. However, given that the same tendencies are manifest in the responses following English (i.e. L1) sentences, it is also possible that bottom-up information (i.e. language-specific cues) is the principal source of language information in L2 lexical processing, over and above top-down language information (i.e. sentence language).

This is not to imply that top-down language information has no impact on less proficient bilinguals' initial processing stages. As mentioned above, poor performance following L2 sentences has been previously reported and can be ascribed to various factors, such as a "less efficient enhancement and suppression mechanism [of relevant and irrelevant contextual information, respectively]" (Schwartz & Kroll 2006:199). The impact of top-down information on less proficient bilinguals' lexical processing is just as strongly defined by a preceding L1 sentence. This is probably best reflected in the accuracy of responses to IHGs only, i.e. in a condition which allows investigation of the relative activation of the non-target lexicon due to the words' different readings in the respective languages. In line with expectations, the data suggests that a preceding L1 sentence elevates the activation levels of the non-target L1 lexicon, resulting in higher cross-linguistic competition and a somewhat impaired L2 word recognition. Alternatively, it could also be that IHGs were rejected in the L2 because participants recognised them as belonging to their L1 first. More importantly, however, the availability of language-specific information seems to reduce such competition effects or false recognition (see above).

With increasing proficiency in the L2, participants seem less misled by, but certainly not oblivious to, the available language information. Accordingly, the data for these more proficient participants do not reveal clear differences in processing stimuli with a varying presentation format, or preceded by a varying sentence language. The former result is predicted by Dijkstra who argues that words that are well known by the

bilingual will be recognised in a much faster and automatised bottom-up way (2005:187); hence, language-specific sub-lexical information such as format may not be part of the bottom-up information that is used by the automatic word recognition system. Note that this observation is consistent with the long-established phenomenon in monolingual reading research, namely the Stroop effect (e.g. Stroop 1935; for a recent account, see also Raz, Moreno-Iñiguez, Martin, & Zhu 2007). In a Stroop task, proficient readers find it difficult to ignore the meaning of a word like WHITE when asked to give its colour ("black"). This provides evidence for the automaticity of reading, i.e. proficient learners display that variation in the form of the word (i.e. colour of a word; or, in the present thesis onset capitalisation) has little effect on their automatic lexical recognition.

There are several implications arising from the finding that L2 processing by higher proficiency bilinguals is not clearly affected by the language of the preceding sentence. First and foremost, the present results reflect the general expectation that high proficiency participants do not experience the same difficulties in processing sentence material in their L2 as their lower proficiency counterparts (e.g. Gernsbacher 1993, 1997; Gernsbacher & St. John 2001). Second, in accordance with several other studies (e.g. Chambers & Cooke 2009; Duyck et al. 2007; Elston-Güttler, Gunter, et al. 2005) the present results suggest that the availability of an L2 language context can considerably reduce, if not eliminate, lexical activation of the non-target language, even though the sentence may be unrelated in meaning to the target (Schwartz & Kroll 2006, experiment one; but see Libben & Titone 2009, for contrasting results). This is probably best reflected in the absence of homograph interference (in the separate IHG analysis above). A slightly different argument can be put forward, however, when we take into account the observation that participants did not seem clearly affected by a preceding sentence in the non-target language (L1). While a similar result has been reported in previous research (de Bruijn et al. 2001), two explanations are possible with regard to this finding. On the one hand, it could be argued that the results point to bottom-up processing as the primary direction of word activation, with generally little direct influence of top-down information (particularly when there is no constraining *semantic* information). Fast, automatised bottom-up processing with increasing proficiency in the L2 would then explain the present results. On the other hand, the findings could also be explained with regard to the bilinguals' language mode (Grosjean 1997, 1998, 2001). Depending on the condition, the bilinguals participating in this study may have adjusted

their language mode from 'monolingual' to 'bilingual', to better deal with the task. Hence, this particular group of higher-proficiency participants did not show explicit effects of varying language input on their responses. This interpretation is in agreement with the task schema account (Green 1998a), which proposes that depending on task requirements the relevant task schemas can be configured in various ways (e.g. into a monolingual as opposed to a bilingual *mode*, to display or allow for code-switching; see also *General Discussion* for further argumentation).

In contrast to real word responses, something that has no actual representation in the bilinguals' mental lexicon (i.e. nonwords) and is therefore processed less automatically is dependent on various sources of information; i.e. both bottom-up (e.g. neighbourhood density) and top-down (e.g. sentence language; for both, see discussion above). It seems unlikely that the presentation format does not wield any influence on nonword processing; instead, I believe that the absence of a format effect can be interpreted as evidence that participants did not follow an orthography-based strategy (i.e. say "yes" whenever they detect language-specific orthography) to execute the task (cf. Thomas & Allport 2000). This conjecture will be further examined in the following analysis of participants' response latencies.

3.2.3.3 RT analysis

3.2.3.3.1 IHG and German control word data

For all response time analyses, the statistical method chosen was linear mixed-effects modelling (for a discussion, see section 3.2.2.4 above). The first of four analyses reported below included both types of existing words, namely English-German homographs and their matched controls. A linear mixed-effects model was fitted to the data, including Item Type, Format, Sentence Language and Proficiency as the main predictors and a number of secondary predictors (see section 3.2.2.3, for detail). Subject and Item were included in the model as random effects. Prior to analysis, predictor variables were centred or residualised - where necessary - to avoid collinearity.

3.2.3.3.1.1 Main predictors

The latency data suggest that the proficiency of an L2 speaker not only influences the speed with which an item is recognised (MCMCmean = -0.372, HPD95lower = -0.498, HPD95upper = -0.250, $p < 0.001$), but also the extent to which a bilingual speaker makes use of existing bottom-up (Proficiency and Format: MCMCmean = 0.236, HPD95lower =

0.085, HPD95upper = 0.397, $p < 0.001$, Figure 30) and top-down information (Proficiency and Sentence Language: MCMCmean = 0.093, HPD95lower = 0.050, HPD95upper = 0.134, $p < 0.001$, Figure 31; Proficiency, Sentence Language and Format: MCMCmean = -0.053, HPD95lower = -0.106, HPD95upper = -0.004, $p = 0.043$, Figure 32) to make a fast L2 decision (see Table E5, Appendix C).

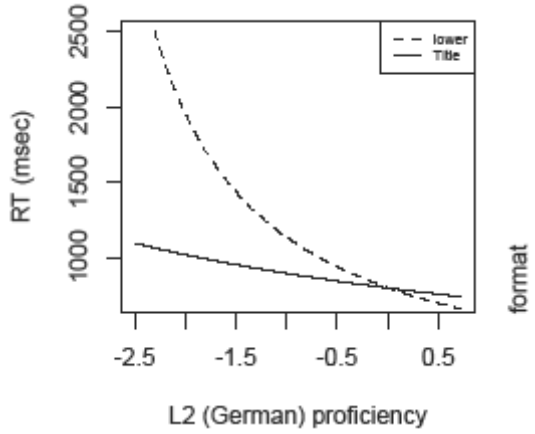


Figure 30. Mean response times for words, in relation to the stimulus' presentation format and L2 (German) proficiency of speakers.

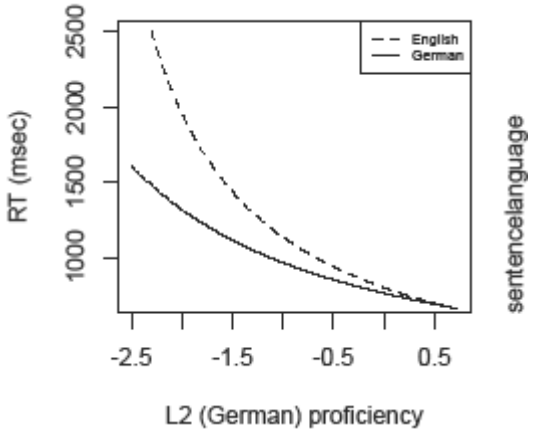


Figure 31. Mean response times for words, as defined by the stimulus' preceding sentence language and L2 (German) proficiency of speakers.

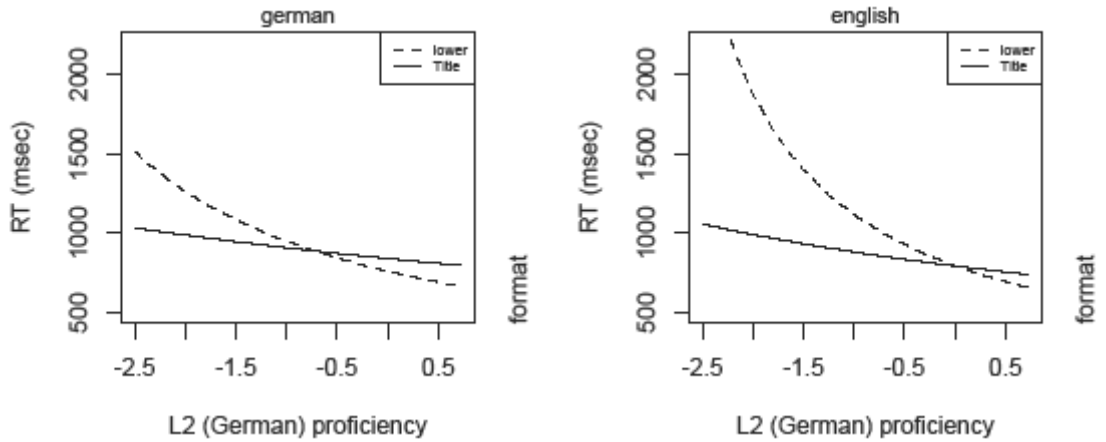


Figure 32. Mean response times for words, by L2 proficiency, stimulus format (lower vs. onset capital) and preceding sentence language (German, in the left panel, and English, in the right panel).

In this regard, low proficiency participants' responses are considerably slowed down when no language-specific cues are provided. In addition, this effect seems particularly pronounced after English (L1) sentence frames. These findings are consistent with the previous error analysis (and to a great extent with my predictions), indicating that for this group of speakers a stimulus' presentation format may be a fundamental source of information in the decision making process. In contrast, more proficient bilinguals' responses become faster (and more accurate), which rules out the possibility of a

speed/accuracy trade off and indicates that these speakers make different use of the available resources.

As part of the same effect, Sentence Language (MCMCmean = -0.097, HPD95lower = -0.121, HPD95upper = -0.074, $p < 0.001$) and its interaction with Format (MCMCmean = 0.125, HPD95lower = 0.098, HPD95upper = 0.155, $p < 0.001$) are also revealed as conventionally significant.

The data also reveal that English-German homographs are identified more quickly than controls (Item Type: MCMCmean = 0.041, HPD95lower = 0.004, HPD95upper = 0.075, $p = 0.021$, Figure 33), especially when following an L2 (German) sentence (Item Type and Sentence Language: MCMCmean = 0.036, HPD95lower = 0.008, HPD95upper = 0.061, $p = 0.011$, Figure 34).

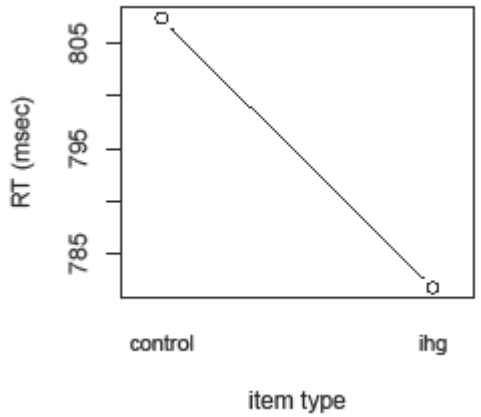


Figure 33. Mean response times for IHGs and controls.

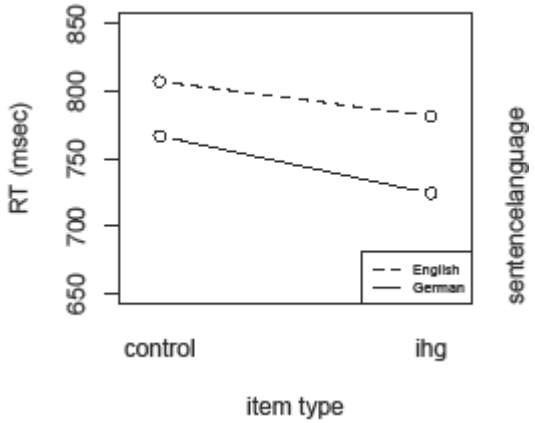


Figure 34. Mean response times for IHGs and controls, depending on the preceding sentence language (German or English).

Accordingly, a task-relevant sentence language (i.e. German) not only allows participants to make a generally faster response, but it is particularly effective in reducing cross-linguistic interference (i.e. the advantage for IHGs over controls is greater in the German sentence language condition than in the English condition). The latter finding is largely in line with previous work: while it supports the idea of zooming-into the language of the task (Elston-Güttler, Gunter, et al. 2005), it also suggests that a semantically unrelated sentence context may be somewhat able to "activate the target [lexical representation] allowing it to compete early for selection" (Schwartz & Kroll 2006:209; but see Libben & Titone 2009, for contrasting results). Notably, the absence of an interaction with Format does not imply that the presentation format had no effect on participants' responses, but rather that responses to either type of word stimulus may have been similarly affected by the presence or absence of a language-specific cue.

This conjecture will be further explored in the separate analyses of IHG and control data below.

3.2.3.3.1.2 Other variables

More frequent (MCMCmean = -0.067, HPD95lower = -0.090, HPD95upper = -0.045, $p < 0.001$, Figure 35) and longer words (MCMCmean = -0.021, HPD95lower = -0.037, HPD95upper = -0.003, $p = 0.018$, Figure 36) are identified more quickly. The latter finding (i.e. the length effect) seems clearly surprising in light of previous research (cf. Ziegler et al. 2001), yet it is also found in the previous analyses of error rates. A tentative explanation of this finding is that shorter words have more neighbours; hence word recognition may have been slowed down due to greater lexical competition than for longer words. This was supported by the high correlation between word length and neighbourhood density (German neighbours: $r = -.31$, $p < 0.001$; English neighbours: $r = -.62$; $p < 0.001$).

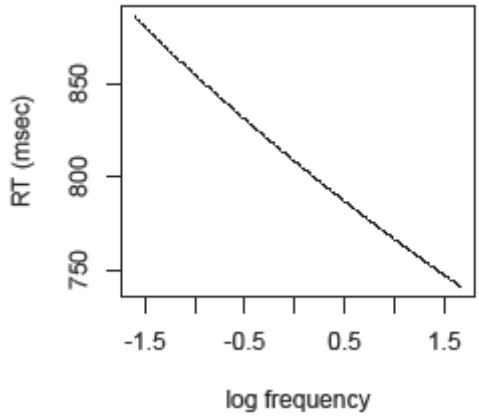


Figure 35. Mean response times for words, by increasing log frequency.

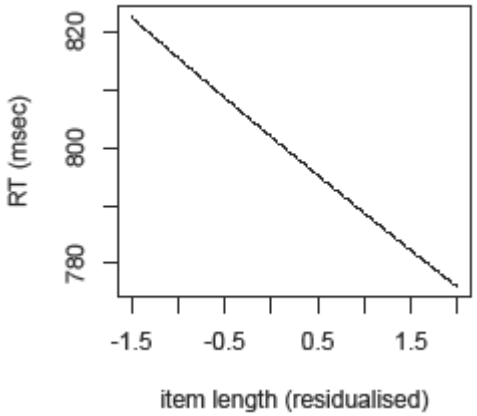


Figure 36. Mean response times for words, by increasing stimulus length.

Another simple effect found in the data is that of Session, with responses in the second experimental session being made more quickly (MCMCmean = -0.131, HPD95lower = -0.149, HPD95upper = -0.114, $p < 0.001$; Figure 37). This may again point to a possible familiarisation with the task.

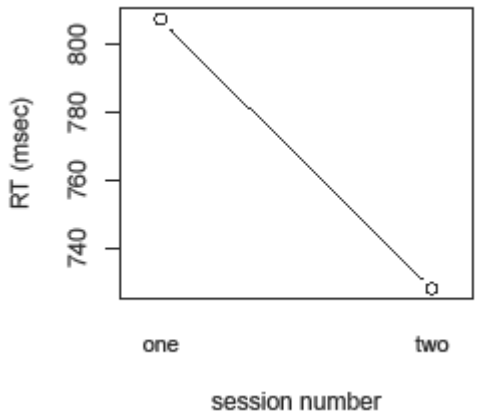


Figure 37. Mean response times for words, depending on the experimental session.

3.2.3.3.2 IHG data

The separate analysis of English-German homographs explores further the impact of the main variables of interest (i.e. Format, Sentence Language and Proficiency) on the activation of the target reading of the IHG. As with the previous analysis, a linear mixed-effects model was fitted to the data, including the predictors discussed in section 3.2.2.3 (with the exception of Item Type).

3.2.3.3.2.1 Main predictors

The present analysis returned the same simple effects compared to the entire word set (Proficiency: MCMCmean = -0.377, HPD95lower = -0.503, HPD95upper = -0.244, $p < 0.001$; Sentence Language: MCMCmean = -0.109, HPD95lower = -0.136, HPD95upper = -0.081, $p < 0.001$), as well as similar interactions (Proficiency and Format: MCMCmean = 0.234, HPD95lower = 0.078, HPD95upper = 0.392, $p < 0.001$, Figure 38; Proficiency and Sentence Language: MCMCmean = 0.069, HPD95lower = 0.036, HPD95upper = 0.104, $p < 0.001$, Figure 39; Format and Sentence Language: MCMCmean = 0.146, HPD95lower = 0.106, HPD95upper = 0.183, $p < 0.001$, Figure 40). The only difference found was in the absence of a three-way interaction between Proficiency, Format and Sentence Language (see Table E6, Appendix C).

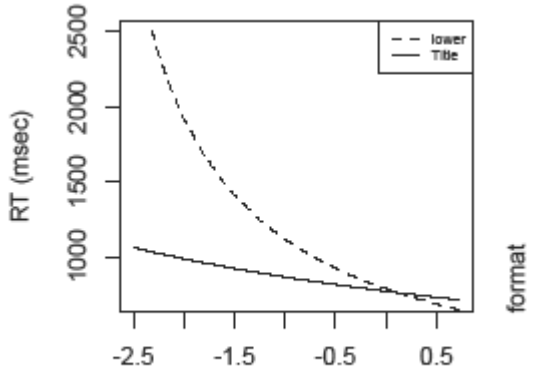


Figure 38. Mean response times for English-German homographs, by increasing L2 proficiency and presentation format.

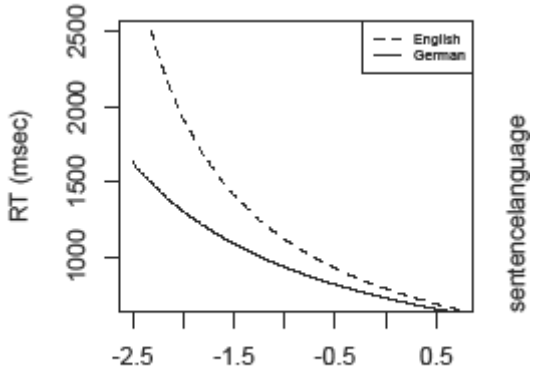


Figure 39. Mean response times for IHGs, by increasing L2 proficiency and preceding sentence language.

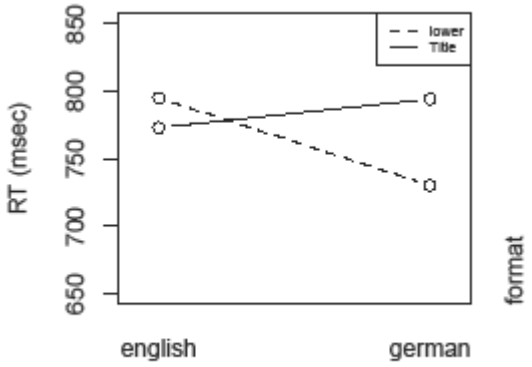


Figure 40. Mean response times for English-German homographs, in relation to the preceding sentence language and stimulus' presentation format.

Looking at Figure 40 above, the interaction between Sentence Language and Format is fairly surprising because it is counter-intuitive and seems to contradict my previous predictions and outcomes. That is, while response latencies after an English sentence are not particularly affected by the presence or absence of a language-specific cue, responses following a German sentence frame are considerably faster when no such language-specific information is available. To better understand this outcome, some additional analyses were performed. Given that onset capitalisation occurs in German only for nouns, one such analysis tested whether the present effect was a result of the not entirely balanced congruency in syntactic category across languages (e.g. Baten et al. 2011; Sunderman & Kroll 2006; for a more extensive discussion in relation to this study, see also the relevant error analysis); however, the obtained results were identical to the present outcomes. The other two analyses looked separately at responses in each of the two language contexts, and in the first experimental session only (to disregard possible familiarisation effects). Both analyses returned a statistically significant interaction of

Proficiency and Format. However, the effect was more pronounced following a German sentence. More specifically, less proficient bilinguals' responses were generally slower when no orthographic cues were provided, but this inhibitory effect was considerably less pronounced after a German sentence. On the other hand, with increasing L2 proficiency participants' responses were less strongly affected by the absence of a language-specific cue. In fact, responses were considerably facilitated in the absence of such information, and more clearly so when following a German sentence (see the three-way interaction in Figure 32, for an illustration of this effect). In light of these additional analyses, it appears that depending on the participants' proficiency level, language-specific cues (or the absence of them) may play a different role in cognitive processing (see the discussion section below for more detail), and that the combination of these effects together is reflected in the interaction between Sentence Language and Format.

3.2.3.3.2.2 Other variables

The only other variables found to have a significant effect in the final model are Frequency (MCMCmean = -0.057, HPD95lower = -0.092, HPD95upper = -0. 021, $p <$ 0.001) and Session (MCMCmean = -0.156, HPD95lower = -0.176, HPD95upper = -0.136, $p < 0.001$). Notably, response times were lower in the second session when compared to the first.

3.2.3.3.3 German control word data

As a complement to the separate analysis of IHGs, the following analysis focused on responses to (German) controls only. A linear mixed-effects model was fitted to the data, including the predictors indicated in section 3.2.3.3.2.

3.2.3.3.3.1 Main predictors

As in the main word data set, Proficiency is both a significant simple effect (MCMCmean = -0.352, HPD95lower = -0.478, HPD95upper = -0.237, $p < 0.001$), and involved in various interactions with the remaining main predictors (Proficiency and Format: MCMCmean = 0.231, HPD95lower = 0.080, HPD95upper = 0.373, $p < 0.001$; Proficiency and Sentence Language: MCMCmean = 0.094, HPD95lower = 0.034, HPD95upper = 0.153, $p < 0.001$; Proficiency, Format and Sentence Language: MCMCmean = -0.084, HPD95lower = -0.152, HPD95upper = -0.011, $p = 0.021$, Figure 41). The direction of these effects was the same as reported for the entire data set (compare Table E7 with Table E5, Appendix C).

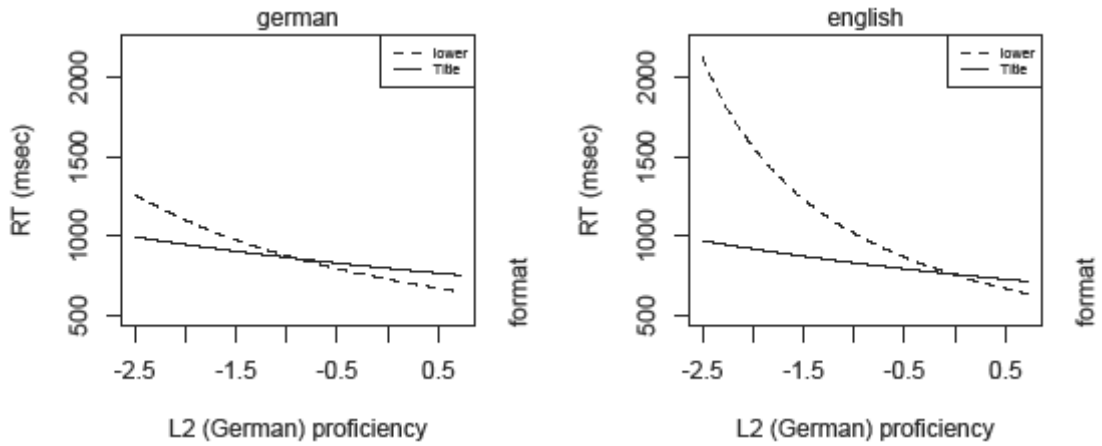


Figure 41. Mean response times for German controls, by L2 proficiency, stimulus format (lower vs. onset capital) and preceding sentence language (German, in the left panel, and English, in the right panel).

Sentence Language was also found as a simple main effect (MCMCmean = -0. 058, HPD95lower = -0. 085, HPD95upper = -0. 030, $p < 0.001$), as well as in an interaction with Format (MCMCmean = 0.119, HPD95lower = 0.081, HPD95upper = 0.160, $p < 0.001$). The present effects replicate the findings reported in the previous analyses discussed above.

3.2.3.3.3.2 Other variables

The only other simple effects returned as conventionally significant are Frequency (MCMCmean = -0.072, HPD95lower = -0.099, HPD95upper = -0.045, $p < 0.001$) and Length (MCMCmean = -0.038, HPD95lower = -0.060, HPD95upper = -0.016, $p < 0.001$; for details, see discussion in the first analysis, section 3.2.3.3.1).

3.2.3.3.4 Nonword data

The fourth and final analysis concerned the speed of nonword rejections. This was done to extend our current understanding (van Heuven et al. 1998; Lemhöfer & Dijkstra 2004; Lemhöfer & Radach 2009) of the type of information (i.e. bottom-up, top-down, both or none) used by bilinguals during the processing of nonwords. Similar to the previous analyses, a linear mixed-effects model was fitted to the data which included correct responses to nonword tokens only. The same (main and secondary) predictors were included in the fitted model, with the exception of Item Type.

3.2.3.3.4.1 Main predictors

The analysis reveals that the speed with which items can be rejected as non-existing L2 words depends on the speakers' L2 proficiency level (becoming faster with increasing L2 proficiency). In the analysis, this is reflected in the simple effect of Proficiency

(MCMCmean = -0.212, HPD95lower = -0.282, HPD95upper = -0.142, $p < 0.001$; Figure 42).

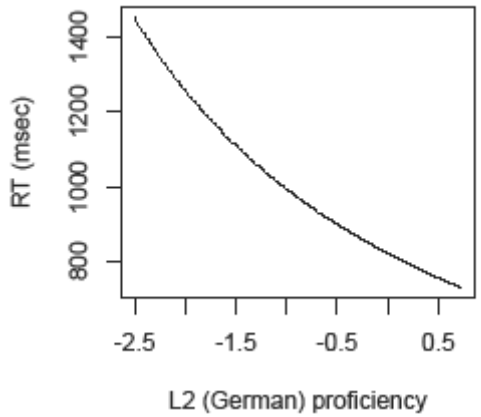


Figure 42. Mean rejection times for nonwords, with increasing L2 (German) proficiency.

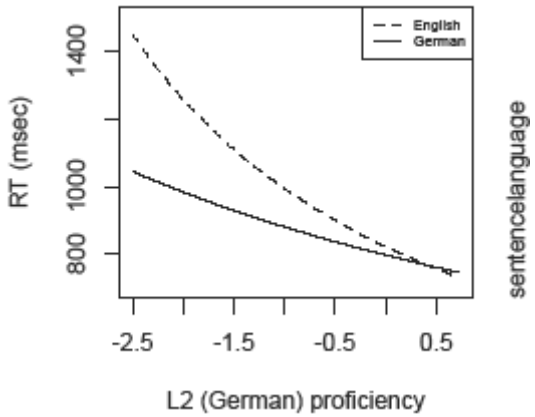


Figure 43. Mean rejection times for nonwords, in relation to the preceding sentence language and participants' L2 proficiency.

In addition, for lower proficiency learners this decision is further influenced by the language of the sentence preceding the target, with slower responses after a task-irrelevant sentence frame (i.e. English; Proficiency and Sentence Language: MCMCmean = 0.092, HPD95lower = 0.071, HPD95upper = 0.113, $p < 0.001$; Figure 43).

The analysis further reveals that while rejections of nonword tokens are generally faster after an L2 (task-relevant) sentence frame (Sentence Language: MCMCmean = -0.038, HPD95lower = -0.055, HPD95upper = -0.021, $p < 0.001$; Figure 44), within this L2 sentence condition German-biased items in Title case take longer to reject than nonwords presented all in lower case (Sentence Language and Format: MCMCmean = 0.067, HPD95lower = 0.042, HPD95upper = 0.090, $p < 0.001$; Figure 45).

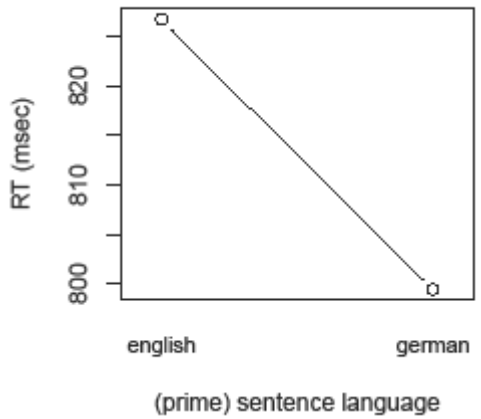


Figure 44. Mean rejection times for nonwords, as defined by the preceding sentence language.

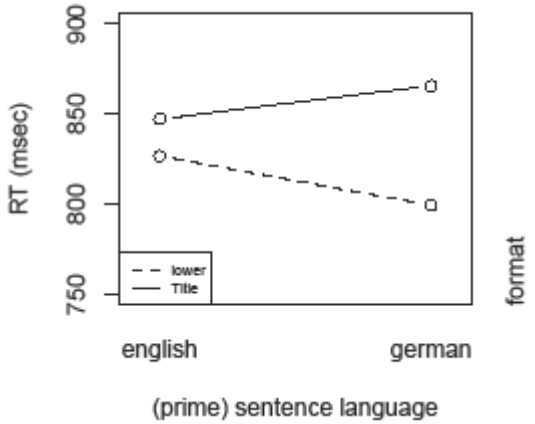


Figure 45. Mean rejection times for nonwords, as defined by the preceding sentence language and stimulus presentation format.

One aspect of bilingual visual processing that can be conjectured from these findings is that simultaneous processing of two languages (made explicit by the experimental setup) imposes an additional cost on the speed with which a decision can be made (e.g. von Studnitz & Green 1997). In addition, less proficient bilinguals seem more strongly affected by the inclusion of the L1 in the experiment, which may indicate that these speakers experience stronger L1 interference and/or exhibit less effective cognitive control. The other finding, which seems in line with the majority of outcomes reported above, is that sub-lexical cues are certainly employed as a "language discrimination mechanism" (Lemhöfer & Dijkstra 2004:547) in lexical processing. Notably, however, the language of the sentence frame seems also to be a major language determining factor, although given the specifics of the present task, it is perhaps not the primary one (see general discussion for further detail).

3.2.3.3.4.2 Other variables

With respect to the secondary variables, four simple effects were returned as significant: the Bigram Frequency and the Length of an item (Bigram Frequency: MCMCmean = 0.000, HPD95lower = 0.000, HPD95upper = 0.000, $p < 0.001$, Figure 46; Length: MCMCmean = 0.066, HPD95lower = 0.049, HPD95upper = 0.083, $p < 0.001$, Figure 47), as well as the number of German orthographic representations in the stimulus' neighbourhood (MCMCmean = 0.020, HPD95lower = 0.015, HPD95upper = 0.026, $p < 0.001$, Figure 48) and whether participants attended their first or second session (suggesting familiarisation with the task; MCMCmean = -0.145, HPD95lower = -0.166, HPD95upper = -0.123, $p < 0.001$, Figure 49).

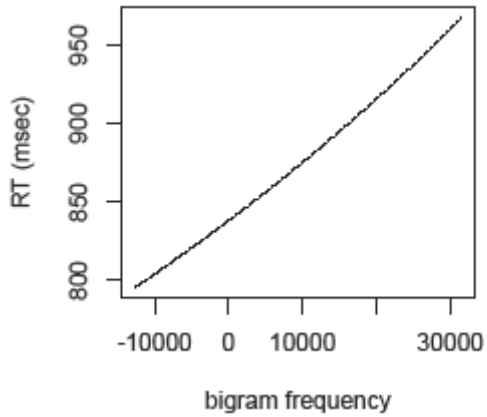


Figure 46. Mean rejection times for nonwords, with increasing bigram frequency of the stimulus.

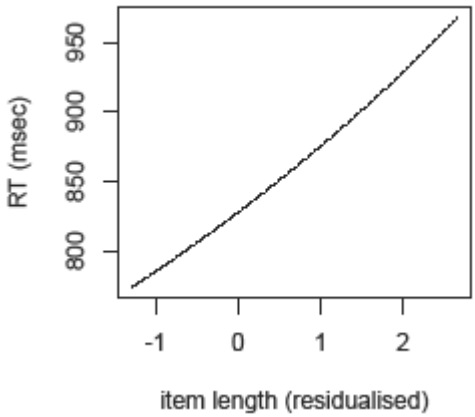


Figure 47. Mean rejection times for nonwords, with increasing stimulus length.

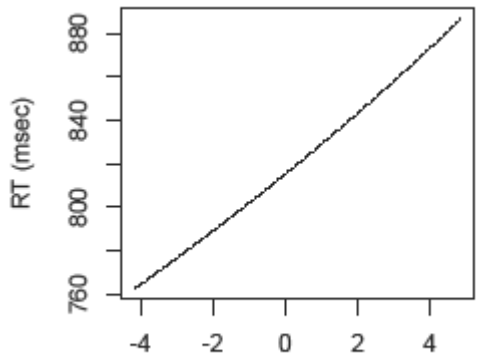


Figure 48. Mean rejection times for nonwords, with an increasing number of L2 (German) neighbours.

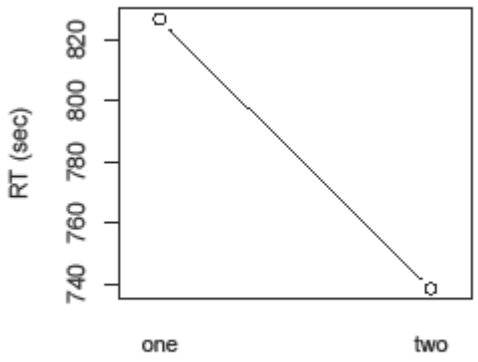


Figure 49. Mean rejection times for nonwords, depending on experimental session.

Whereas the first two effects have been repeatedly observed in other psycholinguistic studies (see Ziegler et al. 2001, concerning the length effect in German versus English native speakers), the other two effects are not surprising either. Overall, these results suggest that the more potentially confusable a nonword token is with existing German words, the more difficult it is to reject. These findings are in line with the previous analyses and discussion (e.g. Lemhöfer & Radach 2009; but see also section 3.2.3.2 above).

3.2.3.4 Discussion of the RT analysis

The response time analyses of only the correct responses in the data confirmed the general patterns obtained in the error rate analyses, providing a more comprehensive insight into (some of) the processes involved during bilingual cognitive processing. One major finding of the present study confirms the argument put forward by other researchers, namely that a task-relevant sentence language (here, German) can be used by bilingual participants to considerably reduce the effects of cross-linguistic interference (e.g. Elston-Güttler, Gunter, et al. 2005; Libben & Titone 2009; Schwartz & Kroll 2006). While there are uncertainties about the precise extent to which a task-relevant sentence can be a major processing facilitator, the present study addresses this issue by identifying further factors which may not only display a similar effect, but also interact with other linguistic information provided in the experiment.

One such factor is the bilinguals' proficiency in their L2. Unlike the claim made by Chambers and Cooke (2009), the present data suggest that proficiency may be perhaps the most crucial factor in determining the success with which cross-linguistic

interference can be reduced. In both word and nonword responses, this factor has been repeatedly shown as having an impact not only on the speed with which L2 input is being processed, but also the type of resources employed to execute the task and the manner in which this is done.

Accordingly, less proficient bilinguals' (real word) responses are affected by both the availability of a task-relevant sentence language, as well as the presence or absence of a language-specific cue. Largely in line with my predictions, the findings suggest that less proficient bilinguals may be employing sub-lexical information as a key source for their language decision. This is reflected in the finding that response latencies to words with an onset capital letter are not affected by the sentence language preceding the target (i.e. the fastest and most accurate responses across all analyses are found to real word targets with onset capitals). As discussed above, this finding is in line with the proposal of an early language determining mechanism. The present data show that such a mechanism may operate by means of the bottom-up activation of something like a language node on the basis of low-level sub-lexical information, with this language node then exercising a top-down influence on lexical activation (e.g. Dijkstra 2005; Lemhöfer & Dijkstra 2004; see discussion of error analyses).

In contrast, when no language-specific sub-lexical information is provided and the sentence language conflicts with the language of the task, low proficiency participants' responses are considerably slowed. Here the data suggest that a preceding L1 sentence elevates the activation levels of the non-target (L1) language, resulting in higher cross-linguistic competition. With no other language information to resolve lexical competition, L2 word recognition is considerably inhibited. A similarly detrimental effect of sentence language is also found for responses following a task-relevant (L2) sentence, however only when no language-specific orthographic cues are provided. That is, (in the form of a speed/accuracy trade off) a preceding German sentence seems to elevate participants' response levels high enough for them to be more inclined to quickly accept an all lower case item as a German word, while at the same time the error data show that they more often than not reject existing words. This intriguing finding seems to imply that less proficient bilinguals are susceptible to both types of language information (i.e. sentence language and language-specific cues) as their language navigator. In addition, while it could be argued that language-specific cues (or rather the lack of them) are the primary source for the lexical decision made by these participants,

it is clear that other information may also be incorporated in the decision making process (see Elston-Güttler, Gunter, et al. 2005; Libben & Titone 2009; Schwartz & Kroll 2006, for suggestions of an interaction of information between different processing levels and/or systems).

As seen in the error rate analyses above, with increasing proficiency bilingual participants seem less clearly affected by the language of the sentence preceding a target. Consequently, the arguments that were put forward in the error analysis to explain the results seem similarly applicable here. That is, even though this was not explicitly measurable, the current results could be taken to imply that with increasing proficiency level L2 sentences are somewhat easier to process (e.g. Gernsbacher 1993, 1997; Gernsbacher & St. John 2001). Consistent with previous research (e.g. Chambers & Cooke 2009; Duyck et al. 2007; Elston-Güttler, Gunter, et al. 2005), the present results could also be taken as evidence that the availability of an L2 sentence language can considerably reduce, if not eliminate, lexical activation of the non-target language, even though the sentence may be unrelated in meaning to the target (see Schwartz & Kroll 2006, experiment 1; Libben & Titone 2009, for a recent debate, based on contrasting results from both these studies). An unexpected, but highly interesting pattern which is revealed in the L2 context is that more advanced bilinguals' responses are faster when the presented words do not carry an onset capital letter. One implication of this finding is that bilingual speakers may be sensitive towards sub-lexical language-cues, irrespective of their L2 proficiency. However, if we assume that with increasing proficiency lexical processing in an L2 becomes more automatised (Schulpen 2003:201–202; S. J. Segalowitz et al. 1998), we can argue that more proficient bilinguals do not necessarily employ this kind of information to activate one of their different languages (at least not in a top-down way, cf. Dijkstra 2005:187).
Another implication of the finding may be that the L2 sentence established some expectation regarding the underlying characteristics of the following target item (e.g. syntactic category, presentation format). As a result, an accessed German word (carrying an onset capital letter) required an additional consistency verification to ensure that the word was a noun (which requires capitalisation). This conjecture seems to be confirmed by the absence of a similar effect in the responses following an L1 context (see discussion of error analysis for further interpretations of this result).

However, given that the number of highly proficient bilinguals in this study was relatively small, more research is needed to come to definite conclusions.

By and large, the speed with which nonwords are accurately rejected seems to support the previously outlined assumptions that both bottom-up (e.g. sub-lexical) and top-down (e.g. sentence language) cues affect lexical processing; however, the more immediate source of information may be the orthographic cues (as reflected in the interaction between Sentence Language and Format). In this context, the effect of Format seems less pronounced after an L1 (i.e. English) sentence, but quite marked after an L2 (German) sentence, i.e. in the context where some expectation regarding Format might have been established. Recent research provides a useful framework to understand these findings. That is, we can consider the present results in context of the MROM (the model deals explicitly with nonword rejections in an LDT; Grainger & Jacobs 1996), where it is assumed that bilingual speakers reject nonwords based on meeting a temporal deadline which can be determined by either the language or task context (Lemhöfer & Radach 2009:46). The pattern of findings is that in single-language tasks, nonwords that are *dissimilar* to the language of processing are rejected more quickly and produce fewer errors. In mixed-language tasks on the other hand, some researchers argue that nonwords in the participants' *weaker* language seem more difficult to reject (Lemhöfer & Dijkstra 2004; Lemhöfer & Radach 2009), and others argue the opposite, i.e. that responses to nonwords in the *stronger* language are more difficult (van Heuven et al. 1998, experiment 3). The first conjecture seems to find support in the present analysis, if we assume that the preceding German sentence accentuated the 'single-language' nature of the task. If we further assume that sub-lexical cues can pre-activate the relevant language node (i.e. provide input to a language discrimination mechanism, Lemhöfer & Dijkstra 2004:547), then nonwords that do not exhibit the language-specific feature (i.e. which are presented all in lower case) should be inhibited more quickly and therefore induce fewer errors. Similarly, if we try to explain the findings on the basis of the summed lexical activity, we must assume that nonwords with an onset capital (i.e. more German-like) create *higher* levels of summed lexical activity in the most activated (i.e. German/L2) lexicon than nonwords presented all in lower case, making the nonwords that carry features of the task language (i.e. L2) more difficult to reject.[45]

[45] Given that for some participants at least the early stages of the experiment would have been single-language tasks (i.e. in the first experimental session, with target presentation following a German

In contrast, neither of the previous findings seems to find clear support in our data, if we assume that an English sentence highlighted the 'mixed-language' nature of the task. This, however, may simply reflect the fact that cognitive processes involved in the 'mixed' condition are more complex (and more difficult to decipher), given that they are affected by a larger number of different factors.[46]

sentence), I conducted a separate analysis of a correspondingly reduced data set. The analysis revealed a significant simple effect of Format and a significant interaction between Proficiency and Format (with more errors made by lower proficiency bilinguals to items with an onset capital letter). Although this result has to be taken with caution (due to a substantially reduced data set), it not only provides further support for the proposed 'language-filtering' mechanism, but it also extends previous research by showing that this effect is more accentuated for lower proficiency bilinguals.

[46] I am aware that despite the inclusion of ambiguous material (IHGs) and English sentence frames in some conditions, participants were asked to make their decision based on one language only, namely their L2. Accordingly, this may not make the type of the current task as clearly categorisable as a mixed-language task (i.e. a generalised LDT, typically used in the mentioned research above); hence, the lack of support for either of the previous findings.

3.2.4 Summary of Experiment One

To summarise, there are several major findings emerging from both the error and latency analyses conducted above. First, Experiment One provides strong evidence that bilingual speakers performing in their less dominant language (L2) are susceptible to sub-lexical language-specific cues. The impact of these cues, however, depends highly on the speakers' L2 proficiency: less proficient bilinguals seem to employ the sub-lexical level information as a key source for their language decision, perhaps by activating the relevant language node which can exercise influence on lexical activation from above (i.e. in a top-down manner). With increasing proficiency, however, language-specific cues appear to be employed in a different way, i.e. indicating the syntactic category of an item. While this has no impact on the accuracy of responses, it slows down the recognition process to some extent.

Second, sentence language was also shown to influence the bilingual word recognition system. Although only evident for less proficient bilinguals, a task-irrelevant L1 sentence language seems to elevate the activation levels of this non-target language, resulting in increased lexical competition. A task-relevant (L2) sentence language has a similarly detrimental effect on these lower proficiency speakers when no sub-lexical language information is provided. More specifically, there appears to be a trend towards a speed-accuracy trade-off where participants are inclined to make a fast response after an L2 sentence, more often than not rejecting existing words with an all lower case.

Finally, both analyses identify various factors affecting the processing of IHGs. The main factors include the L2 proficiency of a bilingual speaker and the congruency of syntactic category across the two languages.

To shed more light on the features discussed here, an identical experiment was conducted with a German-English bilingual population (i.e. in the participants' L1). The following sections discuss any necessary changes to the experimental setup, as well as the obtained results and implications.

3.3 Experiment Two: L1 Lexical Decision Task (German-English Bilinguals)

3.3.1 Methodology

3.3.1.1 Participants

Sixty-nine native speakers of German took part in this experiment. Participants were students enrolled at Ludwigs-Albert University of Freiburg, the St. Ursula Gymnasium and the Kepler Gymnasium in Freiburg. They all had normal or corrected-to-normal vision and no hearing impairment. Participants had varying knowledge of English, and were selected to represent a wide range of proficiency in their L2. In most cases, a basic selection criterion was their university course enrolment and school year. To test their English knowledge individually, and in order to acquire other relevant information, I asked each participant to fill out a language questionnaire and to complete an English language proficiency test after they had completed the experiment (for detailed discussion of both assessment types, see section 3.2.1.4). All participants signed a written consent form. In recognition of their participation, participants went into a draw for one of three major prizes.

3.3.1.2 Materials and Procedure

Materials and procedure were almost identical to Experiment One. All participants completed two experimental sessions which were approximately seven days apart. This time frame was chosen to minimise possible learning effects of a sub-group of chosen experimental items (i.e. interlingual homographs; discussed below). Participants were seated in a spacious office, with standard lighting, approximately 60 cm away from a standard desktop monitor (Dell E190S 19-inch Flat Panel LCD Monitor; Dell OptiPlex 755 Desktop, Intel® vPro™ processor technology). The procedure, instructions, and presentation orders were identical to those of Experiment One (see section 3.2.1.3 above). The only slight deviations from the procedures involved in Experiment One concerned the post-tests. First, participants were required to give the German meanings of the *English* words represented by the IHGs in the experiment (e.g. for HOSE a correct response would be 'Schlauch', meaning 'pipe'). Further, participants carried out the English version of the proficiency test (adopted from Lemhöfer et al. 2004). Finally, participants were asked to fill in a language history questionnaire (adapted from Li et al. 2006), which was sent to them by email, and to return it in an agreed way (i.e. chosen by the participant) within two weeks. The latter was done to ensure that the duration of an experimental session did not exceed the indicated time. Unfortunately, only 81% of

questionnaires were returned to the researcher in due course. This was not a major concern, since the original purpose of the questionnaire was merely to collect some additional information about the participants' language background, and as a backup for the language proficiency test. However, given the large proportion of unreturned questionnaires, I decided not to use this data source in any of the statistical analyses conducted in Study One; instead, proficiency data are based exclusively on the on-line test. (Note that this is also true in Study Two, for the same reasons.) The same researcher (the author) conducted all sessions, and the language of communication between researcher and participant was English. This should not have been suspicious, since the researcher was introduced as coming from New Zealand, and there was no mention of her being a German native speaker.

3.3.1.3 Language proficiency assessments

English proficiency test

In parallel to the German proficiency test (section 3.2.1.4), the English version of the test was a simple off-line vocabulary test requiring participants to decide whether a presented string of letters formed an English word or not, by pressing one of two buttons.

The English proficiency test developed by Lemhöfer and colleagues (2004) was a shortened version of Meara's (1996) vocabulary test. Lemhöfer et al.'s (2004) test consisted of 60 items, including both words and nonwords, in a ratio of two-to-one (for the rationale behind this selection, see discussion in section 3.2.1.4). All items selected by Lemhöfer and colleagues (2004) were employed in the current research. As reported by the researchers, items were between 4 and 12 letters long (mean stimulus length: 7.3) and the calculated frequency of the selected word stimuli ranged from 1 to 26 occurrences per million (mean word frequency: 6.4 o.p.m.). The latter calculation was based on information in the CELEX database (Baayen et al. 1993). All items are listed in Appendix B.

The order of stimulus material, instructions and procedure for the English proficiency test were identical to those for the German test.

Language questionnaire

The questionnaire was identical to the one given to the English L1 participants (see section 3.2.1.4 above), except that some questions had the labels 'English' and 'German' reversed, given the inverted L1 of the participants.

3.3.2 Statistical analysis

The statistical approach (including the analysis type, the procedures and the chosen predictors) employed for this data set is identical to the statistical analysis conducted in the previous experiment (Experiment One; for a detailed description, see section 3.2.2 above).

The initial data set of Experiment Two (i.e. data collected in Germany, from German-English bilinguals) comprised 21528 observations. Prior to data analysis, one participant had to be excluded since she did not follow the given instructions. This resulted in the loss of 312 data points. Further data exclusions were as follows: first, responses that were faster than 300ms were removed, since they would not capture the required cognitive processing (cf. Dijkstra et al. 2010). Second, data from one more participant was excluded. This participant performed the experimental task with a considerably higher error rate (greater than 25% on the entire data set, i.e. including IHGs, controls and nonwords) than any other participant. Finally, four experimental items (with the same cutoff point as in Experiment One) were removed since they elicited high error rates.[47] The remaining data set contained 20614 observations.
For ease of reading and interpretation, the following results are presented and discussed based on plots of lmer model estimates (back-transformed for RTs) and standard errors in the relevant conditions.

3.3.3 Results and discussion

The aim of this study was to explore the impact of sub-lexical information (i.e. in the form of language-specific cues) on bilingual visual processing. The task was conducted in the native language (German) of unbalanced bilingual German-English speakers. As previously, the impact of such information was assessed with respect to the language of a prime sentence, as well as the proficiency level of the bilingual in their L2.

[47] The group of eliminated items contained one IHG and two German controls. Items in the matching presentation format were also excluded, even though they may not have produced the same number of errors (e.g. although 'smoking' produced a high number of errors, but 'Smoking' did not, both items were excluded from the analysis). Data loss due to item characteristics totalled only 1.3%.

The main prediction corresponded largely to my expectations in the previous experiment (Experiment One). That is, if language-specific cues can act as early target language navigators (or a "language discrimination mechanism"; cf. Lemhöfer & Dijkstra 2004:547) then responses to real words with onset capitals should be facilitated, because of the reduction of the activation levels of the non-target language. In contrast, rejections of nonwords with an onset capital should be inhibited, given their greater similarity to existing German words when carrying a language-specific cue.

Clearly, both predictions need further refinement. One such refinement may be made when considering the L2 proficiency of the bilinguals involved in this experiment. First, given that the task is conducted in the participants' L1 and considerably automatised processes are expected to take place for all participants, we might not observe any differences in responses to existing L1 words that are related to proficiency in L2. However, less automatic processing of input (i.e. when responding to nonwords) might reveal a different picture. That is, if the manner in which linguistic input is processed changes with increasing proficiency in an L2 (for evidence, see Experiment One) and if this is irrespective of the language of processing, then we might observe greater reliance on the presentation format in less proficient bilinguals.

Another refinement of the predicted effects concerns the language of the sentence preceding the target. As illustrated in the introduction to this study, traces of L2 interference should be fundamentally limited (if measurable at all) following an L1 sentence frame (cf. Elston-Güttler, Gunter, et al. 2005), but measurable interference should be observed for L2 sentence frames (cf. Titone, Libben, Mercier, Whitford, & Pivneva 2011). In addition, both effects may be strongest in a condition where all aspects of the linguistic context support the (L1) task demands (i.e. L1 sentence frame, followed by the Title case presentation format).

Finally, it needs to be noted that given the nature of the task (i.e. processing is in the participants' L1, and hence is more automatic), the real word predictions may only be revealed in the responses to interlingual homographs (IHGs). This is because only IHGs have an alternative reading in the non-target language, and may therefore not only be processed less automatically, but also compete for selection during the recognition process (e.g. van Heuven et al. 2008). In addition, responses to IHGs may also be most likely (if not exclusively) to reveal any effects of L2 proficiency.

Given these predictions, the same main predictors were included in the most basic mixed effects regression model – for both the RT and error data – i.e. Item Type, Format,

Sentence Language and Proficiency. Subsequent models also included other factors, such as Frequency, Length, Word Class, the number of competing neighbours in both languages (i.e. German and English), Session number, Trial number and the response latency and accuracy to a Preceding Trial (for rationale, see previous discussion).
All analyses were first conducted on the word set as a whole (i.e. including both the German-English homographs and their German control words), followed by separate analyses of the critical stimuli (IHGs), and of the controls. Responses to nonwords were analysed separately. The mean response latencies on correct trials and percent error rates are summarised in Table 4. The calculation of the mean values presented in this table followed the initial data trimming described in section 3.3.2 above, but preceded the statistical modelling reported below. Hence, they merely give a general pattern of response latencies and error rates.

Table 4.
Mean recognition latencies to correct responses (in milliseconds) and percent error rates (in parentheses) for IHGs, controls and nonwords in L1 (German) and L2 (English).[48]

Item Type	German (L1)		English (L2)	
	lower	Title	lower	Title
IHG	622 (5.51%)	612 (3.68%)	690 (7.44%)	682 (4.89%)
control	623 (3.04%)	624 (3.08%)	661 (5.36%)	669 (4.22%)
nonword	676 (1.64%)	701 (2.56%)	727 (1.45%)	764 (2.45%)

3.3.3.1 Error Analysis

3.3.3.1.1 Responses to IHGs and German control words

Overall, the error rate in this experiment was much lower compared with Experiment One, presumably because the task was conducted in the participants' L1. The following error analysis included both types of existing words, namely German-English homographs and their matched control words. A statistical (mixed logit) model was fitted to the data, including the specific procedures and predictors indicated in section 3.3.2.

[48] In Appendix C (Table 15), I present the same data split into high and low proficiency groups (as determined by a median split value), as a better illustration of the relationship between RTs and accuracy for these bilinguals.

3.3.3.1.1.1 Main predictors

The analysis of error rates revealed only two simple effects and no interactions between the identified main predictors (see Table G1, Appendix C). (Note that the values displayed in the figure(s) are predicted values from the model, with other factors set at their intercept value. This is why subsequent effects may differ from the values in Table 4 above). One simple effect is that of Sentence Language, with fewer errors to words following an L1 (German) sentence frame than following an L2 sentence frame (β = 0.457, SE = 0.107, z = 4.290, $p < 0.001$; see Figure 50). While no context studies to date have investigated the effects of the language of a sentence on L1 processing, a similar effect has been reported in a number of single-word processing studies (e.g. von Studnitz & Green 1997; Thomas & Allport 2000; see also Titone et al. 2011, for the first study on L1 sentence reading).

The other simple effect is Proficiency, which reveals that the lower a participant's proficiency in the L2 the more errors they made when responding to existing words in their L1 (β = 0.641, SE = 0.199, z = 3.223, $p < 0.001$). A possible explanation of this finding is that lower proficiency L2 speakers struggle more with the experimental task demands, i.e. they exhibit less control over their two languages, which leads to more errors (see Figure 51). Another possibility is that this result reflects some more general issues or psychological factors related to less proficient bilinguals, such as poor cognitive control, short term memory, language aptitude, and so forth (see Mitchell & Myles 2004:24–26, for a brief overview of identified factors).

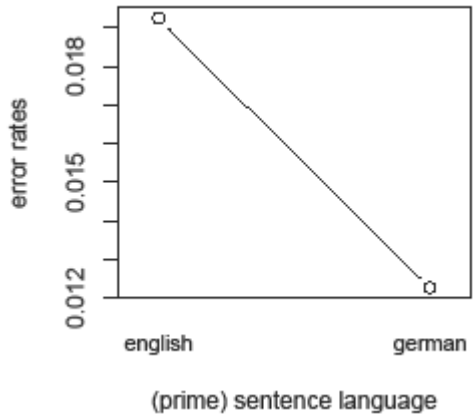


Figure 50. Proportion of incorrect responses (lmer *model estimates*) to words, after an L1 (German) or an L2 (English) context.

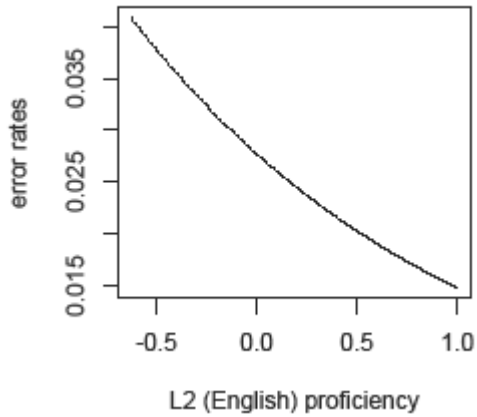


Figure 51. Proportion of incorrect responses (lmer *model estimates*) to words, with increasing L2 proficiency.

An interesting further observation is the absence of a conventionally significant effect of Item Type, which suggests that both German-English homographs and real word

controls were equally well identified. Similarly, the presentation format of a target did not have an impact on participants' responses (as seen in the absence of a simple effect of Format or any interactions involving Format). This null-effect of Format and the absence of an interaction with Item Type were also observed when items that could not be identified as German nouns[49] were excluded from the analysis. A possible explanation of these findings is that bilingual speakers do not rely greatly on this type of bottom-up visual information (i.e. onset capitalisation) when identifying items as belonging to their L1. Further analyses below, however, provide evidence that these German speakers are certainly not oblivious to this prominent marker in their L1.

3.3.3.1.1.2 Other variables

An inspection of secondary predictors revealed that native speakers identified longer words more accurately. This finding is reflected in the simple effect of word length (β = 0.531, SE = 0.167, z = 3.176, p < 0.001; see Figure 52 below). A similar effect was obtained in the first part of this study (i.e. in Experiment One). As discussed before, the present effect may be mainly based on the number of lexical competitors in the items' neighbourhood (cf. van Heuven et al. 1998), with larger neighbourhood density (and lexical competition) for shorter words, decreasing with increasing word length. In addition, given that length has been residualised against frequency, any length effects are more likely to be simple length effects, since I have removed the confound with frequency. Hence, other influencing factors (such as neighbourhood density) had a chance to appear.

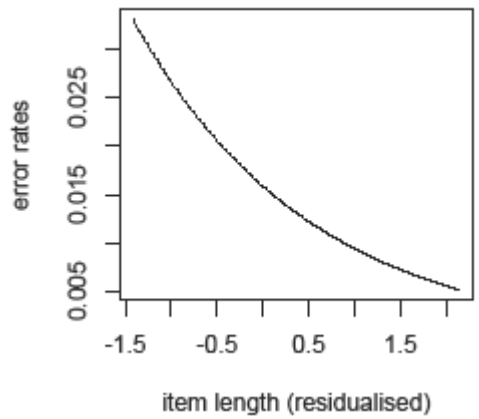


Figure 52. Proportion of incorrect responses
(lmer *model estimates*) to words, with increasing word length.

[49] The reader is reminded that onset capitalisation in German is limited to *nouns* only, but that the stimulus set in the present study included a small number of items that were not nouns (6 IHGs and 16 controls).

We also observe a difference between the two experimental sessions (β = 0.829, SE = 0.136, *z* = 6.112, *p* < 0.001), which most likely reflects familiarisation with the task (see Figure 53). Finally, the effect of German word class showed a trend towards significance (β = -0.654, SE = 0.361, *z* = -1.812, *p* = 0.07; see Figure 54), with fewer errors to nouns (cf. Baten et al. 2011; Elston-Güttler & Friederici 2005; van Hell & de Groot 2008; Sunderman & Kroll 2006).

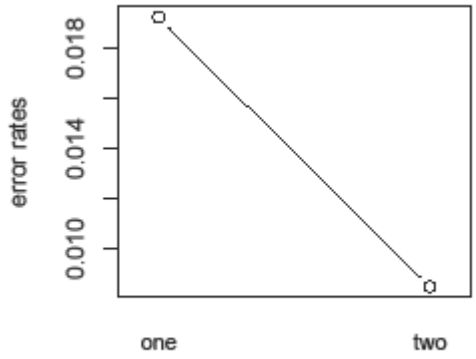


Figure 53. Proportion of incorrect responses (lmer *model estimates*) to words in the two experimental sessions

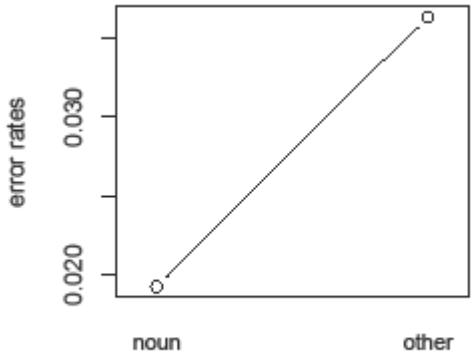


Figure 54. Proportion of incorrect responses (lmer *model estimates*) to words, depending on their L1 syntactic category.

3.3.3.1.2 IHG data

Although the analysis above did not reveal differences in the accuracy of identifying IHGs as opposed to controls (e.g. Dijkstra et al. 1998; Schwartz & Kroll 2006; both experiment 1), certain processing differences were detected in the analysis of response latencies (see section 3.3.3.3.1.1 below). To explore the impact of the main variables of interest (i.e. Format, Sentence Language and Proficiency) on error rates for either item type, separate analyses were conducted on IHGs and controls. The following analysis focuses on responses to IHGs. Again, a mixed logit model was fitted to the data, including the predictors discussed in section 3.2.2.3 (with the exception of Item Type).

3.3.3.1.2.1 Main predictors

This IHG analysis shows very similar statistical effects to the combined analysis of IHGs and controls (see Table G2, Appendix C). In line with the previous analysis, the language of the sentence frame affects participants' responses, with more accurate identifications following a German sentence (i.e. the language of the lexical decision task, and also the participants' native language). This is reflected in the simple effect of Sentence Language (β = 0.366, SE = 0.142, *z* = 2.583, *p* = 0.010; see Figure 55). The only other conventionally significant effect is Proficiency (β = 0.523, SE = 0.219, *z* = 2.381, *p* = 0.017), with more

accurate responses among more advanced bilinguals (see Figure 56), suggesting that these speakers may have greater control of their two languages. Note that a subsequent analysis including only German *nouns* returned the same significant simple effects. Similar to the combined analysis (see section 3.3.3.1.1 above), Format does not show as a simple effect, nor is it involved in any interactions.

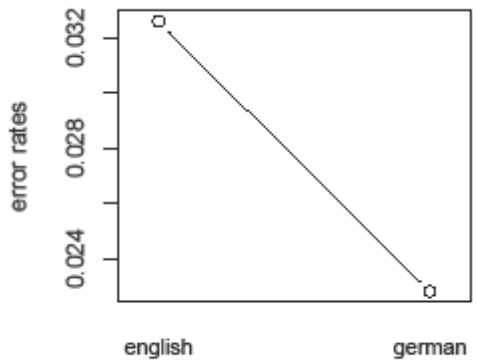


Figure 55. Proportion of incorrect responses (lmer *model estimates*) to German-English homographs after an L1 (German) or an L2 (English) context.

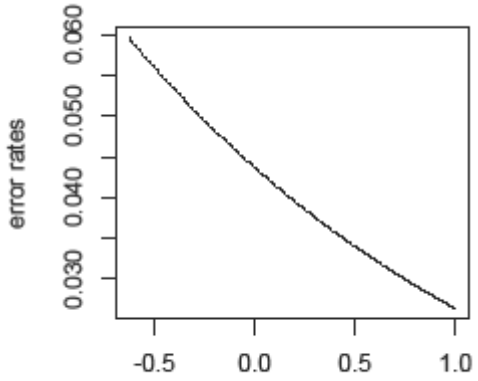


Figure 56. Proportion of incorrect responses (lmer *model estimates*) to German-English homographs, with increasing L2 proficiency.

3.3.3.1.2.2 Other variables

In this subset of data, Length (see previous discussion) and Session number are once again revealed as significant simple effects (Length: β = 0.852, SE = 0.251, z = 3.393, p < 0.001; Session: β = 0.993, SE = 0.146, z = 6.788, p < 0.001). In contrast to the main analysis, the *English* word class of an IHG has an impact on participants' responses, with more correct identifications of German words that are homographs to nouns in English (β = -1.161, SE = 0.546, z = -2.126, p = 0.033). This is a very interesting finding, because the reverse effect was obtained for the English-German bilingual population tested in Experiment One (i.e. more errors were found with German words that are homographs to English nouns). Clearly, this finding provides further support for the parallel activation of a bilingual's two languages (i.e. the language non-selective view). But how can the conflicting direction of the effects be explained? As argued before, one possibility is that a mismatch in the syntactic category of an IHG across the two languages creates a processing disadvantage (for recent evidence, see Baten et al. 2011; Sunderman & Kroll 2006). To test whether this effect would disappear (as it did for the mentioned complementary bilingual group), I reduced the amount of category incongruity (from .28 to .18) by excluding 6 IHGs that were not nouns in German. (This kind of exclusion was the best option, since words categorised as 'other' did not always have a matching

syntactic category across languages. In fact, labelling the IHGs as either 'congruent' or 'incongruent' would not have improved the category congruency ratio.) More specifically, I created a slightly reduced data set, which comprised 33 IHGs that were all nouns in German, but 6 of which had a different syntactic category in English. A separate analysis of this reduced data set revealed the same effect of noun-hood in English. Consequently, the present findings provide clear evidence for the importance of syntactic category congruence and are consistent with the findings of previous research. In addition, it could also be argued that nouns reflect a certain processing advantage compared to other syntactic categories (cf. Devadiga & Bhat 2011). This aspect, however, remains to be further explored in future studies.[50]

3.3.3.1.3 German control word data

As a complement to the separate error rates analysis of IHGs, the following analysis focused on responses to (German) controls only. A mixed logit model was fitted to the data, including the predictors mentioned in section 3.3.3.1.2.

3.3.3.1.3.1 Main predictors

In this subset of data only Proficiency (β = 0.795, SE = 0.260, z = 3.060, p = 0.002) and Sentence Language (β = 0.537, SE = 0.169, z = 3.177, p = 0.001) show as conventionally significant effects. German sentence is again shown as a facilitative factor (Figure 57), whereas an increasing L2 proficiency level results in a smaller proportion of incorrect responses (see Table G3, Appendix C; Figure 58).

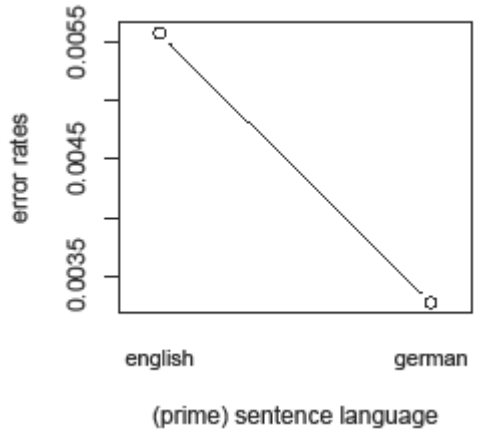


Figure 57. Proportion of incorrect responses (lmer *model estimates*) to L1 words after an L1 (German) or an L2 (English) context.

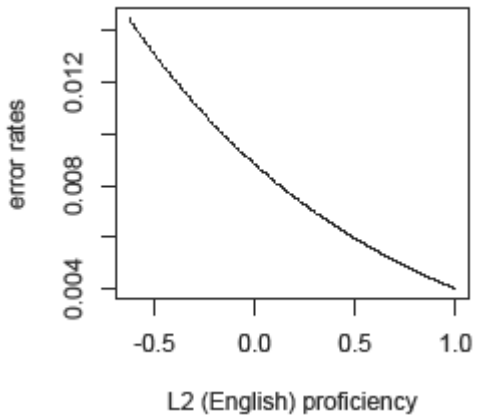


Figure 58. Proportion of incorrect responses (lmer *model estimates*) to L1 words, with increasing L2 proficiency.

[50] A separate analysis also tested the possibility that this effect was due to the nature of the experiment, by inspecting whether (English) word class would enter an interaction with Proficiency or Format, respectively. Neither interaction was returned significant, which provides strong support for the previous discussion, i.e. the necessity of category congruency.

With no other simple effects or interactions, these findings suggest that very few resources need to be drawn upon during the processing and decision-making in the L1, given a fairly limited interference from the L2 lexicon. Alternatively, the scarceness of effects may also be taken as evidence that in an L1 task, L1 effects prevail over any possible L2 effects.

3.3.3.1.3.2 Other variables

Looking at other factors included in the model, we find that significantly more errors were made to German control words that are not nouns (β = -1.155, SE = 0.453, z = -2.551, p = 0.011), and also to controls with larger *English* orthographic neighbourhoods (β = -0.108, SE = 0.045, z = -2.399, p = 0.016). While the second of these results supports the view that bilingual language processing is largely non-selective in nature, the first implies that participants are not oblivious to the grammatical class of a presented stimulus (Baten et al. 2011). A similarly interesting finding is that when only nouns are included in the analysis, participants' responses are not affected by the number of orthographic entries that resemble an *English* word in the stimulus' neighbourhood but instead by the number of *German* neighbours, by exhibiting more errors with an increasing neighbourhood density (cf. van Heuven et al. 1998).

3.3.3.1.4 Nonword data

The analysis of error rates to nonwords was intended to provide further insights into the less automatised recognition processes, investigating what type of information (i.e. bottom-up, top-down, both or none) is utilised by a bilingual speaker conducting a task in their more dominant language. Similar to the previous analyses, a mixed logit model was fitted to the data which consisted of responses to nonword tokens only.

3.3.3.1.4.1 Main predictors

It is only when we look at their responses to nonwords that we find a significant effect of Format for our L1 (German) participants (β = -0.898, SE = 0.424, z = -2.118, p = 0.034; Figure 59), with fewer errors to items presented all in lower case (see Table G4, Appendix C). Interestingly, this effect is further defined by an interaction with Proficiency (β = 1.501, SE = 0.650, z = 2.309, p = 0.021), with lower proficiency L2

speakers being more prone to accept something unknown as an existing German (L1) word, particularly when it is presented with an onset capital letter (see Figure 60).[51]

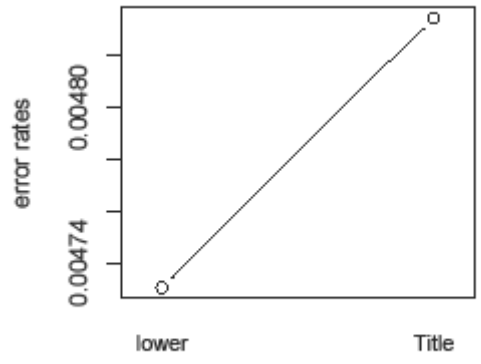


Figure 59. Proportion of incorrect responses for nonwords (lmer *model estimates*), either presented all in lower case (lower), or with the first letter capitalised (Title).

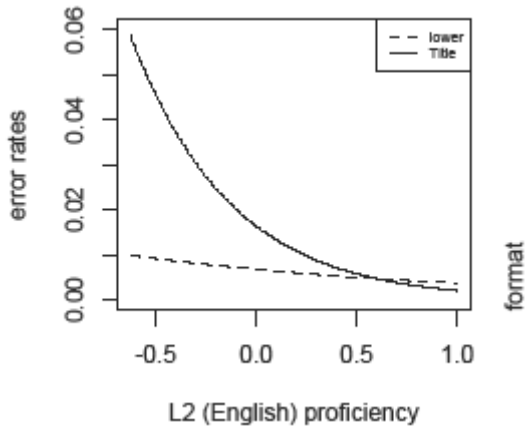


Figure 60. Proportion of incorrect responses for nonwords (lmer *model estimates*), in relation to their presentation format (lower vs. Title) and L2 proficiency level.

One explanation of this finding may be that increased exposure to an L2 (which does not require onset capitalisation to the same extent as the L1) leads to decreased sensitivity towards capitalisation in general. Could it be then that the L1-specific lexicality marker is not as functional as anticipated? Or could it be perhaps that L1 processing is affected by L2 exposure?

A better explanation of the findings is provided when we consider the results of the previous experiment (section 3.2.3, with bilinguals conducting the task in their L2): the pattern of responses with increasing (L2) proficiency appears remarkably similar across the two experiments. More specifically, while very proficient bilinguals do not seem clearly affected by the task requirements and the available linguistic information, less proficient L2 speakers seem more misled by the experimental setup and seem to make different use of resources to execute the task. If this conjecture holds, then we should also be able to observe a proficiency-related difference in the impact of the language of

[51] In accordance with the first experiment, a separate analysis was conducted on a small subset of data, i.e. in the first experimental session, with target presentation following a German sentence. This additional analysis was done to investigate 'single-language task effects' for bilinguals conducting the task in their L1 (cf. Lemhöfer & Radach 2009). The analysis revealed a strikingly similar pattern of results to those just reported. That is, Format was revealed as a simple effect and there was a strong tendency towards significance for the interaction between Proficiency and Format (p=0.083). Accordingly, while items presented with an onset capital letter generated more errors, less proficient bilinguals seemed particularly affected by this bottom-up information (cf. Figure 60). Although this result has to be taken with caution (due to a substantially reduced data set), it provides further support for the notion of a "language-filtering" mechanism (Lemhöfer & Radach 2009:45). This mechanism allows items that are less clearly identifiable as belonging to a particular language to be quickly suppressed in order to avoid errors. In addition, the results can be argued to also extend previous research by showing that this effect is more accentuated for less proficient bilinguals (for implications, see the discussion section below).

the sentence preceding the target, with less proficient bilinguals producing more errors after an English (L2) sentence. In addition, this effect should be primarily evident in the first session, i.e. when participants had not become familiar with the task. A separate analysis, limited to nonword responses in the first session only, revealed the predicted outcome, with a highly significant interaction between Proficiency and Sentence Language (see Figure 61 below).

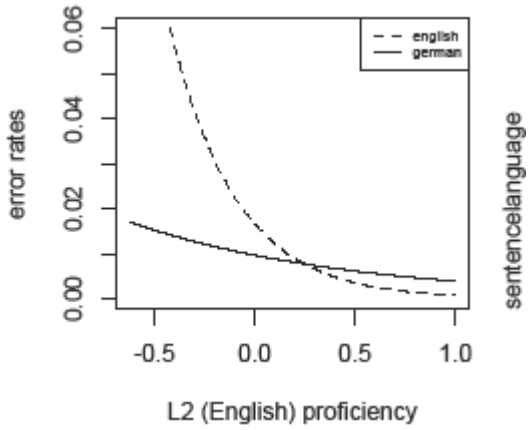


Figure 61. Proportion of incorrect responses for nonwords (lmer *model estimates*) in the *first experimental session*, depending on the preceding sentence language (English vs. German) and L2 proficiency level.

These results suggest considerable parallels between the visual processing in a bilingual's dominant and non-dominant language. That is, when input cannot be processed automatically (because it is not known), bottom-up information (i.e. language-specific cues) seems to be a significant source of language information for less proficient bilinguals, perhaps even over and above top-down language information (i.e. sentence language).

3.3.3.1.4.2 Other variables

Among other variables which have been revealed as significant we find the size of the L1 (German) and L2 (English) neighbourhoods (see Table G4, Appendix C). While the accuracy of German native speakers decreases with an increasing number of existing German words in the neighbourhood of the stimulus (β = -0.352, SE = 0.060, z = -5.834, p < 0.001), the opposite effect is observed as the number of competing English orthographic representations increases (β = 0.084, SE = 0.031, z = 2.751, p = 0.006, see Figures 62 and 63, respectively).[52] The former of the two effects is in line with the well-

[52] Note that these effects are equivalent to the results obtained in the previous study (for discussion and relevant comments, see Experiment One, sections 3.2.3.1 and 3.2.3.2).

established neighbourhood effect, mainly reported in monolingual studies (e.g. Andrews 1989; M. Coltheart, Davelaar, Jonasson, & Besner 1977; Davis, Perea, & Acha 2009; Grainger, Muneaux, Farioli, & Ziegler 2005; Siakaluk, Sears, & Lupker 2002; but see van Heuven et al. 1998, for bilingual neighbourhood effects), and reflects the nature of the task (i.e. being conducted in the L1). In addition, the considerable impact of the non-target language confirms that even in an L1 context, there is a certain amount of activity in the L2 (i.e. the language non-selective view). Finally, the finding that a larger L1 neighbourhood has a smaller impact on participants' responses can be explained on the assumption that the larger the resemblance of a nonword to an English item, the smaller the likelihood that it is an existing German word; hence, the less difficult it gets to reject the nonword.

Similar to the study conducted in the bilinguals' L2 (see section 3.2.3), the current findings confirm the existence of an early discrimination mechanism which affects the processing of stimuli that are more similar to the current language of processing (Lemhöfer & Dijkstra 2004; Lemhöfer & Radach 2009).

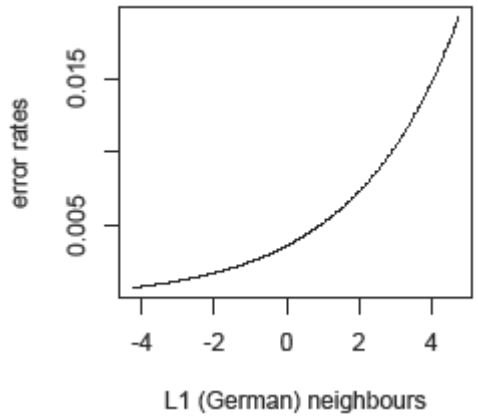


Figure 62. Proportion of incorrect responses for nonwords (lmer *model estimates*), with an increasing L1 neighbourhood.

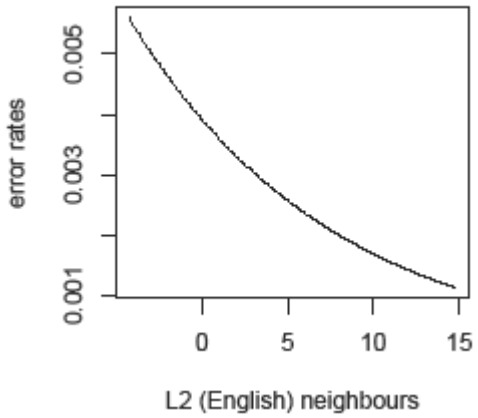


Figure 63. Proportion of incorrect responses for nonwords (lmer *model estimates*), with an increasing L2 neighbourhood.

Finally, nonword rejections seem also affected by the length of a stimulus, being less difficult with an increasing number of letters (β = -0.647, SE = 0.195, z = -3.321, p < 0.001; Figure 64). While a similar effect was obtained for responses to existing words – although in the opposite direction – an analogous argument can be put forward here. That is, the present effect may be mainly based on the number of lexical competitors in the items' neighbourhood (cf. van Heuven et al. 1998), with larger neighbourhood density (and lexical competition) for shorter words, decreasing with increasing word length. This conjecture seems confirmed in the high correlation between item length and

neighbourhood density (German neighbours: r = -.31, p < 0.001; English neighbours: r = -.61; p < 0.001).

Overall, these results suggest that the higher the confusability of the nonword with existing German words, the more difficult it gets to reject it. These findings are in line with the previous analyses and discussion.

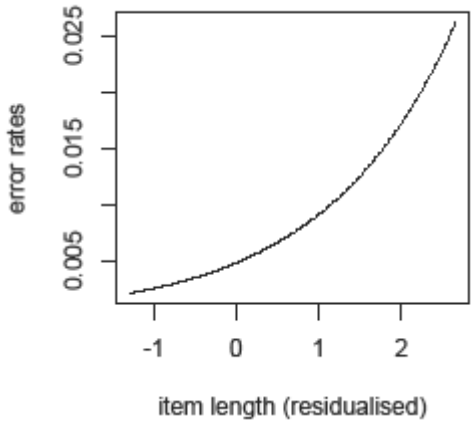


Figure 64. Proportion of incorrect responses for nonwords (lmer *model estimates*), with increasing word length.

3.3.3.2 Discussion of the Error Analysis

Analyses of error rates to words yielded no difference in the accuracy of responses between interlingual homographs and their matched real word control items. This finding is not surprising for two reasons: first, the fact that the current task was an L1 lexical decision task means that the less relevant (i.e. L2) representation of an IHG would have received less activation (e.g. Grosjean 2001) and therefore exhibited less competition in the selection process. It follows that the recognition process of these critical stimuli should have been similar to that of stimuli that were only words in German (i.e. the matched controls), resulting in the absence of an effect. Second, researchers employing interlingual homographs in their studies have reported very mixed results, ranging from facilitation through null-effects to inhibition (see van Heuven et al. 2008, for a comprehensive listing of some of the results). Various factors have been suggested as having an impact on such inconsistencies, including task demands, list composition (van Heuven et al. 2008), as well as the availability of a semantic context (e.g. Schwartz & Kroll 2006) and categorial semantics (Baten et al. 2011). Differing from previous studies in its experimental design (i.e. being conducted in the bilinguals L1) and main variables of interest, the findings of the present study can add new insights to the current understanding of bilingual IHG processing.

Another effect that is consistently found in the responses to words and is in line with the findings of recent research conducted in the bilinguals' L2 (e.g. Chambers & Cooke 2009; Duyck et al. 2007; Schwartz & Kroll 2006), is that the availability of a language context can considerably reduce, if not eliminate, lexical activation of the non-target language. That is, the availability of a task-matching (L1) sentence language in the current study reinforced (i.e. facilitated) an accurate identification of a stimulus as an existing German (L1) word. Because this effect occurred even though the sentence was unrelated in meaning to the target, it could be argued that "a relatively low-constraint sentence is sufficient to more strongly activate the target [lexical representation] allowing it to compete early for selection" (Schwartz & Kroll 2006:209, having obtained a very similar result). However, perhaps a better interpretation of this finding is to assume that an L1 sentence helped (German-English) bilingual participants to zoom into the language of the task (Elston-Güttler, Gunter, et al. 2005). This interpretation is in line with the task schema account (e.g. Dijkstra & van Heuven 2002a; Green 1998a), which suggests that response errors after a task-irrelevant sentence language (i.e. a language switch) appear because of incongruent task schema settings (i.e. *L2* sentence/context language schema versus *L1* target/response schema) in the course of a trial.[53]

Interestingly, although most research to date has focused on highly proficient bilingual populations, some researchers suggest that the context facilitation effect applies mainly to this highly proficient group of speakers (e.g. Elston-Güttler, Paulmann, et al. 2005). While this may be true for processing in the L2 (see also sections 3.2.3.2 and 3.2.3.4, for further insights provided by this study), no such effects were found in the present data, i.e. in a task which was conducted in the participants' L1. It does not come as a surprise that L2 proficiency does not noticeably influence the effect of the language of the sentence preceding the target (L1 or L2), especially considering the current task demands; yet, it provides an interesting addition to current bilingual research. Importantly however, the absence of an interaction between Sentence Language and Proficiency does not mean that all participants behaved similarly because they were conducting a task in their native language; quite the contrary. The data show that

[53] Note that in the original task schema account, Green (1998a) proposes that depending on task requirements the relevant task schemas can be configured in various ways (e.g. into a monolingual as opposed to a bilingual mode, to display or allow for code-switching). Although such a configuration would be appealing for the present study, the presence of further influential factors in the experiment (i.e. stimulus presentation format and participants' L2 proficiency) makes a clear-cut configuration more difficult.

irrespective of sentence language less proficient bilinguals produce significantly more errors than their more proficient counterparts (although minimally so). A possible explanation of this observation is that more proficient bilingual speakers are less affected by the experimental set-up and in better control of their two languages (even though L2 interference cannot be fully ruled out at this stage, and will be further examined in the RT analyses below). In addition, higher L2 proficiency may also be accompanied by other cognitive advantages which can assist in successful task execution.

With respect to the variable of principal interest, the Format of the stimulus, the error analyses do not seem to support the predictions made above. That is, there is a consistent lack of an effect of Format either as a simple main effect or in interaction with other factors. Although this finding suggests that bilingual speakers who process information in their first language do not rely greatly on this (bottom-up) visual information, this initial impression needs revision when we consider responses to nonword tokens.

In contrast to real word responses, analyses of the accuracy of *nonword* rejections provide strong evidence that the bilingual group of speakers employed in this study is *not* oblivious to the presentation format of stimuli (i.e. whether all in lower case, or with the first letter capitalised). This effect is further defined by the interaction with Proficiency, which exposes lower proficiency L2 speakers as being particularly sensitive towards these bottom-up cues, and erroneously identifying items as belonging to their L1 when presented with an onset capital letter.

Recalling the discussion of nonword rejections in the previous experiment, it has been suggested that the sum of (lexical) activity in a bilingual's most activated lexicon (as controlled, for instance, by the language nodes) is a major factor in determining the accuracy of a nonword response (e.g. van Heuven et al. 1998:475). What follows then is that the language-specific sub-lexical information used in this study (i.e. onset capital letters) is an early determiner of which language is initially activated, probably by being linked with the German (L1) language node(s). Accordingly, high levels of summed activity in the L1 cause higher error rates to a German-like nonword (i.e. one carrying an onset capital letter).

The question that remains to be answered is, why are these effects particularly evident for bilinguals who are *less* proficient in their L2? As mentioned above, a tentative interpretation of this finding is that less proficient bilinguals make different use of the

(cognitive and experimental) resources available to them.[54] This possibility can be tested taking a more theoretical approach to nonword rejection, as put forward by the MROM (Grainger & Jacobs 1996). The MROM suggests that the type and speed of a response in an LDT is determined by three response criteria (the local activity criterion *M*, the global activity criterion *Σ*, and the temporal deadline *T*), two of which (*Σ* and *T*) can be modified by certain experimental factors. While the temporal deadline (*T*) is the main determinant of a successful "no" (i.e. nonword) response, errors to nonword stimuli ("false positives") arise with either a high *T* criterion or a low *Σ* criterion[55] (Grainger & Jacobs 1996:522). In the MROM, it is further argued that

> stimuli that generate a high *σ* value [global activity in the orthographic lexicon] in early phases of processing [...] give rise to a lower average value of the Σ criterion. (Grainger & Jacobs 1996:523)

Returning to the context of the current study, if we assume that *T* was constant for all participants (given that the task was conducted in the L1 and more automatic responses were obtained than would be the case in an L2 task), the critical factor to influence nonword responses and evoke an increased proportion of false positives in less proficient bilingual speakers remains the *Σ* criterion.[56] If we further assume that such lower proficiency bilinguals are more sensitive towards (L1) sub-lexical cues or make more use of this kind of information for the activation of their different languages (cf. Dijkstra 2005; Elston-Güttler, Paulmann, et al. 2005; Schwartz & Kroll 2006), then these speakers' global activity levels (in the L1 lexicon) may be slightly more elevated than for more proficient bilinguals (i.e. have a higher *σ* value). As a result, less proficient bilinguals' responses may be based on meeting a somewhat lowered *Σ* criterion (i.e. being predominantly fast guesses), whereas more proficient bilinguals' responses are truly based on meeting the temporal deadline. This conjecture will be further examined in the subsequent analysis of participants' response latencies.

[54] Note that the results obtained in the first part of this study (i.e. Experiment One, section 3.2.3.2), point towards the same overall trend.
[55] A low *M* criterion is also suggested as a possible error source, but it is not important in the present argument because the local activity of nonword stimuli is assumed not to vary across the present participant population.
[56] In Grainger and Jacobs' MROM (1996), the *Σ* criterion represents fast guesses.

3.3.3.3 RT Analysis

3.3.3.3.1 IHG and German control word data

For all response time analyses the statistical method chosen was linear mixed-effects modelling (for a discussion, see section 3.2.2.2 above). The first of four analyses reported below included both types of existing words, namely German-English homographs and their matched control words.

3.3.3.3.1.1 Main predictors

The latency data indicate primarily that the different types of real word stimuli included in the experiment (i.e. IHGs and German controls) are processed differently, depending on certain factors (see Table G5, Appendix C). One such factor is Proficiency, which enters an interaction with Item Type (MCMCmean = -0.046, HPD95lower = -0.070, HPD95upper = -0.021, $p < 0.001$). As can be seen in Figure 65 below, with increased exposure to an L2, responses to uniquely German words are made faster than to words that exist in both of a bilingual's languages (i.e. interlingual homographs). Above all, this finding could be taken as evidence of a developmental pattern of bilingual lexical organisation, with the two languages becoming increasingly separated (see discussion below). We can test this further by conducting separate analyses of both the IHG material and the control stimulus set (see below).

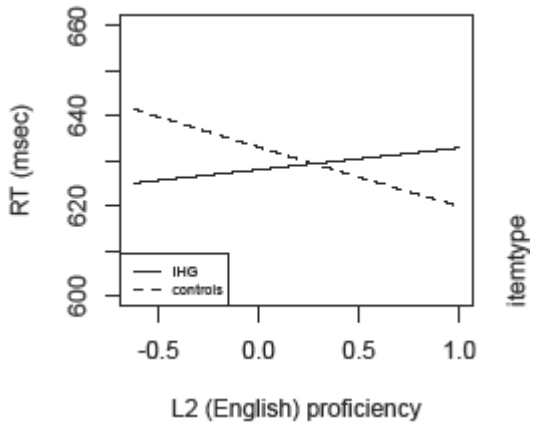


Figure 65. Mean response times for German-English homographs (IHG) and German control words, with increasing L2 proficiency.

The other factor that has an impact on the processing of words is Sentence Language, as reflected in the interaction of Item Type and Sentence Language (MCMCmean = 0.048, HPD95lower = 0.027, HPD95upper = 0.070, $p < 0.001$; see Figure 66). Here the data reveal that the task-relevant (L1) sentence language is particularly helpful in activating the appropriate meaning of an interlingual homograph, hence leading to faster response

latencies (and vice versa). Similarly, decisions about uniquely German words are inhibited when the stimuli are presented after a sentence frame in a language that does not match the language of the task and the participants' L1. These findings not only confirm previously reported "zooming-in" effects (e.g. Elston-Güttler, Gunter, et al. 2005), but also replicate the widely reported language switching costs (e.g. Kolers 1966b; von Studnitz & Green 1997).

The effect of Sentence Language is also found in the interaction with Proficiency (MCMCmean = 0.031, HPD95lower = 0.008, HPD95upper = 0.056, p = 0.012; see Figure 67). While there is a clear difference in the speed of responses after an L2 as opposed to an L1 sentence frame (with faster responses following the L1), lower proficiency L2 speakers display more of this L1 advantage. A possible interpretation of this finding is that bilingual speakers with a reduced exposure to an L2 are better able to zoom into their L1. This may be possible because the activation thresholds for the L2 are set somewhat lower than for more proficient speakers; hence, they experience less interference from the competing L2 lexicon.

Similar to the analyses of error rates, the speed with which real word responses can be given to words of the L1 does not depend greatly on their presentation format, as reflected in the absence of a simple effect of Format or any interactions involving Format. This is consistent with the idea that words that are well known by the bilingual are recognised in an automatised (bottom-up) way, with no explicit need for (e.g. top-down) language clarification (by means of, for instance, the language nodes; Dijkstra 2005:187).

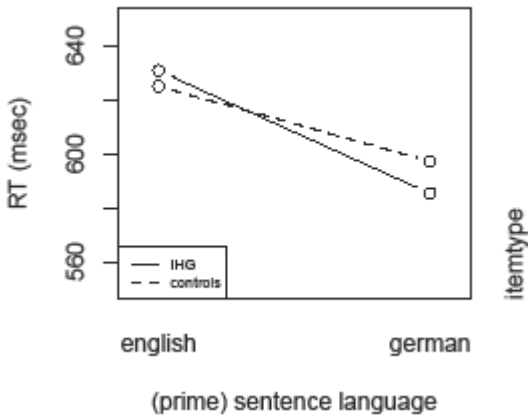


Figure 66. Mean response times for IHGs and controls, depending on the preceding sentence language (German or English).

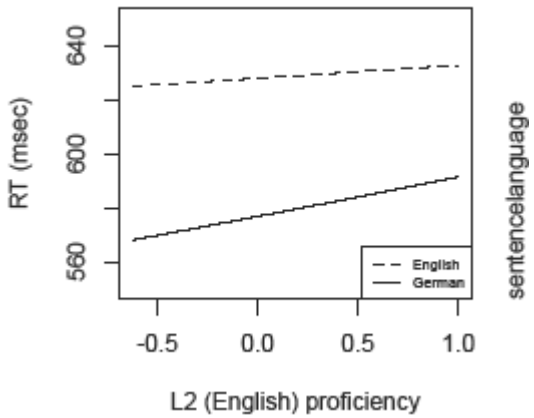


Figure 67. Mean response times for words, by increasing L2 proficiency and preceding sentence language.

3.3.3.3.1.2 Other variables

In line with previous research (see Lemhöfer et al. 2008, for a brief overview of the most recent bilingual evidence), more frequent words are identified more quickly (MCMCmean = -0.051, HPD95lower = -0.076, HPD95upper = -0.025, $p < 0.001$). Moreover, responses in the second experimental session are made more quickly (MCMCmean = -0.098, HPD95lower = -0.113, HPD95upper = -0.083, $p < 0.001$), which again indicates possible familiarisation with the task. Finally, as indicated in previous analyses German *nouns* are recognized more quickly (MCMCmean = 0.087, HPD95lower = 0.039, HPD95upper = 0.134, $p < 0.001$). On the one hand, this finding demonstrates that participants are not oblivious to the grammatical class of a presented stimulus, with nouns displaying a certain processing advantage over other syntactic categories (cf. Devadiga & Bhat 2011). On the other hand, this finding may also reflect to some extent the importance of syntactic category congruence for cross-linguistic material (i.e. only for IHGs). As discussed in the error analysis above, recent research argues that "categorial meaning [of IHGs] modulates the extent of cross-linguistic activations" (Baten et al. 2011:357), i.e. facilitating responses. It may be then that the reported facilitation effect is not based solely on the fact that a stimulus is a noun, but also that it has a matching *categorial* meaning across the two languages. Given that the effect did not interact with Format, it does not seem dependent on the visual appearance of the stimulus. This assumption was further confirmed in a subsequent analysis (which excluded items other than nouns) with the absence of a simple effect of Format, or any interactions therewith.

3.3.3.3.2 IHG data

As anticipated in the analysis above, to explore the impact of the main variables of interest (i.e. Format, Sentence Language and Proficiency) on either of the two word types, separate analyses were conducted on IHGs and controls. The following analysis focuses on responses to IHGs only.

3.3.3.3.2.1 Main predictors

The latency analysis of correct responses to just the German-English homographs returned further refinement of the simple effects and effect sizes reported above (see Table G6, Appendix C). In this subset of data, Sentence Language did not interact with Proficiency but was instead revealed as a simple effect (MCMCmean = -0.127, HPD95lower = -0.144, HPD95upper = -0.110, $p < 0.001$; see Figure 68), with overall

faster responses following an L1 (German) sentence frame. On the other hand, Proficiency entered an interaction with Format (MCMCmean = 0.250, HPD95lower = 0.056, HPD95upper = 0.450, p = 0.014; see Figure 69). This interaction is mainly carried by the different impact of Format on less proficient L2 speakers, with IHGs presented with an onset capital being recognised more easily. This finding is consistent with my previous assumption (as well as the error data), namely that lower proficiency L2 speakers rely more on bottom-up information, even when processing in their L1.

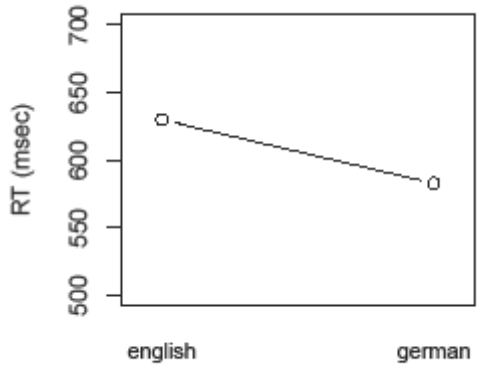


Figure 68. Mean response times for German-English homographs, defined by the preceding sentence language.

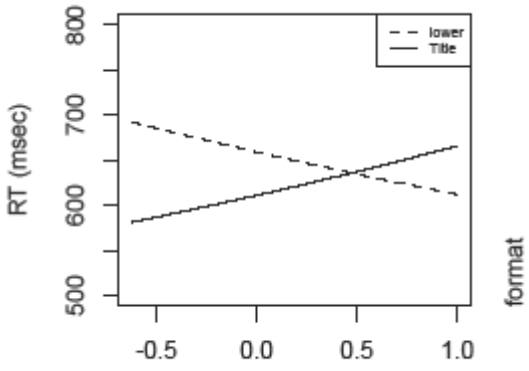


Figure 69. Mean response times for German-English homographs, by increasing L2 proficiency and presentation format.

3.3.3.3.2.2 Other variables

The only two other variables that are found to have a significant effect in the statistical model are Frequency (MCMCmean = -0.072, HPD95lower = -0.116, HPD95upper = -0.028, p = 0.002; Figure 70) and Session (MCMCmean = -0.111, HPD95lower = -0.128, HPD95upper = -0.094, p < 0.001; Figure 71). Notably, response times during the second session were faster compared to those in the first experimental session.

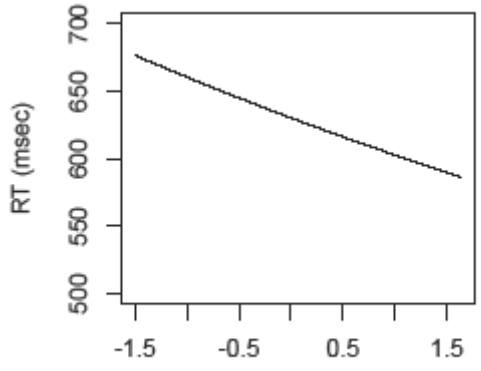


Figure 70. Mean response times for German-English homographs, by increasing log frequency.

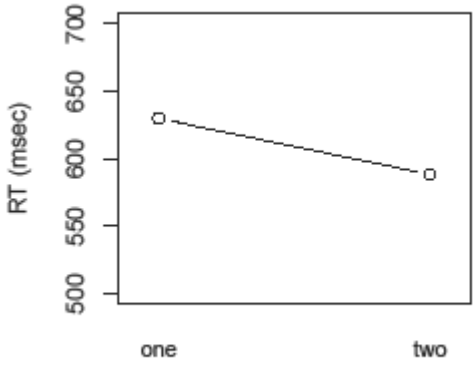


Figure 71. Mean response times for German-English homographs, depending on experimental session.

3.3.3.3.3 German control word data

As a complement to the separate response time analysis for IHGs, the following analysis focused on responses to (German) controls only.

3.3.3.3.3.1 Main predictors

When *correctly* responding to real German words (controls), native speakers seem not to be affected by the presentation format of the stimulus (see Table G7, Appendix C). One of the factors that has an impact on the speed of these subjects' responses is the language of the sentence frame that precedes the target, as defined by the simple effect of Sentence Language (MCMCmean = -0.095, HPD95lower = -0.116, HPD95upper = -0.076, $p < 0.001$; Figure 72). The other factor is Proficiency, involved in an interaction with Sentence Language (MCMCmean = 0.035, HPD95lower = 0.004, HPD95upper = 0.066, $p = 0.028$). While there are no notable differences in processing L1 words after an L1 sentence frame across the participants, higher proficiency L2 speakers seem better able to perform on mixed-language trials, i.e. when a task-irrelevant language (L2) is heard prior to stimulus presentation (Figure 73).

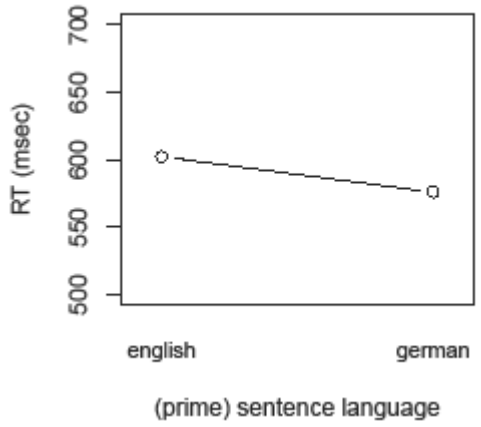


Figure 72. Mean response times for L1 (German) words, defined by the preceding sentence language.

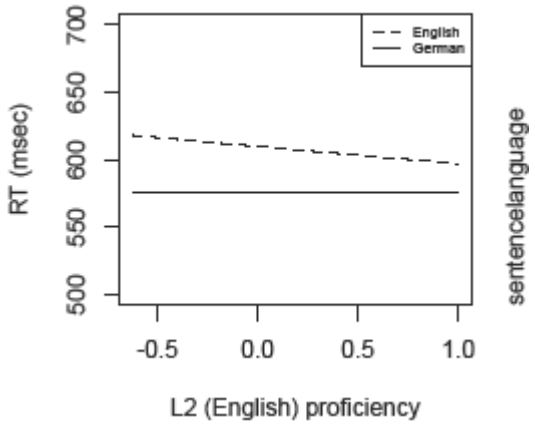


Figure 73. Mean response times for L1 (German) words, as defined by the preceding sentence language and L2 (English) proficiency of speakers.

3.3.3.3.3.2 Other variables

Apart from the commonly observed simple effect of Frequency (MCMCmean = -0.032, HPD95lower = -0.061, HPD95upper = -0.002, $p = 0.030$; Figure 74), there was also a significant effect of L1 (German) Word Class (MCMCmean = 0.127, HPD95lower = 0.071, HPD95upper = 0.180, $p < 0.001$; Figure 75), reflecting faster responses to nouns.[57]

[57] Note that the simple effect of L1 Word Class did not interact with Format, i.e. faster responses are not facilitated by the typical German presentation format for nouns. Also note that Format did not reach significance (or enter an interaction) even when only nouns were included. This confirms again that the

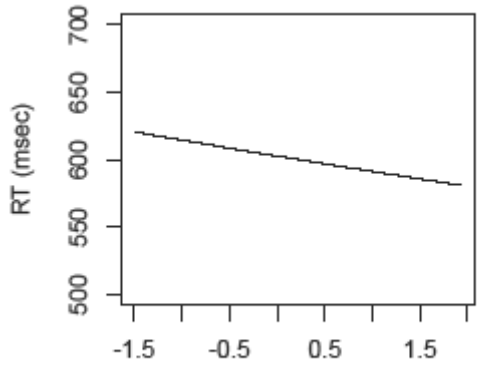


Figure 74. Mean response times for L1 (German) words, with increasing log frequency.

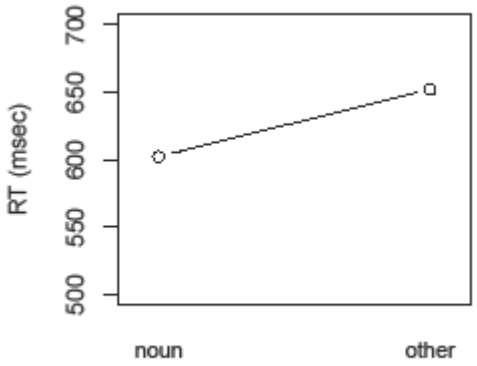


Figure 75. Mean response times for L1 (German) words, depending on the stimulus' grammatical category.

3.3.3.3.4 Nonword data

As in the analysis of error rates, the fourth and final analysis concerned the speed of nonword rejections. This was done to provide further insights into the less automatised recognition processes, investigating what type of information (i.e. bottom-up, top-down, both or none) is utilised by a bilingual speaker conducting a task in their more dominant language. Similar to the previous analyses, a linear mixed-effects model was fitted to the data which consisted of nonword responses only. The same (main and secondary) predictors were included in the fitted model, with the exception of Item Type. Where necessary, predictors were centred or residualised to avoid collinearity.

3.3.3.3.4.1 Main predictors

The pattern of responses observed for correct rejections of nonwords suggests that the speed with which this can be done depends heavily on two factors: Format and the language of the sentence preceding the target (see Table G8, Appendix C). This is reflected in the analysis in the simple effects of Format (MCMCmean = 0.083, HPD95lower = 0.004, HPD95upper = 0.160, p = 0.037; Figure 76) and Sentence Language (MCMCmean = -0.064, HPD95lower = -0.077, HPD95upper = -0.050, p < 0.001; Figure 77), as well as in the interaction of these two factors MCMCmean = -0.036, HPD95lower = -0.054, HPD95upper = -0.017, p < 0.001; see Figure 78).

processing of purely L1 items did not rely on the involvement of bottom-up cues nearly to the same extent as is noticeable with other items (or other bilingual groups).

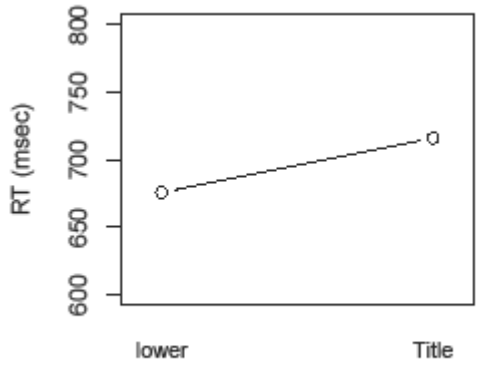


Figure 76. Mean rejection times for nonwords, in relation to the presentation format.

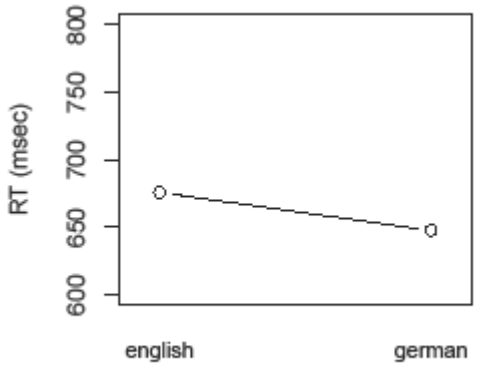


Figure 77. Mean rejection times for nonwords, as defined by the preceding sentence language.

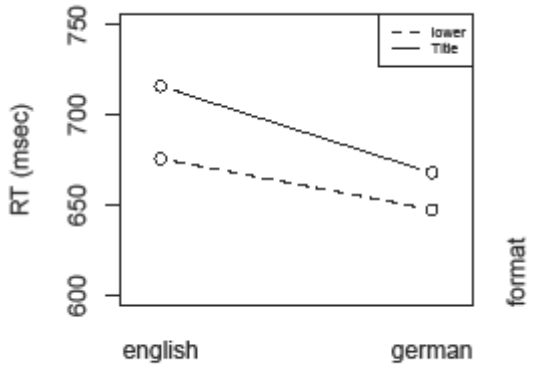


Figure 78. Mean rejection times for nonwords, as defined by the preceding sentence language and stimulus presentation format.

The interaction reveals that the overall slowest responses are being made in a situation where a combination of factors might be expected to require a more thorough 'search' of the lexicon(s), i.e. after an English context and to a German-biased Title case item. The fastest decisions on the other hand are being made after an L1 context to items all in lower case. There are two major aspects of bilingual visual processing that can be deduced from this finding. First, the findings give clear evidence that simultaneous processing of two languages imposes an additional cost on the speed with which a decision can be made (e.g. von Studnitz & Green 1997). More importantly, however, the findings support the idea that sub-lexical cues are a legitimate language discrimination mechanism (Lemhöfer & Dijkstra 2004:547), which has an impact on bilingual visual word recognition, even when processing is done in the L1.

Notably, there was no effect of Proficiency. This suggests that the fast guessing account put forward to explain the pattern of responses obtained in the error data may not be the sole predictor to account for the present results.

3.3.3.3.4.2 Other variables

With respect to the secondary variables, the simple effect of Length was significant (MCMCmean = 0.071, HPD95lower = 0.057, HPD95upper = 0.085, $p < 0.001$; Figure 79), as well Session (suggesting familiarisation with the task; MCMCmean = -0.093, HPD95lower = -0.116, HPD95upper = -0.068, $p < 0.001$; Figure 80), and the number of L1 (German) orthographic neighbours (with more neighbours leading to slower response times: MCMCmean = 0.027, HPD95lower = 0.021, HPD95upper = 0.033, $p < 0.001$; Figure 81). The neighbourhood and length effects have been previously reported in other psycholinguistic studies (e.g. van Heuven et al. 1998; Ziegler et al. 2001, respectively), and the familiarisation effect is unsurprising. The neighbourhood effect shows that the higher the confusability of a nonword with existing German words, the more difficult it is to reject.

Figure 79. Mean rejection times for nonwords, with increasing stimulus length.

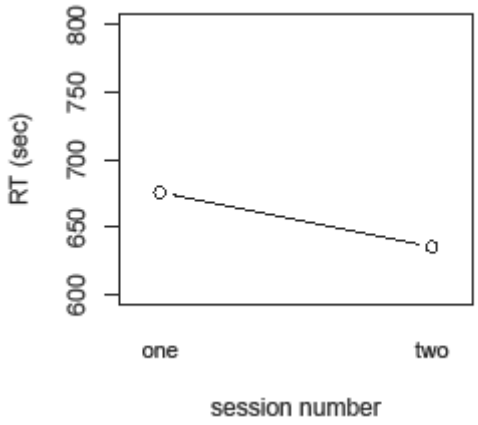


Figure 80. Mean rejection times for nonwords, depending on experimental session.

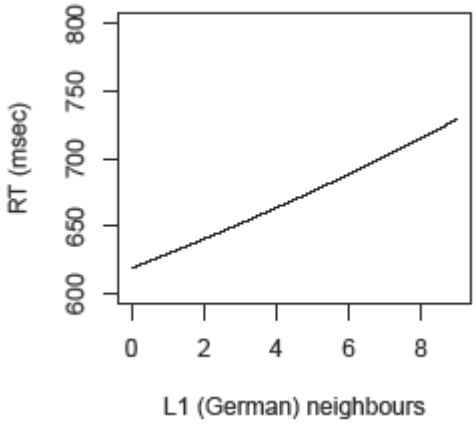


Figure 81. Mean rejection times for nonwords, with an increasing number of L1 neighbours.

3.3.3.4 Discussion of the RT analysis

The earlier analysis of error rates (see section 3.3.3.2) led to some suggestions concerning the processes used in L1 visual word recognition by the unbalanced bilingual participants in the current experiment. The analysis of response times to correct answers to words allows us to further refine the description of these processes.
One such refinement is evident in the responses to interlingual homographs. More generally, we find that responses following a German sentence frame are made more quickly, which reflects the recurrent finding that a mixed stimulus list or experimental setup is more likely to induce response delays, as opposed to a (largely) unilingual experimental context (for overviews, see Los 1999; Lupker, Kinoshita, Coltheart, & Taylor 2003). When we compare participants' responses to these stimuli with those obtained for control items, however, this top-down information seems particularly crucial to bilingual speakers when responding to interlingual homographs, exhibiting facilitation effects. As argued recently by Baten and colleagues, "in recognising interlingual homographs [...], the presence of a sentence context could guide lexical access towards the target language" (2011:352), by raising activation thresholds high enough to limit measurable influence of the non-target language. An issue which is not entirely clear from (the limited number of) previous studies as well as from the present findings is how this top-down information can be defined and by what means it feeds back to the orthographic level (see the general discussion section for more detail). Notably, research investigating the effect of a sentence context on bilingual processing is fairly limited to date and includes a wide range of cross-linguistic material and experimental setups (e.g. Chambers & Cooke 2009; Duyck et al. 2007; van Hell & de Groot 2008; Li & Yip 1998). Because of the mixed findings, much more research is needed to arrive at some definite conclusions. In addition, the three studies that are most relevant to the present work (Elston-Güttler, Gunter, et al. 2005; Libben & Titone 2009; Schwartz & Kroll 2006) focus on L1 interference during L2 processing[58] and also use different experimental techniques. Hence, the current results extend our knowledge by including the study of L2 interference on L1 bilingual processing (see also Titone et al. 2011).
The amount of cross-linguistic interference in an L1 task is also influenced by a participant's L2 proficiency level. The present finding that more proficient L2 speakers

[58] See sections 3.2.3.2 and 3.2.3.4 for a discussion and implications for the reverse part of this study.

take more time to respond to IHGs than to unambiguously German words (i.e. controls) can be explained on the basis of stronger cross-lexical competition for these bilingual speakers and in this particular context. This conjecture finds support in previous research, which has provided evidence for traces of L2 interference in purely L1 tasks (e.g. Duyck 2005; van Hell & Dijkstra 2002; van Wijnendaele & Brysbaert 2002),[59] and is in line with the non-selective view of lexical access. On the assumption that cross-linguistic activation also exists, though to a lesser extent, for bilingual speakers with *less* L2 experience, one might argue that the small processing advantage found for this group for IHGs (reflected in faster responses to IHGs than to controls) might have a different source, i.e. reflecting a developmental lexical organisation. This means that initially low proficiency L2 speakers possess either one lexical entry for the different readings of an interlingual homograph, or (more likely) two lexical entries with very weak links to the L2 semantic representation (cf. the RHM; Kroll & Stewart 1994). In contrast, the same stimuli are represented by two separate lexical entries in more proficient bilinguals, one for each language. The latter conjecture is consistent with previous research on highly proficient bilinguals (e.g. Lemhöfer & Dijkstra 2004) and some of the assumptions made by the BIA+ model (Dijkstra & van Heuven 2002a).

Apart from a hypothetically different lexical organisation, lower proficiency L2 speakers display other interesting features that contrast their visual processing in their native language from that of their more proficient counterparts. On the one hand, and most central to the main concern of this research, the present analyses reveal that these low-proficiency bilingual participants benefit more clearly from bottom-up information, in the form of onset capital letters. The fact that this effect is most notable in responses to interlingual homographs could be taken as evidence that orthographic cues act as a kind of language determination mechanism and selectively inhibit irrelevant lexical candidates (Lemhöfer & Dijkstra 2004:547). An alternative interpretation is that this early orthographic information may be even directly connected to language nodes, acting as an early activator of the relevant language (cf. van Kesteren, Dijkstra, & de Smedt 2012). These assumptions challenge the current structure of the BIA+ model (Dijkstra & van Heuven 2002a), "[which] posits that the orthographic level is blind with respect to language membership and that language nodes exist late in the word

[59] Note that the stimuli employed in these studies differed from the current stimulus set. Hence, the present findings can be taken as an addition to our current knowledge of bilingual L1 processing.

recognition stream with no direct connections to orthography" (Libben & Titone 2009:389).

Being less proficient in an L2 also means that participants will use top-down information (from the language of the sentence preceding the target) in a manner that is more marked when compared to more proficient bilinguals. As shown in the analysis of the entire data set, while the L1 has a more strongly positive effect, the L2 has a slightly more negative impact on these less proficient bilinguals. Given that the sentence frame material was relatively open ended (i.e. low constraint), participants are less likely to have responded based solely on the preceding semantic information (cf. Elston-Güttler, Gunter, et al. 2005). Instead, it is possible that depending on the language of the sentence frame, the relevant *language nodes* were activated. In this case, an incorporation of a feedback mechanism from the language node(s) to the orthographic level is crucial (cf. Libben & Titone 2009; Schwartz & Kroll 2006). Clearly, language nodes are only one way to model the language factor. Others include *language tags* assigned to lemmas (Dijkstra & van Heuven 1998; Green 1998a, 1998b) or patterns of activated features in bilingual memory, created as a result of co-occurrence of different elements of word representations within a language (Grosjean 1997). In either of the latter cases, the notion of 'top-down information' may not be a suitable term in this context.

Another possibility is to interpret the current findings according to the task schema account, as implemented in the BIA+ (Dijkstra & van Heuven 2002a) and the IC (Green 1998a) models. Although the original account (Green 1998a) - and the research supporting that view - focuses largely on experimental paradigms where two intrinsically conflicting *task demands* alternate (e.g. language-exclusive tasks; or language-inclusive tasks with alternating bilingual stimulus material, presented in isolation) it could be extended to include the current experimental setup. According to the task schema account, switch costs (i.e. inhibited performance) emerge "[w]hen the task demands shift from task A to task B, [and the] persisting activation of the prior task, A, becomes a form of competitor priming, interfering with the execution of B" (Thomas & Allport 2000:51).[60] This pattern of responses is reflected in my data as a persistently inhibitory effect of an L2 sentence frame (irrespective of L2 proficiency level of a

[60] Note that several researchers point out that the finding of mixed-language costs is difficult to reconcile in a language non-selective view of bilingual processing (e.g. Lemhöfer & Radach 2009; Thomas & Allport 2000). This and other issues will need some more clarification in future.

bilingual speaker). Yet, conflicting task demands do not seem to capture the entire range of the observed effects, particularly when considering the lower L2 proficiency group (see above). In this case, it seems that the more marked responses were at least in part a result of a different organisation of these speakers' (linguistic) lexico-semantic system compared to more proficient bilinguals, or that these lower proficiency speakers employ different resources available in that system during the processing of visual stimulus material. Finally, it has also been suggested that "a learner's level of proficiency affects a 'default' setting of, or the ability to effectively set, the language task schema in the IC or the task/decision system in the BIA+" (Elston-Güttler, Paulmann, et al. 2005:1605). The assumption brought forward by the researchers is that more extensive knowledge of the L2 means that tasks can be better regulated and that language information can be more effectively used (see also Schulpen 2003). Whatever the exact mechanism is, all of the interpretations above suggest, in principle, that

> [t]hough the task/decision system and linguistic systems are posited to operate separately in the BIA+, there seem to be instances where the effects of both systems work in tandem [...]. (Elston-Güttler, Gunter, et al. 2005:68)

Interestingly, this latter conjecture seems also reflected in the correct rejections of nonwords. Nonword rejections are slowed down not only on trials where the task-irrelevant language (L2) precedes stimulus presentation (cf. task schema account), but also when the nonword carries the prominent L1 marker (i.e. an onset capital letter; likely product of the linguistic system). The finding that the combination of both factors returns the slowest responses, i.e. language-specific orthography does not reduce "switch costs", contrasts with the results of an early study by Grainger and Beauvillain (1987), but is in line with more recent research (cf. Thomas & Allport 2000).
Perhaps the most intriguing effect, however, is found in responses to nonword stimuli following an L1 sentence frame, where stimuli carrying an onset capital letter (i.e. stressing the German-like nature of an item) produce a latency cost. A similar effect has been observed in previous studies (e.g. Lemhöfer & Dijkstra 2004; Lemhöfer & Radach 2009), and the suggestion has been made that there must be some mechanism in place which detects the language membership of a visual stimulus prior to its rejection. As a result, the temporal deadline for rejection is set differently depending on certain stimulus criteria. The present study provides evidence that the suggested mechanism may be in place as early as at the sub-lexical level. While it would make sense that such

language-specific information is coded through language nodes, the current structure of the BIA+ model does not support this possibility.[61] In addition, the present data contrasts the common assumption of virtually all computational models that input detectors are largely case, font, and size insensitive (cf. Paap et al. 1984). Admittedly, however, the language marker employed in this research may reflect a special case, being simultaneously a grammatical marker in German, as opposed to many other languages. Nevertheless, I believe that the present data indicate quite clearly that the first processing stages are *not* insensitive to such information (cf. Chauncey et al. 2008; Grainger et al. 2006; Grill-Spector et al. 1999; Holcomb & Grainger 2006; Petit et al. 2006; Sawamura et al. 2005). However, it remains unclear how exactly the mechanism operates. Future research will need to address the question of the precise point at which the influence of such a marker occurs during the decision making process (e.g. perhaps at a later, post-lexical stage).

[61] According to the BIA+, language nodes are only activated by *words* of the respective language, and only late in the word recognition process. In addition, there is currently neither a 'feedback' mechanism which would allow a direct connection to orthography, nor is there an explicit and efficient mechanism feeding directly forward from the sub-lexical orthographic level to a language node system (although according to Dijkstra & van Heuven 2002a, this could be easily implemented in the BIA+).

3.3.4 Summary of Experiment Two

The findings documented in both the error analyses and response time analyses are similar to one another, presenting a fairly analogous picture of bilingual language processing. Given that the experimental task was conducted in the bilinguals' more dominant L1 and faster, more automatised processes are assumed to take place (cf. Tzeglov & Kadosh 2009), the impact of the variables under investigation was both more subtle, or limited only to very specific circumstances (i.e. when unknown items were encountered and the processing system was slowed down).

What we find in this context is that a task-irrelevant (L2) sentence language delays word recognition. In addition, participants who are less proficient in the L2 are slightly more affected by the impact of this factor.

The effects of L2 proficiency are further reflected in the overall accuracy of responses, with poorer performance amongst less proficient bilinguals. More importantly, however, the data suggest that L2 proficiency is a major determinant of the manner in which bilingual speakers employ sub-lexical language-specific information when reading in their L1. More specifically, lower proficiency bilinguals seem to use this low-level information as a key source for their language decision, when processing both ambiguous words (i.e. interlingual homographs) and unknown words (i.e. nonwords). More proficient bilinguals, on the other hand, do not seem affected in the same way, provided that processing is more automatised. When the processing mechanism is slowed down (i.e. for nonword responses), however, language-specific cues seem equally involved in the decision making process.

Finally, based on the differences identified for processing of IHGs, both the response time and error analyses are also crucial in making assumptions about the developing lexical organisation (i.e. with increasing L2 proficiency).

3.4 General Discussion

The primary aim of the present study was to explore the impact of sub-lexical information (i.e. in the form of language-specific cues) on bilingual visual processing. One objective was to assess the impact of such information with respect to the possible priming of a particular language through a preceding context sentence. The other objective was to investigate the extent to which the proficiency level of a bilingual speaker allowed for the emergence of such effects.

To test these aspects, the study employed interlingual homographs, which were presented to one group of participants with a language-specific marker, and to other participants with no such language indicator.[62] In addition, in one of the two experimental sessions the visual target presentation was preceded by an auditory, task-relevant sentence frame, and in the other session target presentation followed a task-irrelevant sentence frame. The same task was conducted with two bilingual populations who had the reverse language order. Accordingly, one bilingual group was conducting the task in their L2 and the other completed the task in their L1. This was done to obtain a fuller picture of the actual use of the available resources (i.e. both language-specific cues and sentence language) to execute a task. Within each bilingual group, proficiency in L2 varied across participants.

The results of the two experiments suggest that the use of available resources to execute a task depends strongly on the automaticity of lexical recognition. This is primarily reflected in the two bilingual groups' similar use of sub-lexical cues - even though in responses to different stimuli - as well as in a changing underlying function of these cues which comes with increasing L2 proficiency. The experiments also show the extent to which language information in form of a sentence frame in a particular language influences the bilingual processing system.

Because the proficiency of a bilingual speaker has been revealed as a key factor underlying bilingual visual word recognition, the following discussion will begin with an account of the effects of proficiency. Next, the impact of each of the two sources of

---

[62] A deferred statistical analysis considered the possibility that Format effects were actually participant effects, i.e. that it just so happened that the participants in the Title case condition performed differently from those in the lower case condition. The only way to test this was to compare participants' performance in the on-line proficiency test. Accordingly, the test scores of participants assigned to either of the two conditions were tested against each other. Neither of the paired *t*-tests for the two groups of bilinguals was returned significant (German-English: $t(32) = 0.60$, $p = 0.55$, and English-German: $t(29) = -1.54$, $p = 0.13$), which suggests that participants in the two conditions behaved similarly; hence, the Format effect is less likely to be an artefact of certain participant characteristics.

language information will be discussed in turn (i.e. sub-lexical language-specific cues and sentence language, respectively). Finally, the results of both experiments will be considered in light of the current models of bilingual processing.

Proficiency

At a more general level - that is, temporarily disregarding its interaction with available language information - one area where the effects of proficiency on bilingual visual word recognition are most prominent in the present study is when bilingual participants complete the task in their L2 (i.e. English dominant, English-German bilinguals). This is mainly reflected in the responses to interlingual homographs: although responses to these items are faster than to their matched controls, less proficient bilinguals are more inclined to reject something that also exists in their L1 (note that while the obtained facilitation effect contrasts some of the previous findings, e.g. Dijkstra et al. 1998; Schulpen 2003, it is most likely a result of the present experimental setup). This observation is not surprising and several related explanations have been put forward to account for this finding. Schwartz and Kroll, for instance, assume that less proficient bilinguals experience stronger activation of form-related lexical competitors from within their own language (2006:208). More specifically, it is possible that during the initial word recognition stages a highly activated L1 reading of an interlingual homograph suppresses the activation of an L2 reading to such a high extent that the resulting competition for selection becomes greatly inhibited and leads to errors (cf. Schulpen 2003:202). This is particularly feasible in the context of the present study, given that the communication between researcher and participant was limited to English (i.e. the bilinguals' L1). As a consequence, the activation levels of the task-irrelevant L1 would have been clearly elevated, exerting a negative impact on the lower proficiency speakers' performance.

Another explanation for proficiency effects on the recognition of IHGs put forward by previous research concerns the relative cognitive control of a bilingual speaker. The assumption is that the higher the L2 proficiency the more cognitive control a speaker has "over (their use of) the relative activation of both their languages in a bilingual situation" (Schulpen 2003:174). That is, in a pure L2 task, controlling the activation of (words in) the bilinguals' two languages becomes less difficult the more proficient one is in the L2. As Schulpen points out, however, control over the relative language activation in both languages becomes very difficult, if not impossible, when a task contains

material from a task-irrelevant language (here, L1; 2003:175). Data collected in the present study does not fully support the latter proposal, given that no obvious IHG processing differences have been found for more advanced bilingual speakers. Again, it is possible that the nature of the current task provided these participants with sufficient additional (language) information to master control.

While the former and other postulations regarding proficiency have been suggested in the context of L2 processing (see, for instance, Schulpen 2003:202, for other explanations for the occurrence of the proficiency effect), the present study extends the previous findings by including data collected from the reverse bilingual group, namely German-dominant, German-English bilinguals. Interestingly, the pattern of responses for bilinguals conducting the task in their L1 reveals a somewhat different picture. More specifically, while less proficient bilinguals seem to be more strongly affected by the general requirements of the task (as reflected in minimally higher error rates to all experimental items when compared to more proficient speakers), it is the more advanced bilinguals who take more time to respond to IHGs as opposed to the matched German controls. Once more, this observation seems to find compelling support in previous bilingual research (e.g. Duyck 2005; van Hell & Dijkstra 2002; Titone et al. 2011; van Wijnendaele & Brysbaert 2002), which has interpreted this finding exclusively in favour of the language non-selective view. Taking into account, however, that at the same time a somewhat lower level of cross-linguistic activation was evident for less proficient bilingual speakers (as reflected in faster responses to IHGs than to controls), it could be argued that the small processing advantage found for this group for IHGs might have a different source. Based on these results, the tentative interpretation which I would like to put forward is that of possible traces of a developmental lexical organisation. Accordingly, in the early stages of L2 acquisition, low proficiency L2 speakers possess either one lexical entry for the different readings of an interlingual homograph, or (more likely) two lexical entries with very weak links to the L2 semantic representation (cf. the RHM; Kroll & Stewart 1994). With increasing proficiency in the L2, the two different orthographic representations of an interlingual homograph become more strongly separated, until (for highly proficient bilinguals) they have fully established, separate entries each of which is connected to its own semantic representations across languages (cf. Lemhöfer & Dijkstra 2004:546; the latter is also consistent with some of the assumptions made by the BIA+ model, Dijkstra & van Heuven 2002a). Consequently, the weak and/or incomplete links to the L2 semantic

representations may be partly what causes the less apparent cross-linguistic activation for lower proficiency bilinguals when they complete a task in their L1. (And the reverse scenario is what is evident above, when lower proficiency bilinguals complete a task in their L2.)

Finally, another explanation for the effects of proficiency has considered the increase in the automaticity of processes underlying bilingual visual word recognition (cf. Schulpen 2003; Tzeglov & Kadosh 2009). As argued by Dijkstra, words that are well known by the bilingual will be recognised in a much faster and more automatic bottom-up way (2005:187). While this assumption may explain the differences observed for bilinguals completing the experimental task in their L2, much stronger evidence for the effects of automatisation seem reflected in the processing differences between the reverse bilingual groups.

An alternative interpretation of the effects of automatisation is that with increasing proficiency level, bilingual speakers employ some higher level resources to execute the task. This conjecture may be best tested in relation to the interplay between proficiency and sub-lexical language-specific information, as will be discussed in the next section.

Language information: sub-lexical language-specific cues

The key point of interest of the present study was to clarify to what extent bilingual speakers can make use of sub-lexical language-specific information to indicate the language of processing. Previous research has posited that there might be an early language discrimination mechanism, which allows for words (as well as nonwords) to be distinguished as belonging to one or the other of a bilingual's two languages before their actual recognition or rejection (Lemhöfer & Dijkstra 2004:547; for a similar argument, see also Lemhöfer & Radach 2009; van Heuven et al. 1998). Results from both experiments provide strong evidence for the existence of such a mechanism. The present results also go beyond those obtained by previous research by identifying crucial factors that may modulate the effectiveness of and/or restrictions on such a language determining mechanism.

Accordingly, the present data suggest that language discrimination may take place as early as at the sub-lexical level (cf. van Kesteren et al. 2012). For the English-German bilinguals (i.e. completing the task in their L2), this was mainly reflected in their responses to existing German words. More specifically, responses were the fastest and

most accurate when a target was presented with an onset capital letter, indicating its (German) language membership. Interestingly, this facilitative effect of onset capitalisation was only found for the less proficient group of speakers. There are several implications of this finding. First, consistent with the earlier notion that less proficient L2 speakers experience stronger interference from form-related lexical competitors (e.g. Elston-Güttler, Paulmann, et al. 2005; Schulpen 2003; Schwartz & Kroll 2006), these less proficient speakers may be more sensitive to language-specific information – in the form of low-level orthographic cues – to facilitate their language decision. Alternatively, on the assumption that processing of an L2 stimulus is less automatic (hence, slower) for these speakers (cf. Dijkstra 2005), sub-lexical information may have more time to exert an influence on the processing system. In this context, Dijkstra has suggested that

> [it is] unlikely that, for words that are well known by the bilingual, this kind of information is often used in a top-down way (i.e. the bilingual notices a particular [sub-lexical cue] that is unique to a language and then uses this information for his or her language decision) because the automatized bottom-up recognition process will usually be much faster. (2005:187)

While this would explain the steady decline of the impact of language-specific cues with increasing L2 proficiency (i.e. based on increasingly automatic processing; see discussion above), the pattern of responses collected from lower proficiency German-English bilinguals does not fully support my alternative conjecture. That is, we can assume that bilinguals completing a task in their L1 will process L1 word stimuli equally automatically, irrespective of their proficiency in an L2 (except if learning an L2 should have an impact on the processing of the L1; something that remains to be further investigated). It follows that we should not detect any major processing differences between speakers within this bilingual group in relation to the presentation format of a target. Surprisingly, the opposite was revealed; namely, I once again found a facilitative effect of language-specific cues, but only for lower proficiency L2 speakers (in both word and nonword responses). This finding seems to suggest that irrespective of the language of processing (i.e. whether the task was an L1- or an L2-LDT) less proficient L2 speakers make different use of (available) resources to execute the task when compared to their more proficient counterparts. In the context of my experiments this was reflected in a clear processing advantage when stimuli were marked for language, which suggests that for these lower proficiency speakers sub-lexical language-specific cues may be a fundamental source of information in the decision making process. Put differently, lower

proficiency bilingual speakers may be more consistently employing the "early language discrimination mechanism" (Lemhöfer & Dijkstra 2004:547).
The present data suggest further that more advanced bilinguals are certainly not oblivious to this low-level information source, but may use it in a different way. Accordingly, more proficient (German-English) L2 speakers were slowed down when responding to IHGs carrying an onset capital letter as opposed to when no such information was provided. A possible interpretation of this finding is that advanced bilinguals are more sensitive towards another function of onset capitalisation in German, i.e. indicating the syntactic category of an item. As a result, an accessed German word (carrying an onset capital letter) may require an additional consistency verification to ensure that the word is a noun (which requires capitalisation). This conjecture seems to find further support in the data collected from (more advanced) English-German participants. Notably, however, in both cases it is strongly defined by the presence of a German sentence preceding the target. The reader will recall at this point that this is a recurring pattern, i.e. where the discussed format effects appeared, they almost exclusively co-occurred with the effects of the language of the sentence preceding the target. This is not surprising given the present experimental set-up and will be discussed in more detail below.

Language information: sentence language

At a more general level, the finding that is evident throughout the present study is that when a task-relevant language (German) precedes the target, then it facilitates responses, whereas a task-irrelevant sentence language (English) has the reverse (i.e. an inhibitory) effect. This finding is consistent with a growing body of research which investigates the effects of (typically manipulated) context on bilingual visual word processing (e.g. van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker 2011; van Assche, Duyck, Hartsuiker, & Diependaele 2009; Libben & Titone 2009).
Among the explanations that have been provided to account for the facilitative effect obtained during both L1 and L2 processing is the notion that participants may use the most minimal information provided to zoom into the language of processing (i.e. a target language; Elston-Güttler, Gunter, et al. 2005). Such minimal information may be presented in form of "a relatively low-constraint sentence" and result in stronger activation of the target meaning, which allows it to compete early for selection (Schwartz & Kroll 2006:209). In contrast, the inhibitory effect (particularly during L1

processing) seems consistent with the view that switching from one language to another in the course of an experiment takes time (for some of the earliest evidence, see for instance Kolers 1966a; Macnamara 1967; see also de Groot 2011, chapter 6, for an extensive discussion of this and related issues, and van der Meij, Cuetos, Carreiras, & Barber 2011, for a recent EEG study on language switching in sentences). Surprisingly, the present data suggest that bilingual speakers completing the task in their L1 seem more affected by a language change during an experimental session than a correspondingly proficient group of bilinguals completing the task in their L2. A possible explanation of this result is that the reverse bilingual groups had somewhat differing expectancy levels regarding the task-irrelevant language: while English-German bilinguals may have been explicitly suppressing their L1 in order to execute the task in the L2, the German-English group of speakers struggled particularly when an unexpected task-irrelevant L2 sentence appeared; hence, we observe more detrimental effects of language change for German-English bilinguals. Importantly, this finding is not a mere reflection of processing differences observed for German-English versus English-German bilinguals per se, but whether the target is in L1 or L2. It follows then that future research needs to take such differences into account by testing the same group of bilinguals who perform a parallel task in both their L1 and their L2.
A similar argument to that of language switching deficits was adopted by the first study to observe IHG interference during bilingual L1 sentence reading (Titone et al. 2011). The researchers included L2 filler sentences in their otherwise largely L1 task, finding traces of interference for total reading time (eye-movement) measures only. Referring to a similar argument by Wu and Thierry (2010), the authors argued that "a more systematic study of how different task-related or language model manipulations affect the time-course of bilingual language processing" is essential to understand how or why the inclusion of L2 filler sentence may boost cross-language activation (Titone et al. 2011:1424).

The present work provides such a systematic study and provides additional information about the bilingual processing system. More specifically, the experiments presented in this chapter show how language information in the form of a task-(ir)relevant sentence language interacts with other language information provided in the experiment, and how bilingual speakers (may) utilise this information.

The body of findings in the present study suggests that the impact of language information depends primarily on the proficiency of a bilingual speaker. Accordingly, less proficient bilinguals use sub-lexical language-specific cues as the primary source for their language decision. This is reflected in the data in faster and more accurate responses to words carrying an onset capital letter. Interestingly, even though the effects observed for the German-English bilingual group (i.e. completing the task in their L1) appear more subtle, they offer a parallel image to the pattern identified for the English-German bilingual group (i.e. L2 task).

The magnitude of the impact of low-level language-specific information on less proficient bilinguals is further exemplified during an L2 task, when no such information is available following a task-relevant sentence language. That is, a preceding German sentence seems to elevate participants' response levels high enough for them to be more inclined to quickly accept an all lower case word as German, while at the same time they more often than not reject these words. Having observed a similar inhibitory effect in their 'context' studies, an explanation put forward by several researchers involved the argument that lower proficiency bilinguals make generally less efficient use of L2 contextual information available to them (e.g. Chambers & Cooke 2009; Elston-Güttler, Paulmann, et al. 2005; Gernsbacher 1993; Schwartz & Kroll 2006). As a result, the decision making process of lower proficiency speakers may be more reliant on resources other than meaning-based information, namely form-related information (e.g. Elston-Güttler, Gunter, et al. 2005; Schwartz & Kroll 2006). Admittedly, however, this seemingly primary role of sub-lexical language-specific information may have also derived from its order of presentation. That is, while the language of a sentence can be expected to have elevated the activation levels of a particular language – and hence participants' response expectations – the presentation format of a stimulus was conceivably more directly involved in the lexical decision making process because it was the final determinant of a language (and the required response). In addition, because sentence frames were presented auditorily, their overall impact on visual word recognition may have been somewhat less compared to that of a visual language cue (inherent in the presentation format of a stimulus). Both possibilities need further clarification in future research.

More advanced bilinguals are also sensitive to language-specific cues. However, these low-level language markers seem to have a different function during the more proficient speakers' cognitive processing. Within the English-German bilingual group, this is

reflected in faster responses to words that do not carry an onset capital letter, particularly following an L2 sentence. As discussed earlier, a possible implication of the finding is that the L2 sentence established some expectation regarding the underlying characteristics of the following target item (e.g. syntactic category, presentation format). As a result, an accessed German word (carrying an onset capital letter) required an additional consistency verification to ensure that the word was a noun (which requires capitalisation). This conjecture seems to be confirmed by two other findings: (1) the absence of a similar effect in the responses following an L1 context; and (2) a similar trend observed for German-English participants' responses to IHGs.

Finally, when processing is less automatic (i.e. when participants are responding to nonwords), the effects of language information seem to have a similar impact on all bilingual speakers, irrespective of their proficiency. While the same resources are identified as crucial for the decision making process, the direction of the effects has, as can be expected, somewhat shifted. That is, both experiments provide clear evidence for the primary role of sub-lexical language-specific cues, since items carrying an onset capital letter (irrespective of the condition or bilingual group) took the longest to reject. Importantly, the absence of a similar effect in the English-German bilinguals' accuracy analysis indicates that participants' responses were not simply strategic (i.e. saying "yes" whenever they detect language-specific orthography), but required explicit access to lexical information in order to make their response (cf. Thomas & Allport 2000:46). Note that even though lower proficiency German-English bilinguals were inclined to use onset capitalisation as a language determinant for their nonword responses (i.e. in an L1 task), I believe that this is a reflection of the importance of this low-level cue and the resources utilised by lower proficiency speakers, rather than being a processing strategy.

These effects of sub-lexical language information interacted consistently with the language of the sentence frame preceding the target. More generally, the pattern of responses suggests once again that participants used this kind of language information only as a secondary source: English-German bilinguals used a task-relevant sentence language as a means of zooming into the language of processing. However, this only showed up in combination with sub-lexical cues. This is reflected in the slowest responses in a fully German-biased condition (i.e. German sentence frame followed by a Title case item), and the fastest decisions after a German sentence frame when no sub-lexical cues were provided. German-English bilinguals, on the other hand, were

particularly affected when task-irrelevant language material was included. As discussed earlier, there are two major implications of bilingual visual processing that can be deduced from these findings and that are congruent with the response pattern obtained for existing words. First, the findings give clear evidence that simultaneous processing of two languages imposes an additional cost on the speed with which a decision can be made (e.g. von Studnitz & Green 1997). More importantly, however, the findings support the idea that sub-lexical cues are a legitimate language discrimination mechanism (cf. Lemhöfer & Dijkstra 2004; Lemhöfer & Radach 2009), which has an impact on bilingual visual word recognition, even when processing is done in the L1.

A final issue to be resolved is the question of how the present results can be interpreted in the context of current models of bilingual visual word processing. The next sections will attempt to give a tentative explanation in relation to the main theoretical models described in an earlier chapter (see section 2.3.2).

Theoretical explications

Before we turn to the effects of language information and their potential source in the bilingual language processing system, we should consider the prominent finding of the impact proficiency has on bilingual visual word recognition. None of the implemented models to date have explicitly modelled changes in the bilingual processing system with increasing proficiency in an L2 (see also Kroll, van Hell, Tokowicz, & Green 2010, for a recent appraisal of this model; for a theoretical account, see the RHM developed by Kroll & Stewart 1994). This does not mean that none of the models can account for the effects of L2 proficiency; quite the contrary. The assumption modellers usually make is that "an increase in proficiency may primarily be accompanied by stronger L2 representations (and perhaps associated higher resting levels of activation) in the bilingual lexicon" (Schulpen 2003:180–181). Some researchers also theorise that "[less proficient] bilinguals have difficulty in the binding of the activity in the word identification system to the required responses in the task situation" (Schulpen 2003:202). While the present data seem consistent with either view, they go beyond the previous research by showing that both changes in L2 proficiency and changes in the level of automaticity of the processing system seem to entail different requirements posed on a bilingual speaker, and consequently the resources they may utilise to successfully execute the task.

Let us consider the first scenario, i.e. changes in L2 proficiency. Less proficient participants in both experiments showed compelling use of sub-lexical language-specific cues for their language decision. As discussed earlier, this is consistent with the assumption made by Dijkstra who argues that

> language membership could in principle already be determined before recognition because of the presence of 'language cues' in the items themselves. […] In such cases, lexical search might be limited to the relevant target language from the very beginning. (2005:187)

Dijkstra further provides an explanation for why these cues are not employed in the same way by more proficient speakers, assuming that more "automatized bottom-up recognition process[es] will usually be much faster" than a foregoing language delimitation by means of some top-down processes (ibid.). Instead, the present data show that these more advanced bilingual speakers may be employing this sub-lexical information in highly restricted situations and at some higher processing level, presumably to verify the syntactic category of an input word (cf. Green 1998b:101). What we can deduce from both findings is that language-specific sub-lexical cues form part of the linguistic context effects, as defined by the BIA+ model (Dijkstra & van Heuven 2002a:187). The question that arises in this context is how these sub-lexical cues exert influence on the bilingual word recognition system. An early explanation provided by the BIA+ model is that it is the information inherent to an orthographic input that determines the amount of lexical activation across languages. Accordingly, already a minimal deviation from a full overlap of an orthographic input code across a bilingual's two languages can restrict the initial set of activated lexical candidates to one language only (Dijkstra & van Heuven 2002a:182–183). This interpretation rejects the possibility that lexical activation is determined by a word's language membership (by means of a language node). In fact, the authors argue that "language information becomes available rather late during (isolated) bilingual visual word recognition, usually too late to affect the word selection process" (Dijkstra & van Heuven 2002a:186), and that its potential effects would, in any case, be too weak given the one-to-many mapping from the language node to lexical representations.

The absence of a mechanism to feed information from the language nodes back to the orthographic level (or rather the claim of a late language information) has been contested by a number of researchers (e.g. Green 2002; Libben & Titone 2009; Schwartz & Kroll 2006; but see Schulpen 2003:180–183, for contrasting evidence and

argumentation). This brings me to an alternative reading of the current results and how the effect of sub-lexical language-specific cues can be translated within the bilingual language processing system. As implied elsewhere in this study, sub-lexical cues could, in principle, signal language membership via language nodes (or language tags). This means that information on language membership would become available before the actual word identification, so that it could affect a response very early on. This view is consistent with Dijkstra's post-BIA+ postulation (see above), as well as with some empirical evidence (e.g. von Studnitz & Green 2002b), and is based largely on some of the assumptions made by the IC model (Green 1998a, 1998b; a similar mechanism providing top-down modulation of lexical activation is also found in the BIA model, Dijkstra & van Heuven 1998, 2002b). As the reader will recall, the IC assumes that in a language-specific lexical decision task two language task schemas are in competition. (In the current task, one task schema is for German - where real German words have to be mapped on to a "yes" response - and the other is for English, which maps any nonwords or real English words on to a "no" response.) I have stated earlier that (particularly in relation to IHG responses) a correct response is made to a word when activation in the "yes" units (or language task schema) reaches a criterial level above the activation of the "no" units (cf. von Studnitz & Green 2002a). According to the IC, reaching a required criterial level can be achieved by activating the units coding for language membership (i.e. a language tag).[63] Accordingly, words carrying an orthographic language marker would have boosted the activation levels in the "yes" units to the criterial level above that of "no" units, while the reverse effect (i.e. boosting of the level of activation in the "no" units) was likely to occur with no language-specific marking. An alternative (but preferred) account put forward by the same researchers is that participants can vary their decision criterion for a "yes" response when there is a considerable amount of conflicting information (for instance, a German word indicating a "yes" response and the absence of a language-specific cue indicating a "no" response). As suggested by von Studnitz and Green,

[63] Note that on the basis of the *locus of control*, von Studnitz and Green (2002a:2) refer to such an internal mechanism as the *internal* locus of control. In contrast, the assumption that participants can change their response criterion based on signals from the lexico-semantic system concerning language membership has been referred to as the *external* locus of control. While the question of the locus of control with respect to the availability of language specific cues is a particularly intriguing one, the present study had no means to test this aspect. Clearly, this question provides an interesting investigation area for future studies.

> [a]dapatation arises on this view because individuals shift their decision criterion in the light of the stimuli presented. That is, they reduce the amount by which activation in the "yes" response units must exceed that in the "no" response units. (2002a:17)[64]

The findings of the present study are consistent with either of the two accounts (i.e. those discussed within the BIA+ and the IC frameworks), as well as some assumptions made by previous research (e.g. Grosjean 1997; von Studnitz & Green 2002b:249). This is reflected in similarly fast and accurate responses to words carrying an onset capital letter – irrespective of the language of a target preceding sentence frame – as well as in slower (and more erroneous) responses to items with no such language information, particularly following a German sentence frame. The present findings also extend previous research by showing that these effects are more pronounced for lower proficiency L2 bilinguals, presumably because words are less well known by these speakers and lexical processing is less automatic. The same effects then emerge for nonword responses – which brings me to the aspect of automatisation mentioned before – where all participants (irrespective of their proficiency) seem to revert to the same (sub-lexical) resources in order to execute the task. (Despite my assumption that nonwords are processed in a less automatic fashion compared to words, I acknowledge that there may be other factors affecting the processing of nonwords, rather than the level of automaticity; however, currently too little is known about the processing of nonwords to make some final statements.)

A final issue which needs to be clarified is the interaction of a task-(ir)relevant sentence language with the effects of sub-lexical cues. While the finding that the presence of a task-relevant language limits non-target language interference and the availability of a task-irrelevant sentence language has the reverse effect (i.e. increasing cross-linguistic activation) seems consistent with previous studies (e.g. Elston-Güttler, Gunter, et al. 2005; Libben & Titone 2009; Schwartz & Kroll 2006; Titone et al. 2011), it also confirms the previous argument that "context language as such [may] not operate as a very effective factor for lexical selection" (Dijkstra 2005:187). This was evident in both word and nonword responses, for which the emergence of an effect was primarily associated with the presence or absence of a language-specific cue. A possible explanation of this

[64] Note that the BIA+ adopts a highly similar lexicon-external task/decision system, which could account for the current processes in a very similar (if not the same) way. According to De Groot, "[t]his is no coincidence because in developing BIA+ its builders have been strongly influenced by [the IC] model, so much that the two models now strongly resemble one another" (2011:181).

result is the conceivably primary role of sub-lexical language-specific information in bilingual visual word processing, over and above that of sentence language. Alternatively, it could be argued that the order of presentation of language information did not allow the effects of sentence language to emerge as a strong factor (see discussion above). Finally, it could also be argued that given the relatively low constraint sentences used in this study, this top-down information did not reveal the same processing benefits reported elsewhere (e.g. Schwartz & Kroll 2006; but see Libben & Titone 2009, for a contrasting outcome). All of these aspects provide interesting possibilities for further investigation in future research.

To conclude, the present study has provided several insights into the kind of language information utilised by bilingual speakers during visual word recognition. First, the present data are the first to show that diacritical markers (i.e. sub-lexical language-specific cues) are an important ingredient in determining the language of processing. It is further shown that the magnitude of the impact of this sub-lexical information is highly dependent on the automatisation of the word recognition process (this challenges some of the claims made by Schulpen 2003:181). More specifically, bilingual speakers utilise this language-specific information more consistently as an early language discrimination mechanism when lexical processing is less automatic, i.e. in responses to nonwords and with low L2 proficiency. Finally, the current findings also demonstrate that the language of a sentence preceding the target had a relatively minor but significant impact on the overall processing outcome. This was primarily reflected in some interference from the non-target language when a task-irrelevant sentence frame was included in the experiment; however, such cross-language effects were largely attenuated in the presence of language-specific information. Although more research is needed to further confirm the conjectures postulated in the present research, these initial results may be informative with respect to the kind of resources employed by bilingual speakers to reduce cross-language effects and how effectively these resources are employed with developing proficiency in an L2.

Although the present study provides important new insights into the early (sub-lexical) processes underlying bilingual visual word recognition, a number of questions remain open. For instance, the current study focuses largely on interactions at the orthographic level. Similarly, some of the cross-language effects observed and discussed in the

present chapter have also been studied at a different level, namely at the (sub-)lexical phonological level, although to a somewhat lesser degree. To advance the theory, we next need to understand how information between the sub-lexical orthographic and phonological levels is transmitted across languages, and whether the competitive interaction between these two levels can affect visual word recognition in a measurable way. Another aspect identified in the present study is the tentative notion that L1 processing is affected by L2 exposure. This theoretically very important aspect requires further clarification by comparing two complementary bilingual groups (e.g. German-English and English-German) with varying proficiency level, who both perform a task in their L1 and in their L2. Finally, the finding of the present study that less automatic processing entails the use of different resources to execute a task is something that may be further confirmed in the amount of phonological involvement during visual word recognition. These and other issues are the focus of the next chapter, which aims at shedding more light on the cross-language effects inherent to that sub-lexical phonological processing level.

Chapter 4
THE ACTIVATION OF SUB-LEXICAL PHONOLOGY AND ITS INTERACTION WITH/MODULATION OF ORTHOGRAPHIC PROCESSES

4.1 Introduction

Little is known about the interaction between orthography and phonology in bilinguals. One way of testing it is by looking at processes involved during silent reading. To address this question, two experiments have been designed for this study, exploring three related theoretical issues: (1) Is L1 and/or L2 phonology activated when bilinguals read silently in either language; and if so, to what extent does the activation of some phonology affect the reading process? (2) Can phonological decoding vary across the different languages of a bilingual user, depending on how distinct the two languages are from each other; and if so, would this be reflected with respect to the language of processing (i.e. L1 or L2)? (3) How does phonological decoding develop with proficiency in the second language (L2)?

The following sections will begin with an overview of our current knowledge about the activation of L1 and L2 phonology and how they may interact across languages. These sections will be followed by Experiments Three and Four, which address the identified research questions from a bilingual perspective with a reversed language order, respectively. The chapter will conclude with a general discussion section, relating the empirical data to current models of bilingual visual word processing.

4.1.1 From orthographic representations to phonological retrieval – 'the neglected role of phonology'

The majority of bilingual studies investigating the relative importance of phonology have focussed to date on the cross-linguistic assessment of two effects typically found in earlier monolingual research: the pseudohomophone effect and the homophone effect (for a review of relevant monolingual literature and findings, see section 2.2.1.2 above). With regard to the pseudohomophone effect, some researchers have shown that word recognition in an L2 (e.g. of the French word *sauce*) can be facilitated when the target is preceded by an L1 homophonic nonword prime (e.g. the Dutch pseudohomophone *soos*; Brysbaert et al. 1999). Other research obtained evidence that the rejection of an L1

pseudohomophone in an L2 lexical decision task is inhibited in comparison to regular nonwords (i.e. not homophonic to either the L1 or L2; Nas 1983, experiment 2). One implication of these early results was that phonology is automatically activated in bilingual visual word processing, even when this is not explicitly required by the task. Given that the pseudohomophones employed in this previous research were only homophonic according to L1 spelling-to-sound (or, grapheme-phoneme correspondence/GPC) rules, another important implication of the research was that "L1 GPC rules were processed even though the participants were performing a task in their L2" (Duyck 2005:1341). In other words, the findings can be taken as evidence that the widely assumed non-selective account of the bilingual language system also incorporates the (sub-)lexical phonological level and the GPC system.

Subsequent work employing pseudohomophones has provided further insights. First of all, research has shown that similar priming effects can be obtained during L1 word recognition when the pseudohomophone prime is homophonic according to the L2 GPC rules (van Wijnendaele & Brysbaert 2002). In addition, there is some evidence that the observed priming effects may occur irrespective of a bilingual's proficiency in an L2 (Duyck, Diependaele, et al. 2004; but see Brysbaert et al. 1999; Gollan, Forster, & Frost 1997, for contrasting results). Finally, research has also shown priming of targets in both L1 and L2 using cross-lingual pseudohomophones which are also translation equivalents of the target. However, when the pseudohomophone prime shares an associative link with the target, priming effects emerge only if the word recognition task is being carried out in the L2 (Duyck 2005). For instance, in a series of experiments, Duyck (2005) showed that the L2 English target *night* could be primed by a masked L1 Dutch prime which was a pseudohomophone (*nagt* [naxt]) of the target's L1 translation equivalent (*nacht* [naxt]). A similar result was obtained when the prime was an L1 (Dutch) pseudohomophone (*dach* [dax]) of an L1 word (*dag* [dax], meaning 'day' in English) which was in turn semantically related to the L2 target (*night*). In contrast, an L2 (English) pseudohomophone prime (e.g. *mowse*) whose real word partner in L2 (*mouse*) was associatively related to the translation equivalent (L2 *cat*) of the L1 target (*kat* in Dutch) did not considerably facilitate the recognition of that target, although priming effects were obtained in the reverse scenario (i.e. where the pseudohomophone and semantic associate were in L1 and the target in L2).

A largely equivalent pattern of findings was collected in research using interlingual homophones as the critical stimuli. For instance, in an early study by Doctor and Klein (1992), English-Afrikaans bilingual participants had to decide whether a visually presented stimulus, which in some cases was an (imperfect) interlingual homophone (e.g. *brick* [bɹɪk] vs. *briek* [bɾik], meaning 'to stop' in English, or *lake* [leɪk] vs. *lyk* [leɪk], meaning 'to look'), was an existing word in either of their two languages. One of the findings of the study was that participants produced less accurate and slower responses to interlingual homophones than to matched control items.

Subsequent studies obtained very similar results using different language pairs, stimulus lists and/or experimental tasks (e.g. Dijkstra et al. 1999; Duyck 2005, experiment 5; Lemhöfer & Dijkstra 2004, experiments 1 & 3; Smits, Martensen, Dijkstra, & Sandra 2006; Smits, Sandra, Martensen, & Dijkstra 2009). The collective results from these studies were taken to support the previously mentioned language non-selective access view. That is, a visually presented orthographic input is assumed to give access to cross-language phonological and semantic representations which, depending on the experimental task and demands, result either in an inhibitory effect (due to competition of the lexical codes) or in facilitation (in various priming paradigms, such as translation or associative priming).

More recently, however, a purely non-selective view of phonological retrieval was challenged, primarily in relation to whether the task was conducted in the L1 or L2 of a bilingual speaker (Haigh & Jared 2007; for a similar argument, see Jared & Kroll 2001). Using English-French homophones (e.g. *sank-cinq*; note again the imperfect homophony of many supposed interlingual homophones) in an English L1 lexical decision, the authors observed no cross-language effects for bilinguals with little regular use of their L2 French (for a similar result using a priming paradigm, see Duyck 2005, experiment 6), even when the participants were explicitly prompted with French material either prior to or during an experimental session. The interpretation put forward by the authors acknowledged two different scenarios, namely "either that bilinguals reading in their first language did not activate phonological representations in their second language at all or that these representations were not activated sufficiently to have much impact on lexical decision performance in their first language" (Haigh & Jared 2007:634). Given the large body of evidence supporting non-selective access (see above and the introduction to Study One), the authors' second interpretation seems more likely.

Another unexpected finding of Haigh and Jared's (2007) study was the emergence of a strong *facilitatory* effect for the same experimental material, when the task was conducted with bilinguals with French as their L1, i.e. in the participants' L2 (English). In an attempt to reconcile their results with contrasting evidence from previous monolingual and bilingual research (i.e. consistently revealing robust *inhibitory* effects; see above and section 2.2.1.2), the authors assigned a key role to the (minimal) amount of cross-linguistic competition at the orthographic level. More specifically, the authors assume that when bilinguals were presented with an L2 word (English; e.g. *sank*), its phonological representation fed information back to the orthographic level, co-activating the non-target L1 representation (French; e.g. *cinq*). However, they argue that the amount of activation "provided too little competition for English orthographic representations because their orthography was too dissimilar" (Haigh & Jared 2007:638). Hence, this competition did not produce a significant inhibitory interlingual effect (cf. Lemhöfer & Dijkstra 2004).

There are, however, several problematic issues which have not been addressed by Haigh and Jared (2007) in their interpretation. For instance, although the authors assign an important role to the orthographic (dis)similarity of their cross-linguistic material[65], no actual measurements of similarity were made. Accordingly, the list of critical stimuli includes French-English homophones that appear to be orthographically more similar (e.g. *il-ill*, *faire-fare*; and again, the imperfect homophony is striking) and some that are a lot less similar across the two languages (e.g. *oui-wee*, *chaque-shack*). It may be that the limited number of critical stimuli employed in their research (i.e. 20) did not allow the authors to explicitly control for this factor.

Another critical issue which has been left unexplained is the fact that the researchers manipulated the frequency of their items, with the frequency of the French (here, L1) reading of the homophone kept considerably higher than the English (L2) one. Given that the relative frequency of the cross-language material employed in a task has been repeatedly shown to affect bilingual processing (see Lemhöfer et al. 2008, for a brief overview), it is surprising that no interpretations encompassing the effects of frequency were offered, particularly in the L2 context.

[65] Note that other researchers have assigned a similar key role to cross-linguistic orthographic (dis)similarity in bilingual visual processing, using either similar or different experimental material (e.g. Dijkstra et al. 2010; Doctor & Klein 1992; van Assche et al. 2011).

Finally, Haigh and Jared (2007) also consider a feedback mechanism for the interpretation of their results. As outlined in an earlier chapter (section 2.2.1.2.2), feedback effects have been typically associated with strong competition between two orthographically dissimilar representations activated by means of an overlapping phonological code. This competition is assumed to be the cause of inhibited performance (e.g. Stone et al. 1997; Ziegler et al. 1997). Surprisingly, Haigh and Jared's argument suggests that the authors assume no such inhibitory competition despite a high degree of dissimilarity of the activated orthographic representations (i.e. *cinq* and *sank*). On the contrary, the authors seem to support the idea of target language facilitation because "[the items'] orthography was too dissimilar" (Haigh & Jared 2007:638). This view is inconsistent with the claims made by virtually all previous research. Perhaps a better way to understand Haigh and Jared's results is to assume that the information which was fed back to the orthographic level was not sufficient to spread across languages (see the discussion in the present study). Clearly an important issue, the effects of feedback (and feedforward) activation are discussed further in the next section.

4.1.1.1 Feedforward and feedback effects in bilinguals

As we have seen in the previous section, there are two major findings from existing bilingual research that focuses on phonological processing. First, most research suggests that spelling-to-sound rules from the language not being attended to are automatically applied during target stimulus processing. This finding is in line with the view that there is non-selective access to both languages. Second, performance on a target stimulus seems to be influenced by phonological similarity between that stimulus and a word from the nontarget language. This phenomenon has been more specifically referred to as the *feedforward consistency effect*. According to this view, when a visual input maps on to multiple pronunciations it causes competition between the activated codes, which in turn inhibits performance.

A related issue, first suggested just over a decade ago, is the notion that the relationship between orthography and phonology in the visual word recognition system does not operate in a strictly unidirectional manner, i.e. proceeding from print to sound (e.g. Grainger & Ziegler 2007; Jacobs et al. 1998; Stone et al. 1997). Instead, it has been suggested that orthography and phonology have a bidirectional influence on visual word identification, including feedforward and feedback consistency effects (for a detailed review, see Chapter 2, section 2.2.1.2.2). Recognising the possibility of such dynamics

across two languages - in bilingual visual processing - Brysbaert et al. acknowledged that "the question is whether simultaneous phonological recoding according to different languages is possible on a pure feedforward basis or requires feedback mechanisms from higher levels of processing [such as those suggested by Stone and colleagues]" (1999:146).

Despite an early and recurrent recognition of the feedback effects in purely visually-based tasks (e.g. Brysbaert et al. 1999; Duyck 2005; Haigh & Jared 2007), only one study to date has explicitly investigated whether the bidirectional processing hypothesis would hold in a cross-linguistic context, i.e. in bilingual word recognition (Schwartz et al. 2007). In their study, Schwartz and colleagues asked English-Spanish bilingual speakers to read out cognates (words that typically have overlapping orthography, phonology and semantics across languages; e.g. English *fruit* – Spanish *fruta*), which were divided into two blocks; one block contained experimental items only from the L1, and the other block was exclusively limited to the L2. To assess how the requirement to name a cognate reflects the activation of orthographic and phonological representations across languages, the researchers manipulated the degree of cross-linguistic overlap of their items, which were either more or less similar in their spelling and/or pronunciation. Apart from the typical finding that cognates were recognised and named more quickly than non-cognates in both the L1 and the L2, the authors found that when the orthographic forms were similar in the two languages but phonology was not (e.g. English *horror* ['hɒɹə] versus Spanish *horror* [o'roɾ]), naming latencies were slower (inhibitory, feedforward effect). When the orthographic forms were dissimilar, naming latencies were slower for stimuli with a high degree of phonological overlap (e.g. English *sweater* [swetə(r)] – Spanish *suéter* [sweteɾ], an inhibitory feedback effect), although the reported response delay failed to reach conventional statistical significance.[66] First and foremost, the authors interpreted these results as supporting language-independent lexical access. Moreover, even though actual evidence for feedback activation is contestable in their study, Schwartz et al. took their findings together with the strong

[66] Note that the feedforward consistency effect has been frequently reported in early monolingual research, which also used the naming task (Ziegler et al. 2008:643). Hence, it is not surprising that Schwartz and colleagues obtained a similar effect in their bilingual study. Similarly, the absence of a clear feedback effect may reflect the difficulties of previous research in obtaining such an effect. Alternatively, given that the task was completed in both the participants' L1 and L2, several uncontrolled variables (e.g. Spanish word frequency measure, neighbourhood density, etc.) may have also contributed to the absence of an effect.

evidence for feedforward activation to suggest that the degree of consistency between orthographic and phonological codes influences cross-lingual competition (2007:206).

Given that Schwartz et al. present the first explicit investigation in a bilingual context of the bidirectional consistency effects previously observed in monolingual contexts[67], it is worth looking in detail at a number of issues concerning their study. One such issue concerns the experimental task. While some researchers have previously employed a naming task to tackle the question of bidirectional consistency effects in a monolingual setting (see Ziegler et al. 2008, for a recent review of relevant studies), original evidence for this phenomenon (as well as numerous replications of it) comes from an exclusively visual task (e.g. Stone et al. 1997; Ziegler et al. 1997). Empirical evidence from tasks using an entirely visual paradigm is particularly valuable, since in addition to revealing the existence of feedback consistency effects, this evidence provides further support for automatic phonological decoding of visual stimulus material. Accordingly, one major aim of the present study is to determine whether word identification is similarly influenced by the differences in the consistency of the orthography-to-phonology mappings across languages, when bilinguals are not explicitly required to access phonology.

Two other aspects of Schwartz et al.'s (2007) study worth further consideration are closely related to this task issue, as well as to each other: one is the finding that the same pattern of results emerged in the bilinguals' two languages, and the other is the fact that the researchers employed a combination of languages (English and Spanish) that is not only less commonly used in bilingual research, but also a language pair that differs considerably in their orthographic depth. Given that both issues are central to the interests of the present study, they will be discussed in fuller detail in the next section.

4.1.1.2 The orthographic depth hypothesis and the language of the task

Recapitulating the findings of past research, Ota, Hartsuiker, & Haywood argued recently that "there are indications […] that phonological mediation in bilingual visual word

[67] Strictly speaking, Schwartz et al.'s (2007) research may not qualify as an exact extension of the bidirectional consistency effect to the bilingual case, because of their item choice. That is, phonologically or orthographically 'inconsistent' items employed in monolingual studies have typically multiple divergent meanings, depending on their form or pronunciation. In contrast, Schwartz et al. used items with highly similar meanings across languages, limiting the 'inconsistency' of items to their orthographic and/or phonological representations only. While the researchers do not explicitly discuss their item choice, they argue that these items are well suitable to investigate the competitive activation dynamics between the orthographic and phonological level across languages. For further discussion of this issue, see section 4.2.1.1.

recognition is conditioned by the orthographic systems of the two languages involved" (2010:385). Interestingly enough, the researchers referred specifically to language pairs that do not share the same script; for instance, a non-alphabetic and an alphabetic language pair (e.g. Chinese and English, respectively), or two alphabetic languages with markedly different scripts (e.g. Hebrew and English; see, for instance, Gollan et al. 1997; Kim & Davis 2003; Ota, Hartsuiker, & Haywood 2009).

Although research has shown that when a bilingual language combination shares the same script "the grapheme-phoneme conversion (GPC) rules of the two languages also seem to affect each other" (Ota et al. 2010:385; see also discussion above), to my knowledge no studies to date have investigated whether and to what extent these cross-linguistic dynamics are conditioned by the orthographic depth of the languages involved.

Orthographic depth refers to the degree of grapheme-to-phoneme correspondence in a language, or the consistency of an alphabetic writing system. The degree of grapheme-to-phoneme correspondence can be seen as a continuum, along which the different languages are distributed, depending on their particular mapping system. Thus, the distribution ranges from languages with homomorphic (many-to-many) mappings (e.g. English, French) to languages with isomorphic (nearly one-to-one) mappings between letters and speech sounds (e.g. German, Spanish). This distribution of languages which share the same (alphabetic) script, yet have different grapheme-to-phoneme and phoneme-to-grapheme mappings, has frequently been referred to in the context of the *Orthographic Depth Hypothesis* (ODH; Katz & Frost 1992). Here, homomorphic and isomorphic languages are referred to as 'deep' or 'opaque' versus 'shallow' or 'transparent' languages (cf. Schepens, Dijkstra, & Grootjen 2012).

Recent evidence for the importance of the orthographic depth of a language comes from extensive linguistic and psycholinguistic research on monolinguals. More generally, researchers have observed that the consistency of alphabetic writing systems has an impact on the development of reading skills and the strategies employed during reading (e.g. Borgwaldt, Hellwig, & de Groot 2005; Perry & Ziegler 2002; Vaessen et al. 2010; Ziegler & Goswami 2005). More central to the focus of the present study, however, is the fact that researchers draw a direct connection between the orthographic depth of a language, the relative contribution of phonological decoding processes and other processing strategies during visual word identification; the nature of the phonological

decoding process (also referred to as reading strategy in this study) is what has sparked an on-going debate.

For instance, early accounts of the ODH argued that the more consistent relationship between script and sound in transparent orthographies strongly encourages the use of phonological decoding processes. Because reliance on a grapheme-phoneme decoding mechanism in opaque orthographies can often lead to errors, the recognition of words should rely more on the input's visual orthographic structure. Hence, this ('strong') interpretation of the ODH assumed a stronger involvement of phonological decoding strategies in transparent than in opaque orthographies (e.g. Katz & Feldman 1983; Katz & Frost 1992).

Given the large body of evidence in favour of automatic phonological activation, more recent discussions have moved away from the idea that phonological decoding is largely limited to more transparent orthographies, to suggest that it is equally available for processing in all languages, irrespective of their orthographic depth (for a comprehensive discussion of the ODH and an extensive overview of relevant research, see Frost 2005:277ff.). Instead, what is assumed to differ for processing in languages with varying orthographic depth is the "nature of the phonological recoding process" (Ziegler et al. 2001:379). Perhaps the most crucial piece of evidence for this proposal comes from cross-language naming studies which were able to show that identical words were processed differently in languages that differ in their orthographic depth, i.e. German and English (e.g. Goswami, Ziegler, Dalton, & Schneider 2001; Ziegler et al. 2001; see also Ziegler & Goswami 2005, for a review of relevant studies). Based on their findings, the researchers suggested that reading aloud in a language with a *less consistent* orthography may entail phonological processing units larger than phonemes (e.g. syllables, onsets, rhymes) and other, more flexible decoding strategies (e.g. using lexical properties or semantics; cf. Lemhöfer et al. 2008). Similarly, researchers collecting evidence from entirely visual tasks have speculated that readers of such languages "are encouraged to access their lexicon with underspecified phonological representations" (Frost 2005:285). In contrast, readers of orthographically consistent languages may rely primarily on distinct grapheme-phoneme decoding strategies or, alternatively, "lexical access [may] be based on a relatively detailed phonological representation" (ibid.).

Given that language differences relating to orthographic depth have been identified in monolingual processing, a question that immediately springs to mind is whether bilingual speakers are similarly affected by the orthographic depth of their two languages. It is possible that bilinguals construct reading strategies that are effective for reading in both languages. This conjecture would assume one common GPC system. Alternatively, even though it would be less economical, bilingual speakers may have different sets of reading strategies at their disposal, one for each language, which they are able to apply depending on the language of processing. Such a proposal would assume two (possibly separate) GPC systems.

To address the question of an effective reading strategy, and hence the nature of phonological recoding in bilinguals, an identical experiment was conducted with two bilingual groups whose language pair was fairly distinct in terms of their orthographic depth, i.e. German (shallow) and English (deep). The two bilingual groups differed with respect to their dominant language; in Experiment Three the bilinguals had English as their L1, and in Experiment Four the bilinguals' L1 was German. Each bilingual group underwent half of an experimental session in their L1 and the other half in their L2.

In addition, if an effective reading strategy affects the degree of detail generated in a phonological representation, then a consequence of a particular reading strategy in bilingual processing may be that it has an impact on the bi-directional flow of information (i.e. between the orthographic and the phonological level) across two languages. This in turn may be reflected in the amount of inhibition and/or facilitation between items from different languages. This aspect was explored in this study by including cross-linguistic material which was manipulated at the level of orthographic and/or phonological overlap across languages.

4.1.2 Proficiency

A final issue is the role of L2 proficiency on the manifestation of the issues discussed above. This may appear as a rather challenging task, given that there seems to be a lack of agreement on the more general activation status and the effects of L2 phonology to date.

For instance, in an early study Gollan and colleagues argued that less proficient bilinguals (as defined by a high number of errors they made) "rely more heavily on phonological computation of L2 words" (1997:1137). Using a masked translation

priming paradigm with Hebrew-English bilinguals, the authors determined that L2 (English) target recognition was facilitated by the presentation of an L1 (Hebrew) cognate prime compared with a (phonologically dissimilar) control prime. In addition, the researchers found that this effect was much larger with less balanced bilinguals than with those who were more balanced. Given that the different writing systems meant that relying on the orthography of one language when processing the other was not of any use, the effect must have originated in the cross-linguistic overlap at the phonological level. This was shown to have a greater impact on less proficient bilinguals.

While subsequent research could not fully substantiate the hypothesised over-reliance of less proficient bilinguals on phonology in L2 reading (e.g. Brysbaert et al. 1999), more recent research has claimed the opposite, i.e. that "effects of L2 phonology are more likely to be observed in proficient bilinguals who can rapidly activate lexical codes from their second language" (Schwartz et al. 2007:120). Interestingly, Schwartz et al. did not employ different bilingual groups in their study. Instead, they conducted the same experiment in their English-Spanish bilinguals' L2 and L1, obtaining very similar response patterns in both languages. Hence, their claim that L2 phonology effects might be more likely detectable in more proficient bilinguals. Even though their bilingual participants were reportedly fairly proficient in their L2, the lack of a complementary bilingual group with L1 and L2 reversed makes their argument seem less convincing.

Finally, another argument presented by previous research is that the effects of phonology do not seem to be modulated by L2 proficiency at all (Duyck, Diependaele, et al. 2004). Using a masked phonological priming paradigm, the researchers obtained a similarly large cross-lingual priming effect for both balanced and unbalanced Dutch-French bilingual participants.

Despite invaluable first insights from these studies, it is clear that more research is needed to arrive at some definite answers. As discussed in the previous section, one aspect that needs to be considered in this context is the orthographic depth of the examined language pairs. If the nature of the phonological decoding process differs across orthographies, bilinguals in the studies described above may have been employing different strategies, either depending on the language of processing or the characteristics of their bilingual GPC system; hence, the conflicting results.

Another aspect which past research has not paid much attention to is the changing effects of phonology with gradual increases in L2 proficiency (instead, the studies typically employed fairly advanced bilingual speakers). One exception is the study by

van Wijnendaele & Brysbaert (2002), which acknowledges the potential processes at the sub-lexical phonological level during L2 acquisition, even though the researchers do not explicitly test it. They argue that

> from the beginning of the language acquisition process, a great deal of the phonology of the new language is available to the lexicon. Gradually, the inconsistencies in the letter-sound mappings between both languages are incorporated in the network (in the same way that inconsistencies within a language become incorporated), until, in the end the letter-sound mappings of the new language are not only mastered but also start to have an influence on the letter-sound conversions of the native language. (op.cit.:624)

If this depiction is a true reflection of a (developing) bilingual GPC system, then a study including bilingual participants with varying L2 proficiency may display differences in processing for these bilinguals. Accordingly, one of the central aims of the present study was to provide evidence for this account.

In addition, the present study employed a complementary group of bilinguals (i.e. German-English and English-German), who both completed the same task in their L1 as well as their L2. This was done to assess whether the obtained effects are related to the orthographic depth of a bilingual language pair, and/or to language dominance.

4.2 Experiment Three: Progressive Demasking in English-German Bilinguals

4.2.1 Methodology

4.2.1.1 Materials

During an experimental session participants had to identify target words in two blocks of trials; one block was completed in the participants' first language (English), and the other in their second language (German). On each trial, a word gradually appeared on a computer screen out of a background of visual noise. Participants pressed a response button as soon as they thought they identified the word. They then typed in the word they thought they had identified. The typed-in word provided a measure of accuracy for each participant.

Based on previous research using cognates as critical stimuli (cf. Dijkstra et al. 1999, 2010; Lemhöfer & Dijkstra 2004; Schwartz et al. 2007) and as an extension of that research, the selected critical stimulus material consisted of English-German cognates. The use of cognates for the current purposes may be superior to items employed in previous studies investigating the activation of phonology across languages (i.e. interlingual homographs or homophones). This is because by using cognates, the researcher can largely limit imbalances in the frequencies of a word form in the two languages (for instance, the German-English homograph *Made*, meaning "maggot", is far less frequent than its English correspondent *made*; the German-English cognate finger/Finger, however, is similarly frequent in both languages).

The set of cognates employed in this study was selected to provide contrasts in either or both of orthographic and phonological similarity across the two languages (i.e. being either more similar [+] or less similar [–]). This stimulus setup resulted in four experimental conditions: +O+P, +O-P, -O+P, and -O-P (see Table 5). For example, the English-German cognate pair *finger*/*Finger* would be categorised as belonging to the +O+P condition because of the high orthographic and phonological overlap of its members, whereas the cognate pair *book*/*Buch* ([bʊk] vs. [bu:x]) was categorised as -O-P due to the distinct orthographic and phonological codes of the cognate pair members (see below for examples in the remaining categories and a description of how these classifications were obtained).

This experimental design required careful target item selection and preparation, and involved the following stages: construction of a list of potential target words (English-German cognate pairs); selection of item pairs based on separate orthographic and

phonological similarity measurements; selection of matched control words for comparative analyses with critical stimuli. Each stage is described in more detail as follows.

Preliminary cognate pair composition

Initial item construction was done in the following way. First, an extensive list of German-English cognate pairs was created, mainly based on items provided by a researcher at the University of Wolverhampton (Mulloni 2007). A first selection criterion was to exclude any words that contained characters (e.g. German letters or diacritics such as *ä*, *ö*, *ü* or *ß*) which would give away which language they belonged to. Further, to ensure that lower proficiency learners of German would be familiar with included items, the existing selection was matched against an entry in the vocabulary list from two elementary German learners' course books (Albrecht et al. 2005; Perlmann-Balme & Kiefer 2004). Finally, as in previous research conducted within this experimental area (e.g. Dijkstra et al. 1999), only words that contained three, four, five or six letters were included in the experiment. Meeting these criteria resulted in 157 English-German cognate pairs.

Note that the definition of cognates varies slightly between researchers. In linguistics, for instance, cognates are defined as words that derive historically from the same ancestor language. In most psycholinguistic research, however, cognates may be "any two words with shared aspects of spelling, sound, and meaning across two languages" (Sunderman & Schwartz 2008:527). Given that the majority of psycholinguistic research employing cognates focused mainly on their form-meaning overlap, the latter definition may present a certain benefit (ibid.). For the purposes of my study, the selected items were believed to be etymological descendants (although not exclusively Germanic) which shared meanings across the two related languages (English and German), but varied in their degree of orthographic and/or phonological overlap (c.f. *finger*/*Finger; book*/*Buch*).

Note also that due to specific characteristics of the two languages, German words tend to be longer than their English translation equivalents. In terms of the current study, this meant that a number of cognate pair members differed in length, particularly when orthography varied across the two members (i.e. in the -O-P and -O+P conditions). Average word lengths and other relevant item characteristics are displayed in Table 5 below (see also Appendix D, for a full material list and item characteristics).

Measuring orthographic similarity

Following the approach adopted by Schwartz and colleagues (2007), the first aim was to obtain an objective measure of the orthographic similarity (OS) within each of the 157 cognate pairs using the algorithm described by van Orden (1987) and based on Weber's (1970) measure of graphic similarity (GS) for word pairs. This GS computation assigns different weights to (a) the number of pairs of adjacent letters shared by word pairs in order, (b) the number of pairs of adjacent letters shared by word pairs in reversed order, (c) the number of single letters shared, (d) the average number of letters in the two words, (e) the ratio of the word length of the shorter word to the longer, (f) whether the initial letter in the two words are the same, and finally (g) whether the final letter in a word pair are the same (see Appendix D for both formulae). As van Orden (1987) points out, one result of Weber's method is that words differ from one another in terms of how similar they are to themselves. For instance, the word *meet* is calculated to be more similar to itself (with a GS score of 1050) than the word *meat* is to itself (with a GS score of 975).

To accommodate his revised assumption that all words are *equally* similar in spelling to themselves, van Orden proposed the OS algorithm, which is based on the ratio of the GS of the members of a given word pair and the GS of one of the member words when compared to itself (1987:196).

The OS algorithm was originally proposed using monolingual (English) material only, but since my aim was to compute OS across two different languages, it was important to avoid letter inventory differences and diacritical marking (such as the previously mentioned *ä*, *ö*, *ü* or *ß*). When computing the OS score, all words, regardless of language, were entered in capital letters (also reflecting the presentation format of stimuli; see below). This was taken as a precaution against the influence of orthographic rules in German, i.e. the capitalisation of the initial letter in nouns. The following examples illustrate some characteristics resulting from the adopted measuring techniques. Independent of their word length, the calculated OS for both English-German cognate pairs *hand-Hand* and *finger-Finger* reached the maximum similarity score (1.00). When cognate pair members are not spelled exactly the same way across the two languages, the OS score decreases. For instance, the cognate pairs *friend-Freund*, *long-lang* and *half-halb* would typically receive increasingly lower scores. Two further word characteristics are particularly worth noting in these three cognate pair examples, which have an impact on their overall OS score: shared letters in the final position and word length.

Consider the cognate pair *half-halb,* which received the lowest OS score of the three (0.66). Primarily, this is a result of the fact that the members of this pair differ in their final letter. In addition, since the OS score is directly proportional to the length of cognate pair members, shorter words receive lower scores. As a result, the cognate pair *half-halb* receives the lowest OS score, being exceeded by the cognate pair *long-lang* (0.72; higher OS score due to the cognate members' shared final letter) and finally by the cognate pair *friend-Freund* (0.74). The latter would receive the highest OS score of the three not only due to the cognate pair words being longer but also since its members share their final letter. Finally, note that because the GS of words compared to themselves differ from word to word, as shown above for *meet* and *meat*, OS scores can differ depending on which word provides the denominator for the ratio calculation (so for instance the comparison wood/Wald returned an OS of 0.56 using the GS of wood/wood but 0.52 using the GS of Wald/Wald). Therefore the OS measure used in this research was the average of two OS values, using the GS of each word in the ratio calculation. Table 5 presents the average based on the means obtained from the two OS calculation measures (for OS scores of the individual cognate pairs, see Appendix D). The OS scores for the entire set of word pairs were fairly evenly distributed across a range from 0.05 to 1.00, and the average orthographic similarity for the first selection of 157 cognate pairs resulted in 0.60 (the score obtained for the median was the same, i.e. 0.60). Cognate pairs whose orthographic similarity was above this mark, i.e. 0.61 or greater, were classified as being more similar, and pairs with a similarity score of less than 0.60 were classified as being less similar. Note that because this categorisation was performed within the cognate word category, even pairs classified as relatively dissimilar for the purpose of the study were more similar across the two languages than non-cognate filler items and their translations.

Measuring phonological similarity

To obtain a measure of each cognate pair's phonological similarity, I first opted for a tool to provide me with an *objective* measurement. The only available tool to suit my purposes at the time was ALINE (Kondrak 2002). This on-line tool was developed for computing phonetic similarity between words. Although at first promising, this tool proved to be inadequate for my purposes, not least because it was set up for comparisons within rather than between languages, and so was unable to adequately take into account cross-linguistic phonetic and phonological differences.

I therefore reverted to a different method, successfully employed in previous research (Dijkstra et al. 1999, 2010; Schwartz et al. 2007; van Assche et al. 2011): the *subjective* phonological similarity rating. This method involved multiple preparation stages, each of which is described in the following. First, 21 filler items were added to the previous set of 157 cognate pairs, resulting in a total of 178 stimuli. The entire stimulus set was then recorded by two highly proficient bilingual German-English speakers (one male and one female), both currently living in New Zealand. The order of the recordings was randomised. One speaker read one member of a cognate pair, and the other speaker read the other member of a cognate pair. In total, each bilingual speaker read half of the cognate material in one language, and the other half in the other language. Four playlist versions were constructed using the recorded stimuli. Across these four playlists, half of the cognate pairs were spoken in English first, and the other half in German. In addition, half of the cognate pairs were spoken by the male speaker first, followed by the female speaker; and vice versa.

To obtain a measure of subjective phonological similarity of cognates, naïve listeners rated the perceived phonological similarity of the cognate pairs on a scale from 1 to 7, with 1 indicating that the cognate pair members were not similar at all and 7 indicating that they were highly similar (cf. Dijkstra et al. 2010; Schwartz et al. 2007). Two groups of raters (native speakers of German and of English) were used, and the four playlist versions were counterbalanced over these two groups. The English rater group consisted of twelve monolingual English speakers, and the German group of fourteen native German speakers with as little exposure to English as possible (note that it is close to impossible to find truly monolingual German speakers). To reduce inter-rater variability [$t(11) = 2.66$, $p = 0.02$], each listener's ratings were transformed into z-scores.[68] Within each of the two rater groups, cognate pairs whose phonological similarity was above the mean[69], i.e. with a value that is the average of the z-scores in the ratings of the 12 raters greater than zero, were classified as being more similar, and pairs below that mean were classified as being less similar. Again, this categorisation was performed within word pairs classified as cognates, so that even pairs classified as

[68] Prior to obtaining a normalised z-score, calculation of both the mean and the standard deviation of a listener's ratings is required. Consequently, the formula for the normalised rating score for each individual listener (n=24) is the following: (RAW SCORE - MEAN)/(STANDARD DEV). The data of two (German) raters was excluded due to missing ratings.

[69] There was no particular reason for choosing the mean over the median value given its close proximity. Deferred *t*-tests performed for each rater group confirmed that the two measures were not significantly different [English raters: $t(307) = 0.04$, $p = 0.97$; German raters: $t(307) = 0.35$, $p = 0.73$].

relatively dissimilar for the purpose of the study were more similar than non-cognate filler items and their translation. Subsequently, the two groups' ratings were compared and for the pool of items with high phonological similarity only items were chosen which were consistently rated as positive across the two groups (i.e. the average of the German raters' z-scores and the average of the English raters' z-scores both had to be positive). The same procedure was followed to designate items as negative, i.e. whose pronunciation was perceived as highly distinct. Out of the original 157 items, 128 items remained for further selection. Table 5 reports the average of the means obtained from both group ratings. For the complete set of materials, including the similarity measures for each cognate pair, see Appendix D. At this point, the remaining 128 items were matched against their OS calculations carried out in the previous step, to obtain four distinct cognate conditions. This was done as follows.

Categorising critical stimuli and selecting matched controls

The two similarity measurements described above were used to create four experimental conditions (see Table 5). An equal number of the most distinct cognate pairs was assigned to each condition (20 cognate pairs per condition; i.e. employing a total of 80 of the 128 cognate items remaining above). Cognate pairs across the different conditions varied with respect to their orthographic and/or phonological similarity, being either more (+), or less (–) similar across the two languages. For instance, the English-German cognate pair *finger*/*Finger* was placed in the +O+P condition, since the words' pronunciation and spelling were both highly similar across the two languages. The cognate pair *nation*/*Nation* was categorised as +O-P, given the identical spelling of its members across the two languages, but differing pronunciation. In contrast, sharing pronunciation but not spelling, the cognate pair *mouse*/*Maus* was placed in the -O+P condition. Finally, the cognate pair *book*/*Buch* was categorised as -O-P, since the pronunciation and the spelling of one of the pair members had only a weak resemblance with that of the other pair member. Appendix D contains a complete list of the 80 critical word pairs from all four categories (as well as the matched control material).

As a next step, the selected item pairs had to be matched with a corresponding number of control words. The CELEX database (Baayen et al. 1993) was used to access lexical information covering a word's number of letters, English or German log frequency per million (depending on whether an item was an English or a German word), English bigram frequency, German bigram frequency, number of German neighbours and the

number of English neighbours (see Table 5). Subsequently, 160 control words (e.g. *single*/*lieber*; *look*/*Luft*) were selected using the WordGen programme (Duyck, Desmet, et al. 2004), which uses the CELEX database (Baayen et al. 1993) as a resource. Characteristics of the set of control words matched those of the critical stimulus set as closely as possible.

Half of the control words were English words and the other half were German words, with one control word matching each member of a cognate pair (see Table 5 and Appendix D). Again, none of the German word items included language-specific letters or diacritics. Apart from matching lexical characteristics, the selected words also shared at least one letter with their critical stimulus equivalent (e.g. *talk*/*Lage* share the a in second position with the test pair *hand*/*Hand*; cf. Brysbaert et al. 1999; Duyck 2005; Haigh & Jared 2007).

Overall, the majority of the selected items (including both the cognate word pairs and the matched control words across the two languages) were nouns. However, some words belonged to a different syntactic category, such as an adjective, adverb or numeral, and in very few cases a verb. In addition, and due to experimental constraints, some items were category ambiguous, i.e. could be both noun and adjective (e.g. 'ARM' spelled all in capital letters could mean both 'poor' and 'arm' in German).

Finally, a further set of sixteen German and sixteen English words was selected, none of which appeared elsewhere in the experiment. This stimulus set was used in the practice session, which preceded the actual experimental trials, giving participants a chance to become familiar with the task. Given this purpose and the fact that the practice set was very small, I did not attempt to match it to any of the main characteristics of the main set, in terms of including cognates, or being within certain frequency ranges. One exception was the length of words, which was kept in accordance with the main set, i.e. within the range of 3-6 letters.

Table 5.

Examples of stimulus material by cognate condition and their matched controls, including various item characteristics and their counts. *Note*. ENG=English; GER=German; NB=Neighbourhood; Freq=frequency.

Target Items (by condition; *N* = 20)		**Lexical Properties**						**Similarity Ratings**	
Type	**ENG GER[a]**	**Letter Length**	**English NB Count[b]**	**German NB Count[b]**	**Log Freq[c]**	**English Bigram Freq[d]**	**German Bigram Freq[d]**	**Ortho-graphic**	**Phono-logical**
+ O +P	finger	4.65	6.95	3.60	2.17*	9925	13057*	.91	.88
	Finger	4.70	3.75	6.30	1.76	14311	10061***		
+ O + P control	single	4.65	6.80	2.35*	1.92*	8139*	9386*	N/A	
	lieber	4.70	3.55	2.85**	1.76	13239	6562***		
+ O – P	nation	4.55	6.95	3.40	2.30	8774	12853*	.73	-.47
	Nation	4.75	4.00	4.15	2.03	15048	8086		
+ O – P control	notice	4.55	6.85	1.55***	2.14	6991	7405*	N/A	
	Nutzen	4.75	4.15	3.10	1.86	15261	6650		
– O + P	shoe	4.15	8.75†††	2.05†††	1.97†	7412	7592	.40	.85
	Schuh	4.20	5.00	3.80	1.83	14737†††	6048†††		
– O + P control	shed	4.15	8.70	2.10	1.90	6519	8210	N/A	
	Stufe	4.20	5.00	3.80	1.67	13867	5975		
– O – P	book	3.85	1.90†††	10.50†††	2.36**†	6481	6102	.27	-.72
	Buch	4.05	3.80	4.55	2.14	4638†††	11436†††		
– O – P control	look	3.85	2.55	10.50	2.07**	6248	6032	N/A	
	Luft	4.05	3.65	4.45	2.08	4968	10648		

[a] Items in this column are examples of each category. Data in the other columns are based on the entire set of items in the category.
[b] Neighbourhood size (M. Coltheart et al. 1977), calculated using the WordGen program (Duyck, Desmet, et al. 2004) on the basis of the CELEX lexical database (Baayen et al. 1993).
[c] Mean log frequency per million words, according to the CELEX lemma database (Baayen et al. 1993).
[d] Mean summated bigram frequency of test and corresponding control targets, calculated using WordGen (Duyck, Desmet, et al. 2004)

Note. Deferred dependent samples *t*-tests were performed between stimulus material. The first tests compared cognates in each of the four conditions with their corresponding controls (e.g. +O+P with +O+P control). Separate tests were performed for language material in each of the two languages. Where *p*-values reached a significant level, the following significance codes are used: <.001 '***', <.01 '**', <.05 '*', as marked at a corresponding value of a lexical property.
Similarly, a second series of tests compared the values of lexical properties within cognate conditions (i.e. +O+P with +O-P; -O+P with -O-P). Again, material in the two languages was tested separately. Where *p*-values reached a significant level, the same denoting system was used as previously, but using the dagger symbol (†) instead of asterisks.
A third series of *t*-tests compared the log frequencies of cognate pairs within one particular condition (e.g. +O+P: *finger* with *Finger*; -O+P: *shoe* with *Schuh*). The only difference was found for the +O+P cognate pair, with p=.034.
Finally, the relatively small differences between similarity ratings across two critical conditions (e.g. the .91 O rating for +O+P and the .73 O rating for +O-P) were confirmed in deferred *t*-tests as minimally different. Ideally, there should not be any difference; however, the fact that there are differences reflects the general constraints on language material selection.

4.2.1.2 Participants

Forty-four native speakers of English completed an experimental session of approximately thirty minutes. Participants were students enrolled at the University of Auckland. They all had normal or corrected-to-normal vision. All participants had varying knowledge of German, and were selected to represent a range of L2 proficiency, based in most cases on their course enrolments. As a test of their individual German knowledge and in order to provide other relevant information, each participant filled out a language questionnaire and completed a German proficiency test[70] at the end of their experimental session (for detailed discussion of both assessment types, see section 3.2.1.4). All participants signed a written consent form. Participants received a gift voucher for their participation.

4.2.1.3 Procedure

The experimental procedure followed closely that of Dijkstra and colleagues (1999, 2010), described earlier. Participants were seated in a sound-proof booth, with dim lighting, approximately 60 cm away from a Philips 170B4 17-inch LCD Monitor, which was connected to a hidden personal laptop computer (Dell Latitude C640). They were first asked to read all instructions for the experiment on the computer screen. The instructions explained that the task was to identify words that would gradually appear on the computer screen out of a background of visual noise. Participants were instructed to react as soon as they identified the appearing word, but to avoid making errors (i.e. not to guess). After a short practice session, the actual test began. Depending on the condition (English or German), a trial started off by presenting either "NEXT WORD" or "NEUES WORT" on an otherwise empty screen. The participant then pressed a foot pedal to start the next trial. Two horizontal lines appeared exactly above and below the centre of the screen, and participants were instructed to focus on the gap between these lines. After 1500ms, the screen was cleared completely and a mask was presented at the centre of the screen. The mask consisted of a series of Xs, matching the length of the target word. During each trial, which consisted of multiple cycles, the mask first appeared for 300ms and was then followed by the target word, which was displayed for 15ms. The target word appeared in the same position as the mask. The first cycle was

[70] As explained in detail in Chapter 3 (section 3.2.1.4), the test used in both studies (and for both languages) was a *vocabulary* test in the form of a non-speeded lexical decision task. Although strictly speaking this test may not reflect participants' language proficiency, I will refer to it as a *proficiency* test for the sake of consistency, as well as in accordance with Lemhöfer et al. (2004).

immediately followed by a second cycle in which the mask was presented for 285ms and then the target for 30ms.

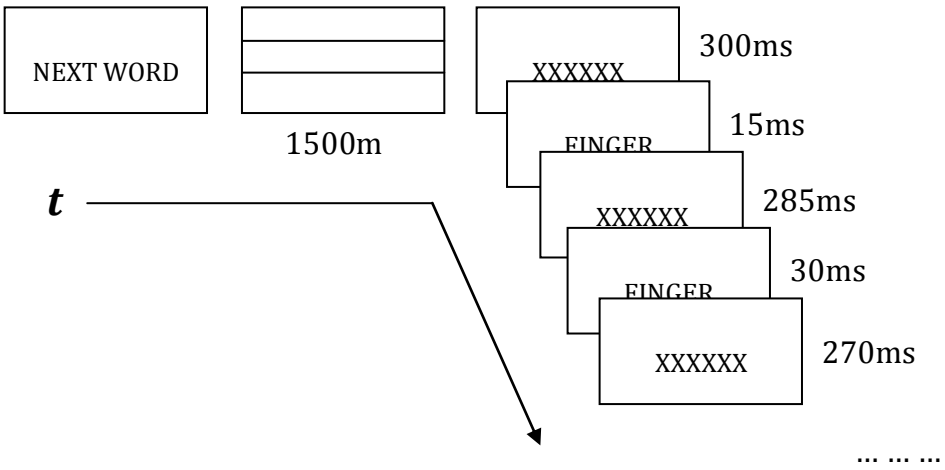


Figure 82. Schematic of an experimental trial in Study Two (Experiments Three and Four).

This pattern of cycles continued, with the mask duration decreasing by 15 ms on each cycle, and the target duration increasing by 15 ms. Thus, while the total display duration of the target and mask pair was always held constant, the ratio of the mask and target display duration progressively decreased (see Figure 82 for illustration). The participants' experience was of the target words emerging progressively from the mask. This progressive demasking procedure (cf. Dijkstra et al. 1999; Dufau, Stevens, & Grainger 2008; Grainger & Segui 1990; van Heuven et al. 1998) continued until the participant indicated that they had identified the word, by pressing any one of the five keys on a button box with millisecond timing accuracy. The response was timed-out after 6000ms (which equalled the total of a single trial duration) if the participant had made no response. Immediately after the participant had pressed a button (or after the time-out), a dialogue box appeared on screen prompting the participant (depending on the condition, either in English or German) to enter the identified word. After the participant had entered the word, the next trial started.

An experimental session was divided by a short break into two blocks of trials. Each block contained 80 trials, including both the critical stimuli (i.e. cognates) and matched controls. One block of trials was conducted in English and the other in German and included items in that language only, so there was no language switching involved within one particular block of trials. To control for a possible language effect, the order of the blocks (English-German or German-English) was counterbalanced across participants. Participants were allocated randomly but evenly to each block order.

Stimulus material was pseudo-randomised; however the order of stimuli across the blocks of trials was kept constant. This ensured that effects of sequential order within a block of trials (e.g. practice or fatigue effects) would be likely to affect each language condition equally. Finally, each experimental session was constructed in such a way that no participant would see a critical target word (i.e. cognate) in one language as well as its equivalent in the other language. For instance, if a participant saw the English cognate pair member 'cat' in one block of trials, they would never see its German equivalent 'Katze' anywhere in the experiment.

The experiment was run in E-Prime (Schneider et al. 2002a, 2002b) on a Windows personal laptop computer connected to an external monitor (for model specifications, see above). Masks and stimuli appeared all in upper case, black Courier New (18 point font size) at the centre of the computer screen on a white background. Presenting stimuli all in upper case ensured that participants would focus on the identification of a word and not on the appropriateness of its format (as noted earlier and in the previous study, all German nouns conventionally require their initial letter to be capitalised). The same response configurations were set up for all participants, and were independent of whether the participant was left- or right-handed. Between trials participants were asked to rest the index finger of their dominant hand on any chosen response button.

Participants were tested individually. To keep the entire experiment as stable as possible, the same researcher (the author) conducted all sessions, and the procedure was exactly the same for all participants (except for differences in block presentation described above). Communication between participant and researcher was done in the bilinguals' L1 (English), which should have reduced potential language expectations regarding the task. The progressive demasking task lasted no more than 30 minutes and was presented in two blocks, as described above. After the second block, subjects carried out a German proficiency test (Lemhöfer & Broersma 2011; Lemhöfer et al. 2004) and were asked to fill in a language history questionnaire (adapted from Li et al. 2006). Both the online proficiency test and the offline questionnaire were identical to those employed in Study One (see section 3.2.1.4).

4.2.2 Statistical analysis

The statistical analysis employed in this study was identical to that in Study One, i.e. mixed-effects modelling using the statistical platform R (Bates, Maechler, & Dai 2007; R Development Core Team 2006). The only difference lay in the selection of the primary predictors, which in the present study were Status (cognate versus control), Language (German versus English) and Proficiency (in the bilinguals' L2), as well as Orthographic Similarity and Phonological Similarity (for both factors, either Positive or Negative; as used by Schwartz et al. 2007). Each initial statistical model also included the same longitudinal predictors (i.e. Trial Number, Previous RT and Previous Accuracy) and secondary predictors employed in Study One (i.e. Frequency, Length, the number of German and English neighbours and Word Class), with the exception of Session, which was not relevant to this experiment since there was one experimental session only. Finally, all analyses reported in this study include participants and items as random effects, with intercepts varying by participant and item (cf. Study One). Random slopes were included where these provided a better statistical model. Where necessary, predictor variables were either centered or residualised to remove collinearity between predictors.

Prior to data analysis of Experiment Three (i.e. data collected in New Zealand, from English-German bilinguals), four participants were excluded (two participants had an L1 other than English; one participant reported being diagnosed with synaesthesia; and one participant was considerably older, i.e. more than twice the age, than the remaining bilinguals). Consequently, the initial data set comprised 6400 observations. Closer data inspection resulted in further exclusions: responses that were faster than 500ms were removed, since they would not capture the required cognitive processing (cf. Dijkstra et al. 2010). In addition, three control items were excluded since they induced an error rate higher than 20% (op.cit.), and another control was removed since it was mistakenly used in two different cognate conditions. The test items matching these controls were not excluded for two reasons: (1) The primary interest of this study was in the cognate responses and not in the replication of the (robust) cognate facilitation effect. Removal of such items could have put the emergence of an important effect at risk. (2) Previous studies with a similar research focus have limited their item exclusions to erroneous items only (e.g. Dijkstra et al. 2010; Schwartz et al. 2007); in fact, I am not aware that any of the reviewed literature excluded matching item pairs. The exclusions resulted in a

data loss of 2.2% (data loss due to item characteristics came to 1.6%), with the remaining data set comprising 6258 observations.

Next, where the dependent variable was RT, I inspected its distribution, which revealed a marked non-normality. Following a comparison of three types of RT transforms (log transform, inverse transform and inverse square root transform), the response durations were log transformed, since this transform was most successful in attenuating the observed non-normality.

Two types of analysis were performed on the given data set: an error rate and a response latency analysis. In the former case, i.e. where the dependent variable was binomially distributed, a *Generalised Linear Mixed Model* (henceforth, mixed logit) was fitted to the data. In this analysis, statistical significance of the fixed effects is assessed using Wald's Z-test (see Jaeger, 2008, for further detail and discussion of some problematic issues). Where the dependent variable was continually distributed, a minimally adequate *Linear Mixed-Effects Model* was fitted to the data containing only correct responses and subjected to model criticism. Potentially harmful outliers were removed and the model was refitted. The measure of statistical significance of the fixed effects in the final model was based on Markov Chain Monte Carlo (MCMC) sampling (10,000 iterations; Baayen et al. 2008).

For ease of reading and interpretation, the following results are presented and discussed based on plots of lmer model estimates (back-transformed for RTs) and standard errors in the relevant conditions.

4.2.3 Results and discussion

The primary aim of this experiment was to assess the competitive activation dynamics between orthographic and phonological representations of cognates across the two languages of L1 English-L2 German bilingual speakers, in an experimental environment that does *not* explicitly require the activation of a phonological representation. This question was addressed in light of the proficiency level of the bilinguals, as well as in the context of a particular language of the task (either L1 or L2).

My main predictions are as follows: given that the cognate facilitation effect has been successfully replicated in the visual modality using various experimental paradigms and increasingly elaborate measuring techniques (see, for instance, van Assche et al. 2011;

Dijkstra et al. 2010, for a most recent overview of previous accounts), I expect faster responses to cognates than to their matched controls.
In addition, given that the effect has been reported in both the participants' L2 (e.g. van Assche et al. 2011; Dijkstra et al. 1999, 2010; Lemhöfer et al. 2008; Lemhöfer & Dijkstra 2004) and in the L1 (e.g. van Assche et al. 2009; Midgley, Holcomb, & Grainger 2010; Titone et al. 2011), cognate facilitation may be expected to show in the present study irrespective of the language of processing.

As exemplified by many of the above researchers, cognate effects can be modulated by various factors. One such factor is the degree of cross-linguistic orthographic and phonological consistency. While some recent studies have acknowledged this fact by taking this factor into consideration, the present study attempts to shed further light on the precise nature of such dynamics, i.e. by exploring the (adapted) bidirectional consistency effect in bilingual visual word recognition.
As outlined in the introduction, the activation of cross-linguistic phonological representations - referred to by Schwartz and colleagues (2007) as the *feedforward effect* - has been already identified in a number of studies with varying research focus. Accordingly, we should find clear evidence for the feedforward effect despite the fact that phonology is not explicitly required in the present task. This should be reflected in the responses to cognates which are orthographically highly similar across the bilinguals' two languages, with slower identification of items which map on to two, highly distinct phonological representations (e.g. *nation*: ['neɪʃ(ə)n] vs. [na'tsjo:n]). In addition, given that these effects have been more consistently found in the L2, the anticipated competitive dynamics at the phonological level may not be manifest to the same extent when participants identify cognates in their L1.
Perhaps a more intriguing question, however, is whether we can find evidence for the *feedback effect* in the visual modality. If the occurrence of such an effect is subject to the orthographic depth of the language of processing, and consequently the corresponding GPC rules applied by the bilingual reader, then we might expect similar results to those observed by Schwartz et al. (op.cit.). That is, when participants are processing in their (deep) L1, cognates which have highly distinct orthographic realisations across languages should not be identified differently from each other, irrespective of their degree of phonological similarity. This is because the "impoverished phonological code" (Frost 2005:285) being activated by the input may not provide sufficient information

back to the orthographic level for the activation of a competing orthographic representation from the non-target language to arise. In contrast, if different GPC rules are employed in more shallow languages, and if these rules are available to a bilingual speaker reading in a(n) (shallow) L2, then a "relatively detailed phonological representation" (ibid.) activated by the input may not only be better able to send information back to the orthographic level, but also to spread activation across languages. What this means in context of the present study is that we may indeed observe response delays to cognates with highly distinct orthographic representations which share a similar phonological code across languages (e.g. [maʊs] = *mouse*/*Maus*); however only when participants are identifying cognates in their (shallow) L2.
Finally, the feedback effect may emerge as a consequence of increasing proficiency. Given that no previous work has explicitly addressed this question to date, it is unclear what effect proficiency may have on the GPC system of a bilingual speaker, or its development. It may be that two separate GPC systems are acquired, one for each language, and that they are available accordingly. It may also be, however, that a new GPC system is formed; one that can contend with the requirements of either language (cf. van Wijnendaele & Brysbaert 2002). If the former is a truer reflection of the bilingual processing system (i.e. two separate GPC systems), then the predicted feedback effect should be limited to processing in the L2, being more pronounced for more proficient speakers. The latter proposition (a joined GPC system), however, may affect both processing in the L1 and in the L2; hence, apart from less pronounced effects for less proficient bilingual readers, similar effects should be revealed irrespective of the language of processing.

Given these predictions, analyses were first conducted on the word set as a whole (i.e. including both the English-German cognates and control words in the respective language), and were subsequently followed by a subset of data, which included only the critical material, i.e. cognates. The mean response latencies on correct trials and percent error rates are summarised in Table 6. The calculation of the mean values presented in this table followed the initial data trimming described in section 4.2.2 above, but preceded the statistical modelling reported below. Hence, they give a general pattern of response latencies and error rates only.

Table 6.

Mean recognition latencies for correct responses (in milliseconds) and percent error rates[71] (in parentheses) for cognates and controls in L1 (English) and L2 (German).

Status	English (L1)	German (L2)
cognates	1866 (0.88%)	2216 (1.32%)
(non-cognate) controls	1967 (0.70%)	2425 (1.46%)

4.2.3.1 Responses to cognates and control words

To determine whether the present study was able to replicate the cognate effect (see Table 6) and what factors played a role during the identification process, the first analysis conducted in this experiment compared responses obtained for cognates with those to matched control words. Accordingly, separate statistical models (i.e. linear mixed-effects and mixed logit; see discussion above) were fitted to response time and error data, including Status (cognate vs. control), Language (L1 vs. L2) and Proficiency (in L2) as the main predictors and a number of secondary predictors, outlined in section 4.2.2. Participants and Items were included in the model as random effects. Prior to analysis, the factor *length* was residualised against *frequency*, to account for the relationship between word frequency (as measured using the CELEX database; Baayen et al., 1993) and word length (in letters). Where necessary, factors were centred to avoid collinearity. Where variables did not reach significance as predictors (i.e. their regression weights were not significantly different from zero) and were not involved in any interactions, they were dropped from the model. Significant effects revealed in the final model fitted to participants' latency data are reported first, followed by an account of the significant effects following a model comparison fitted to the error data.

In line with expectations, the analysis of response latencies (see Table E9, Appendix D) revealed Status as a significant simple effect (MCMCmean = 0.056, HPD95lower = 0.028, HPD95upper = 0.083, $p < 0.001$, Figure 83), with overall faster responses to cognates (Note that the values displayed in the figure(s) are predicted values from the model, with other factors set at their intercept value. This is why they differ from the values in Table 6 above). This effect was further defined by an interaction with Proficiency (MCMCmean = -0.059, HPD95lower = -0.087, HPD95upper = -0.029, $p < 0.001$, Figure

[71] Given that the nature of the task does not permit many errors (i.e. requiring a full identification of the gradually appearing word), the overall error rate is very low. The few instances where a response has been classified as an error reflect this fact, including merely letter substitutions, deletions or incorrect letter insertions. None of these errors are believed to have been a result of a typing mistake, since participants were explicitly informed that they can take as much time as they need for typing.

84), showing that cognate facilitation is particularly effective for lower proficiency bilingual speakers, and subsides with increasing proficiency in an L2. A very similar finding has recently been observed in an eye-tracking study (Libben & Titone 2009).

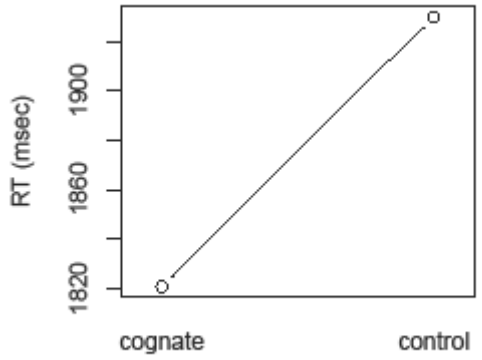


Figure 83. Mean response times for cognates and controls.

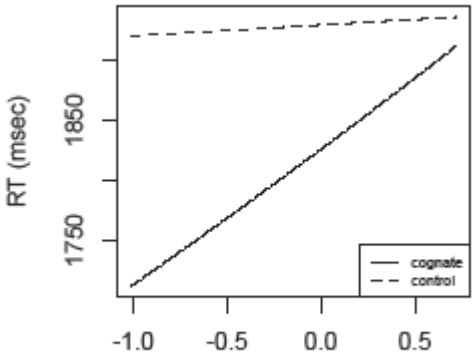


Figure 84. Mean response times for cognates and controls, with increasing L2 proficiency.

Responses to cognates are also affected by the frequency of the stimulus (MCMCmean = -0.053, HPD95lower = -0.095, HPD95upper = -0.011, p = 0.015, Figure 85). That is, while cognates are processed faster than their matched controls when both item types are of low frequency, this difference seems to recede with increasing (log) frequency of the items. A very similar finding was recently reported by Dijkstra and colleagues (2010:296), who interpreted this observation as a cumulative facilitation effect for low-frequency cognates. That is, low-frequency cognates seem to benefit more from the co-activation of their cross-linguistic counterparts. Given that this interaction was not further defined by Language, the present results extend the previous findings by suggesting that this phenomenon is irrespective of the language of processing, i.e. either L1 (English) or L2 (German).

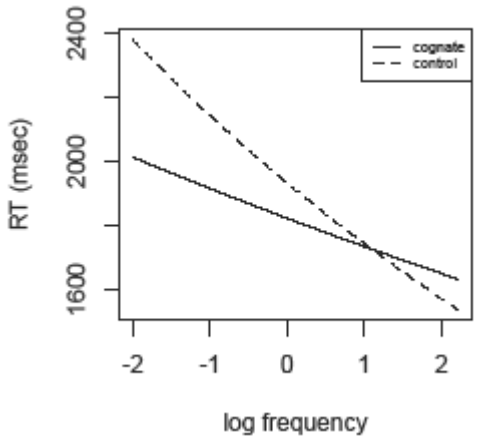


Figure 85. Mean response times for cognates and controls, in relation to the items' frequency.

Finally, the latency analysis also revealed that responses to both item types are similarly affected by the language of the task, where processing is faster in the L1 (MCMCmean = 0.112, HPD95lower = 0.088, HPD95upper = 0.137, $p < 0.001$, Figure 86). In addition, Language enters an interaction with Proficiency (MCMCmean = -0.137, HPD95lower = -0.167, HPD95upper = -0.108, $p < 0.001$, Figure 87), which reflects the fact that lower proficiency speakers perform better (i.e. identify words faster) when the task is conducted in their L1 as opposed to the L2. This may be partly because processing in the L1 is more automatic than in the L2 (cf. Tzeglov & Kadosh 2009), but it may also reflect the fact that their L2 vocabulary knowledge is less extensive or that the lexical links in the L2 are weaker than in the L1 (cf. Schwartz et al. 2007).

Another interesting aspect of this interaction is that the lower proficiency speakers are faster in their identification of L1 words than their higher L2 proficiency counterparts. A possible explanation for this finding concerns the relative vocabulary size of the participants, together with the assumption that language access is largely non-selective. That is, taken that a consequence of an increasing L2 proficiency level is an increase in the bilinguals' L2 vocabulary size, we might expect higher global activation levels for more proficient bilingual readers, i.e. activating a larger amount of (neighbouring) lexical representations from the non-target language. It follows that higher global activation increases the amount of cross-linguistic competition. Hence, the time required to resolve this competition may be what causes the response delay for more proficient bilinguals; even though identification is done in their L1.

Figure 86. Mean response times for items, depending on the task language (L1 or L2).

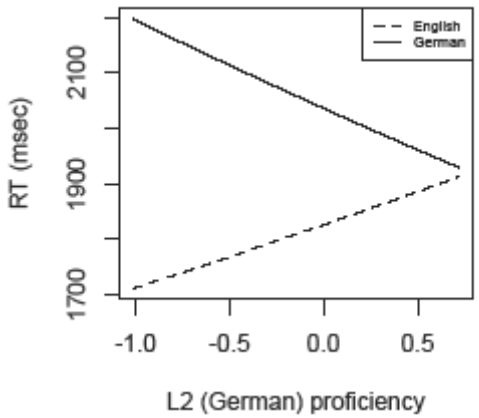


Figure 87. Mean response times for items by task language and with increasing L2 proficiency.

Among the secondary predictors, the only significant effect was found for Frequency (MCMCmean = -0.052, HPD95lower = -0.087, HPD95upper = -0.019, $p < 0.001$), with

faster responses with increasing log frequency of an item. This finding is in line with some recent relevant work (e.g. Dijkstra et al. 2010; Ziegler et al. 2008).

As can be seen in Table 6, very few errors were made in the present task. It may to some extent be for this reason that the analysis of the accuracy data revealed only one (marginally significant) simple effect, namely Language (β = -0.642, SE = 0.334, z = -1.920, p = 0.055; see Table E10, Appendix D), with superior identification in the L1. The absence of other simple effects or interactions may seem surprising at first, however, the nature of the experimental task may largely account for the present finding. That is, the task demands in progressive demasking require participants to uniquely identify a stimulus. One might therefore expect that any variables having an impact on the identification of items will predominantly, if not exclusively, emerge in the response latencies, rather than in the error rates. A similarly reduced number of effects has been reported in previous work using this paradigm (e.g. Dijkstra et al. 2010, experiment 3; Grainger & Segui 1990, experiment 2; van Heuven et al. 1998, experiment 1), although none of the researchers have commented on the difference of the obtained effects. Instead, some investigators have avoided performing a separate error analysis altogether (cf. Lemhöfer et al. 2008).

4.2.3.2 Cognate conditions

To investigate in more detail the competitive activation dynamics between orthographic and phonological representations of cognates across languages, an analysis of only the cognate items (i.e. excluding the controls) was performed. Unlike in the analysis above, the statistical models fitted to the data included Orthographic Similarity and Phonological Similarity as two of the main predictors.[72] The remaining main predictors were Language and Proficiency. The statistical models also included the secondary predictors employed in the previous analysis, and Participants and Items as random effects. Results from the latency analysis are outlined first, followed by the results of the error analysis.

[72] The reader will notice that these predictors were not included in the full data set analysis. This is because the primary focus of this study was to compare the sub-lexical dynamics for items that shared the same meanings across languages, but differed to some extent in the degree of phonological and orthographic overlap across representations (cf. Schwartz et al. 2007; but see Dijkstra et. 2010, for a slightly different research focus). I did therefore not obtain the respective ratings for the control items.

Table 7.

Mean recognition latencies for correct responses (in milliseconds) and percent error rates[73] (in parentheses) for the four cognate conditions in L1 (English) and L2 (German).

Cognate condition	English (L1)	German (L2)
+O+P	1863 (1.5%)	2168 (1.26%)
+O-P	1897 (1%)	2246 (2.02%)
-O+P	1825 (0.75%)	2268 (0.75%)
-O-P	1879 (0.25%)	2183 (1.25%)

The analysis of response latencies revealed Language as the only significant simple effect among the primary variables of interest (MCMCmean = 0.089, HPD95lower = 0.017, HPD95upper = 0.160, p = 0.012; see Table E11, Appendix D), reflecting faster responses when identifying cognates in English (L1) as opposed to in German (L2). More central to the primary interest of this study, however, was the emergence of a three-way interaction between Language, and the Orthographic Similarity and Phonological Similarity of the cognates (MCMCmean = -0.128, HPD95lower = -0.252, HPD95upper = -0.002, p = 0.043, Figure 88).

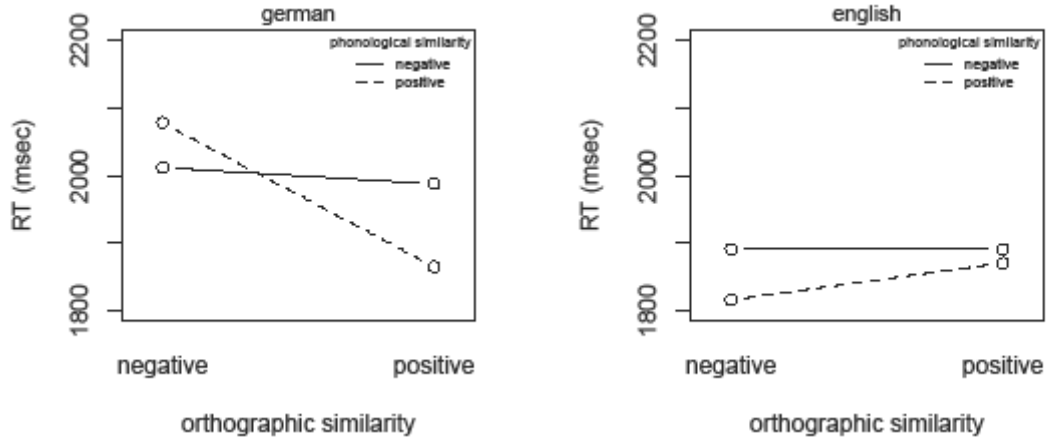


Figure 88. Mean response times for the four O-P cognate conditions, by orthographic similarity (positive vs. negative), phonological similarity (positive vs. negative) and task language (L2 German, in the left panel, and L1 English, in the right panel).[74]

There are several implications for this finding. First, overall faster responses in the English task may be indicative of the more automatic nature of processing in the bilinguals' L1 and hence faster identification (see discussion above for alternative causes leading to faster processing in an L1). Second and more central to the subject of the

[73] Given that the nature of the task does not permit many errors (i.e. requiring a full identification of the gradually appearing word), the overall error rate is very low. The few instances where a response has been classified as an error reflect this fact, including merely letter substitutions, deletions or incorrect letter insertions. None of these errors are believed to have been a result of a typing mistake, since participants were explicitly informed that they can take as much time as they need for typing.

[74] For a clearer visual comparison with Schwartz et al.'s (2007) results, see Figure 109 in Appendix D.

study, is the distinct identification speed of the four cognate types in the L2. The pattern of responses that emerges from the analysis is that cognates with a high degree of orthographic overlap are identified more slowly when their corresponding phonological representations are more distinct across languages (+O-P, e.g. *nation*: ['neɪʃ(ə)n] vs. [na'tsjo:n]). This is in line with previous research and provides evidence for the so-called *feedforward effect*, i.e. an orthographic representation activates the phonological codes from both a bilingual's languages, and it is the competition of these two phonological codes at this level that is assumed to slow down their identification (Schwartz et al. 2007:116–118). Another pattern of responses which is consistent with Schwartz and colleagues' work is found in the responses to cognates with quite distinct orthographic representations across languages (e.g. *mouse*/*Maus*, *book*/*Buch*). Although responses are somewhat slowed down when the orthographic representations map on to a highly similar phonological code (-O+P, e.g. [maʊs]), there is a large degree of overlap of such response times with those obtained when there are larger differences in the phonological match across languages (-O-P, e.g. [bʊk] vs. [bu:x]). Hence, the evidence for a *feedback* activation mechanism is much weaker, if it can be sustained at all.

To shed more light on this issue, I performed additional analyses on the L2 data only, for the +O/-O distinction evident in the left panel of Figure 88, i.e. by separating the data into orthographically highly similar (+O) as opposed to highly dissimilar (-O) cognate material. The analysis of the highly reduced data set revealed a trend in the direction of my conjecture. That is, the statistical analysis including only cognates with highly similar orthographic representations across languages returned a trend towards significance for the interaction between Phonological Similarity and Language (MCMCmean = -0.062, HPD95lower = -0.139, HPD95upper = 0.014, p = 0.095; Figure 89). In contrast, the analysis including only the -O cognate material did not reveal such effects.[75]

[75] Note that both analyses also returned a significant interaction between Language and Proficiency, as well as a simple effect of Language and Frequency. While they need to be mentioned for the sake of completion, they are not of primary interest in the current discussion.

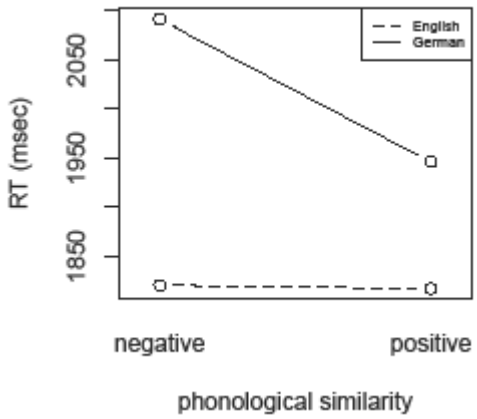


Figure 89. Mean response times for the +O cognate material, by phonological similarity (positive vs. negative) and task language (German L2 vs. English L1).

While the lack of a conventionally significant effect may be explained by the fact that the data set for the subsequent analyses was considerably reduced and may have therefore lacked sufficient observations, the analyses help further substantiate the pattern of observations obtained in the L1. Namely, in contrast to word identification in the L2, there is no real evidence that the activation dynamics between the orthographic and phonological representations across languages are of an interactive, not to mention of a competitive nature, when processing is in the L1.[76] Although this finding does not come as a surprise – assuming more automatic (hence, faster) processing in the L1, which may leave less time for cross-linguistic effects to emerge – it is in contrast to Schwartz et al.'s (ibid.) findings who once again found evidence for feedforward activation. A possible explanation of the contrasting results is in the experimental task employed by the researchers, i.e. a naming task, which explicitly requires access to the phonological representation of items and therefore provides an enhanced opportunity to observe cross-linguistic interference. This explanation seems in line with more recent research which failed to obtain cross-linguistic interference effects in other visual tasks conducted in the bilinguals' L1 (e.g. Haigh & Jared 2007; Jared & Szucs 2002; Midgley et al. 2010; see relevant discussion below). Consequently, the findings of the present experiment emphasise the necessity of varying tasks to better understand the complexity of bilingual processing.

[76] Note that there was also no simple effect of Phonological Similarity when another separate analysis was conducted, including the L1 data only. Hence, this is further evidence that what appears to be an effect in the opposite direction of the predicted feedback effect in Figure 88 is not actually statistically confirmed.

The only other interaction which emerged as significant in the latency analysis was that between Proficiency and Language (MCMCmean = -0.107, HPD95lower = -0.148, HPD95upper = -0.067, $p < 0.001$, Figure 90). As seen in the previous analysis, this interaction is mainly carried by the contrasting speed of responses for lower proficiency bilinguals in their respective languages, i.e. faster in L1 and slower in L2. We also see again that low proficiency L2 bilinguals are faster in identifying words in their L1, compared to their more proficient counterparts (see discussion above).

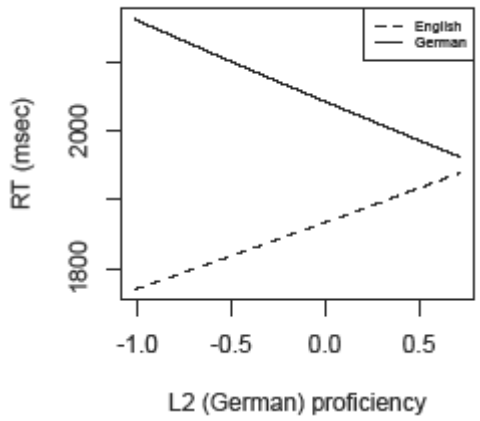


Figure 90. Mean response times for cognates in relation to the language of the task and increasing L2 proficiency.

The last effect to be reported in the analysis of response latencies was found among the secondary predictors, namely Frequency (MCMCmean = -0.066, HPD95lower = -0.101, HPD95upper = -0.035, $p < 0.001$). In this context, more frequent cognates were responded to faster than their low-frequency counterparts (cf. Dijkstra et al. 2010). Unlike in the full data set (see section 4.2.3.1 above), Frequency did not interact with any of the primary predictors in this analysis.

Similar to the full data set, a different result pattern was obtained for the error analysis (see Table E12, Appendix D). That is, only two variables were identified as showing a tendency towards significance: Language ($\beta = -0.923$, SE = 0.546, $z = -1.689$, $p = 0.091$) and the English neighbourhood density ($\beta = -0.085$, SE = 0.049, $z = -1.749$, $p = 0.080$). There were no significant interactions of variables.

4.2.3.3 Discussion

At a more general level, the present experiment sought to replicate the cognate facilitation effect, addressing a number of factors that may have an impact on its occurrence. The results of the statistical analysis identify two such factors, namely the frequency of the items and the bilinguals' L2 proficiency. With respect to the former of the two, i.e. the frequency of the cognates, it does not come as a surprise that it is the low frequency cognates that show a processing advantage over the matched control items. As argued by Dijkstra and colleagues, facilitation occurs for these items only because the identification process benefits a lot more from the co-activation of their (translation) equivalents in the non-target language than either in case of the high-frequency cognates, or the low-frequency items with no such closely corresponding translation entries (2010:296).

While L2 proficiency has also been recently identified as an influential factor (e.g. Libben & Titone 2009; Midgley, Holcomb, & Grainger 2009), it may seem less obvious why cognate facilitation would have a larger effect on less proficient bilinguals. If we construe this finding in accordance with Libben and Titone's interpretation, then we might argue that the occurrence of a facilitatory effect is a result of language non-selectivity, being reduced with increasing L2 proficiency of a bilingual speaker (2009:388). This interpretation may appear counterintuitive, since language selectivity is generally associated with faster (not slower) responses; hence, it largely contradicts the current findings. Alternatively, the reverse could be argued assuming that increasing proficiency in an L2 yields generally higher global activation levels and therefore more cross-linguistic competition (and possibly interference) from items in a cognate's immediate neighbourhood. A facilitation effect would then be more likely to be observed for less proficient bilinguals, as demonstrated by the current data. Finally, the present findings could also be taken as evidence that during their first encounters cognates may still be stored as two separate entries with rather weak (if any) connections between the two languages. This interpretation would imply a developmental pattern of bilingual lexical organisation, with the two languages becoming increasingly combined.

Previous research has shown not only that the cognate facilitation effect is modulated by lexical or external factors (such as the two above), but also that the recognition of cognates can be similarly affected by more subtle, sub-lexical dynamics (e.g. van Assche

et al. 2011; Duyck et al. 2007; Schwartz et al. 2007). The investigation of the precise nature of such dynamics during silent reading was the central aim of the current study. As illustrated above, the present findings confirm prior results (Schwartz et al. 2007) in providing compelling evidence for feedforward activation from orthography to phonology across languages. That is, identification of cognates in the +O+P condition (e.g. *finger*) took *less* time when compared to cognates from the +O-P condition (e.g. *nation*). In line with my predictions, however, a clear facilitation effect occurred only when participants were performing the task in their L2. As argued by Haigh and Jared (2007), a possible explanation of this finding is that when bilinguals read in their first language, the activation of the non-target (L2) language may not be strong enough to affect lexical processing; hence performing the task in the L1 did not reveal clear processing differences (see also Midgley et al. 2010, for a similar discussion). Furthermore, the current results differ from earlier research in showing that bilingual visual processing may be dependent on the orthographic depth of the language of processing, and consequently the corresponding GPC rules applied by the bilingual reader (cf. Lemhöfer et al. 2008). This is demonstrated in the contrasting response pattern for bilinguals identifying cognates in their L1 as opposed to the L2, particularly in relation to the feedback mechanism. That is, when participants performed the task in their L2, a trend in the data appeared indicating that cognate identification in the -O+P condition (e.g. *mouse*/*Maus*) was delayed when compared to cognates in the -O-P condition (e.g. *book*/*Buch*). Although this trend is indicative of feedback activation from a phonological representation to the orthographic level (i.e. co-activating a non-target orthographic representation, which leads to competitive dynamics at the orthographic level), the statistical analysis was not fully conclusive (for a similar result, see Schwartz et al. 2007). In contrast, performance in the L1 yielded no such effects. The absence of a feedback effect in the L1 could be explained under the assumption that processing in a deep language (L1) involves different activation patterns compared to a more shallow language (L2). That is, if skilled readers in deep orthographies construct merely "underspecified phonological representations" (Frost 2005:285) in the process of accessing their mental lexicon, then although there may be some information flow back to the orthographic level (cf. Stone et al. 1997), it may impede the activation to spread across languages. Reading in a more shallow language, on the other hand, may entail the application of different GPC rules and hence the construction of a more defined phonological representation. Accordingly, the activated phonological code may not only

be better able to send information back to the orthographic level, but also to encourage cross-linguistic activation. The lack of clear effects could be explained by the fact that the participating group of bilinguals may not have been proficient enough, in order to effectively employ the L2 GPC rules.

Finally, the reported results go beyond prior studies by showing that the sub-lexical interactive dynamics between the orthographic and phonological representations seem to take place largely irrespective of the L2 proficiency of a bilingual speaker, and that they can be observed despite phonology not being explicitly required. Admittedly, the latter may be highly dependent on the task demands, as suggested recently by Ziegler and colleagues (2008).

To shed more light on the discussion above, an identical experiment was conducted with a complementary, i.e. German L1–English L2 bilingual population. More specifically, given that orthographic depth differed from L1 to L2, it is not clear whether the obtained findings reflected an 'orthographic depth effect' or simply differences between L1 and L2 processing. Clarifying this issue was the main motivation for Experiment Four. The following sections discuss any necessary changes to the experimental setup, as well as the obtained results and implications.

4.3 Experiment Four: Progressive Demasking in German-English Bilinguals

4.3.1 Methodology

4.3.1.1 Participants

Forty-four native speakers of German completed an experimental session of approximately thirty minutes. Recruited participants were students enrolled at Albert-Ludwigs University of Freiburg. They all had normal or corrected-to-normal vision and no hearing impairment. Participants had varying knowledge of English, and were selected to represent a range of L2 proficiency, based on their course enrolments. As a test of their individual English knowledge and in order to provide other relevant information, each participant completed an English proficiency test at the end of their experimental session and was asked to fill out a language questionnaire (for detailed discussion of both assessment types, see section 3.2.1.4). All participants signed a written consent form. In recognition of their participation, participants went into a draw for one of three major prizes.

4.3.1.2 Materials and Procedure

The critical stimulus material consisted of the same eighty German-English cognate pairs as those used in Experiment Three, distributed across four experimental conditions (i.e. +O+P, +O-P, -O+P, and -O-P), as described in section 4.2.1.1. Their corresponding matched control items were also taken from the previous experiment (see section 4.2.1.1). All items are listed in Appendix D.

Participants were seated in a spacious office, with standard lighting, approximately 60 cm away from a standard desktop monitor (Dell E190S 19-inch Flat Panel LCD Monitor; Dell OptiPlex 755 Desktop, Intel® vProTM processor technology). The procedure, instructions, and presentation orders were identical to those of Experiment Three and the same native German-speaking researcher conducted all sessions. Unlike in the previous experiment, however, participants communicated with the researcher exclusively in German (i.e. in their L1). This was done to reduce potential language expectations regarding the task.

Following the experimental session, participants carried out the English version of the proficiency test and were further asked to fill in a language history questionnaire (adapted from Li et al. 2006), which was sent to them by email. They were given two weeks to return the questionnaire in the same (or a preferred) way. The latter was done

to ensure that the duration of an experimental session did not exceed the indicated time. Only one questionnaire was never returned to the researcher.

4.3.2 Statistical analysis

The statistical approach (including the analysis type and the chosen predictors) employed for this data set is identical to the statistical analysis for Experiment Three (for a detailed description, see section 4.2.2 above).

The initial data set comprised 7040 observations. Prior to data analysis, the data of one participant had to be excluded since she did not pay attention to the task by repeatedly looking elsewhere around the experimental room instead of focussing on the computer screen.[77] This resulted in the loss of 160 data points. Further data exclusions were as follows: first, responses that were faster than 500ms were removed, since they would not capture the required cognitive processing (cf. Dijkstra et al. 2010). Second, seven items were excluded since they resulted in an error rate higher than 20% (op.cit.); the items included three cognates and four controls. Finally, one further control item was removed since it was mistakenly used in two different cognate conditions. For the same reasons as outlined in the previous experiment (i.e. Experiment Three), these items' matched counterparts were not excluded from further analysis. The exclusions resulted in a data loss of 2.9% (data loss due to item characteristics came to 2.8%), with the remaining data set comprising 6680 observations.

For ease of reading and interpretation, the following results are presented and discussed based on plots of lmer model estimates (back-transformed for RTs) and standard errors in the relevant conditions.

4.3.3 Results and discussion

The aim of this experiment was to find out whether the competitive activation dynamics between orthographic and phonological representations of cognates reported in the previous experiment could also be observed for bilingual speakers with a reverse language background (i.e. German as L1 and English as L2). As mentioned above, the main motivation was to explore whether the obtained effects can be ascribed to the languages differences in orthographic depth, or whether they reflect more generally

[77] The data from the participant who did not return the emailed questionnaire was not removed, because the questionnaire was not my primary source of information on proficiency.

assumed differences between L1/L2 processing (cf. Lemhöfer et al. 2008; Midgley et al. 2009). In line with the previous experiment, this question was addressed in light of the proficiency level of the bilingual, as well as in the context of a particular task language (either L1 or L2).

The predictions for this experiment were largely in line with the predictions made in the previous experiment (i.e. Experiment Three). That is, we should yet again obtain a clear cognate facilitation effect, irrespective of the language of processing.
As shown by previous and the current research, the magnitude of the effect can be modulated by both internal and external factors. In case of the latter (i.e. external factors), we may find that the cognate effect changes with increasing L2 proficiency of a bilingual. Similarly, the stimulus' frequency (i.e. an internal factor) may have an impact on the cognitive processing of cognates and their identification speed.
For the remaining (internal) variables of interest, i.e. cross-linguistic phonological and orthographic similarity, the following predictions can be made in relation to the (adapted) bidirectional consistency effect: my expectation is that *feedforward effects* should be once again obtained in the present task, despite the fact that phonology is not explicitly required. Accordingly, responses to cognates that are orthographically highly similar across the bilinguals' two languages are expected to be delayed when the items map on to two, highly distinct phonological representations (e.g. *nation*). In addition, such effects may be stronger when processing in the L2 (see previous experiment, for arguments behind this reasoning).
Evidence for *feedback effects* in the visual modality, on the other hand, may be partly dependent on the orthographic depth of the language of processing, or alternatively of the GPC system utilised by the German-English bilingual speaker. As discussed above, this conjecture is mainly based on the response trend obtained in the (shallow) L2, in Experiment Three. If this trend was a true reflection of the processes involved when reading in a more shallow language then the present experiment, unlike Experiment Three, should provide clear evidence for feedback activation in the bilinguals' L1. This should be reflected in the responses to cognates which have two distinct orthographic realisations across languages: items that are phonologically highly similar (e.g. *mouse*/*Maus*) should instigate larger cross-linguistic competition, leading to slower responses when compared to items whose phonological representations vary to a greater extent across languages (e.g. *book*/*Buch*).

In contrast, processing in the (deeper) L2 may not reveal such effects. Although English-German bilinguals reading in their (deep) L1 showed no feedback effects (see Experiment Three), the presence or absence of these effects may be conditional on the kind of GPC rules that are available to and applied by a bilingual reader with a reverse language order, when reading in their L2. Consequently, the lack of feedback effects will be indicative of the use of L2 GPC rules by German-English bilinguals. Finding evidence for feedback effects, however, will imply that similar (L1) GPC rules are applied, regardless of the language of processing.
Lastly, given that the previously reported bidirectional consistency effects were found irrespective of the bilinguals' L2 proficiency, I do not expect the feedforward or feedback effects to be modulated by the proficiency of the present bilingual group.

Analyses were again first conducted on the data set as a whole, i.e. including both the German-English cognates and control words in the respective language. Subsequent analyses were performed on a subset of data, including only the critical cognate material. Table 7 below summarises mean response latencies on correct trials and the respective error rates. The calculation of the mean values presented in this table followed the initial data trimming described in section 4.2.2 above, but preceded the statistical modelling reported below. Hence, they give a general pattern of response latencies and error rates only.

Table 8.
Mean recognition latencies for correct responses (in milliseconds) and percent error rates (in parentheses) for cognates and controls in L1 (German) and L2 (English).

Item Type	German (L1)	English (L2)
cognates	1882 (1.43%)	1916 (1.83%)
(non-cognate) controls	1989 (1.41%)	2126 (2.51%)

4.3.3.1 Responses to cognates and control words

As with Experiment Three, the first set of analyses tested the significance of the cognate facilitation effect (see Table 7), together with an assessment of the kind of factors that have an impact on their processing. First, an RT analysis was carried out fitting a linear mixed-effects model to a data set comprising only accurate responses to cognates and their matched controls. Subsequently, an error analysis was performed by fitting a mixed logit model to the full data set, including both the accurate and inaccurate responses. Both statistical models included Status, Language and Proficiency as the main

predictors (for further model specification, see Experiment Three). Where variables did not reach significance as predictors and were not involved in any interactions, they were dropped from the model. Significant effects revealed in the final model fitted to participants' latency data are reported first, followed by the significant effects obtained from the final model fitted to the error data.

In line with the previous experiment, the analysis of response latencies revealed Status again as a significant simple effect (MCMCmean = 0.069, HPD95lower = 0.044, HPD95upper = 0.094, $p < 0.001$; see Table G9, Appendix D), with overall faster responses to cognates (see Figure 91). In addition, this effect was further defined by an interaction with stimulus frequency (MCMCmean = -0.045, HPD95lower = -0.081, HPD95upper = -0.008, $p = 0.013$, Figure 92), being mainly carried by the difference in the identification speed of low-frequency items; with faster identification of cognates. As discussed in the previous experiment, the latter finding replicates a recent observation by Dijkstra and colleagues (2010), revealing that the frequency of cognates is a major contributor in determining the extent to which the co-activation of cross-linguistic representations can actually facilitate recognition. The present result also extends the previous findings by showing that this effect is found irrespective of the language of processing (either L1 or L2) or the order of language dominance of a bilingual speaker (as shown in the emergence of an identical effect in both Experiments Three and Four).

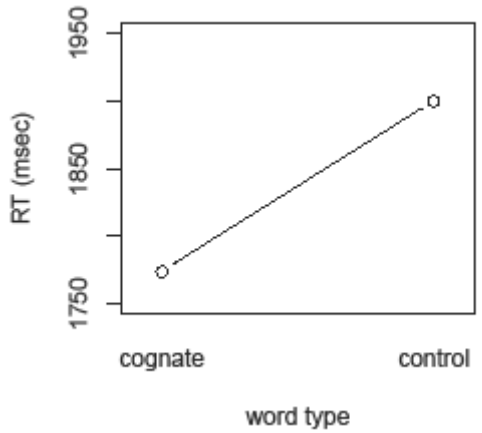


Figure 91. Mean response times for cognates and controls.

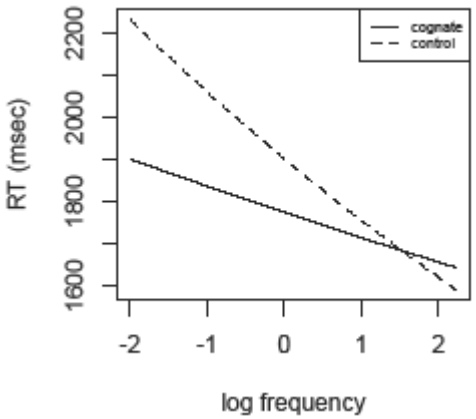


Figure 92. Mean response times for cognates and controls, in relation to the items' frequency.

The speed of responses to both item types was also affected by the language the bilingual has to perform the task in, with faster identification in the L1 (MCMCmean = -0.077, HPD95lower = -0.108, HPD95upper = -0.045, $p < 0.001$, Figure 93). The Language effect was further defined by an interaction with L2 Proficiency (MCMCmean =

0.075, HPD95lower = 0.041, HPD95upper = 0.110, $p < 0.001$, Figure 94), which reveals that, as might be expected, the lower their proficiency in the L2 the longer it takes participants to identify L2 words.

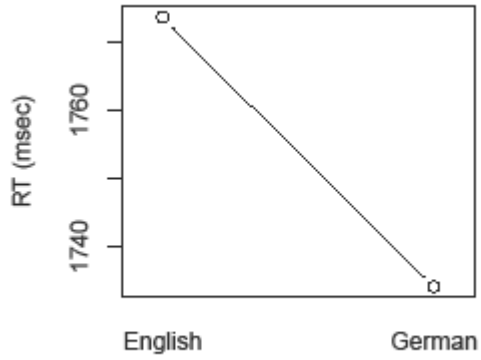


Figure 93. Mean response times for items, depending on the task language (L2 or L1).

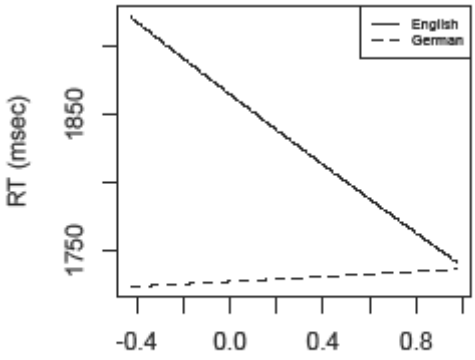


Figure 94. Mean response times for items by task language and with increasing L2 proficiency.

Finally, a number of secondary predictors were returned significant, all of which have been identified in various other studies. The significant effects include the Frequency (MCMCmean = -0.036, HPD95lower = -0.064, HPD95upper = -0.008, $p = 0.015$, Figure 95), Length (MCMCmean = 0.020, HPD95lower = 0.003, HPD95upper = 0.037, $p = 0.023$, Figure 96) and Word Class of items (MCMCmean = 0.041, HPD95lower = 0.018, HPD95upper = 0.065, $p < 0.001$, Figure 97), as well as the number of German neighbours (MCMCmean = -0.004, HPD95lower = -0.008, HPD95upper = -0.001, $p = 0.025$, Figure 98).

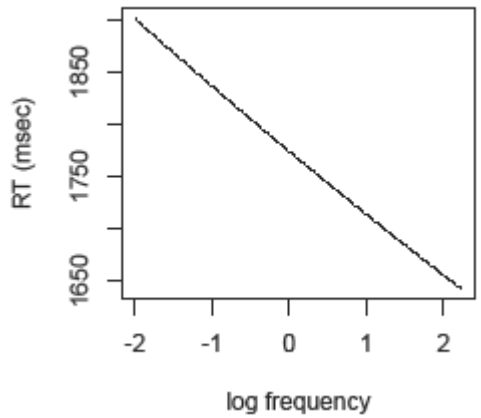


Figure 95. Mean response times for items with increasing log frequency.

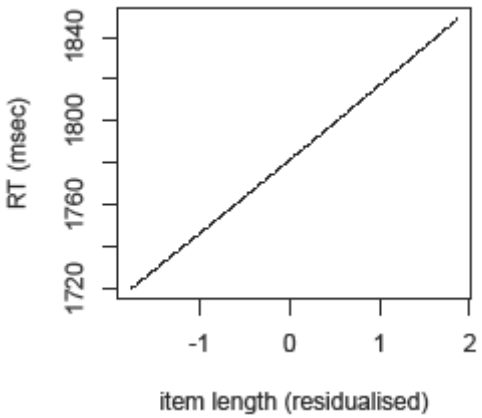


Figure 96. Mean response times for items with increasing length

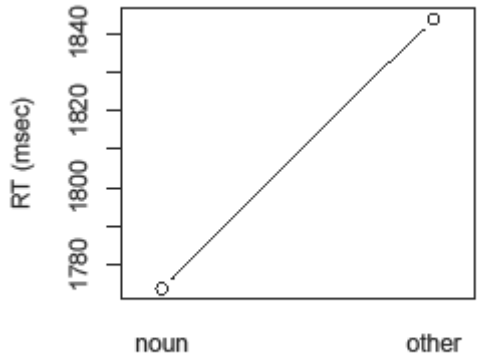


Figure 97. Mean response times for items, in relation to their syntactic category.

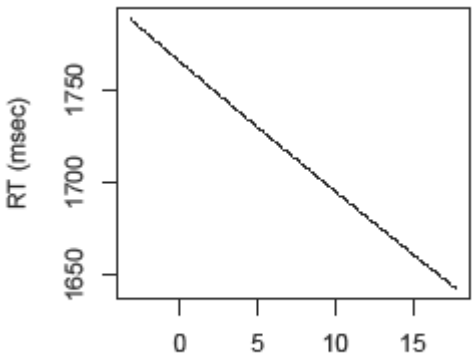


Figure 98. Mean response times for items, with an increasing number of German (L1) neighbours

With regards to the error analysis, Table 7 above reflects the fact that only a minimal proportion of items was falsely identified. This floor effect means that, similar to Experiment Three, only very few significant effects were returned in this analysis (see Table G10, Appendix D). The simple effects included Word Class (β = -0.588, SE = 0.232, z = -2.531, p = 0.011), with more correct responses to nouns, as well as the German and the English Neighbourhood Densities (β = 0.083, SE = 0.041, z = 2.016, p = 0.044 and β = -0.051, SE = 0.022, z = -2.323, p = 0.020, respectively). In the latter two cases, better identification was achieved with an increasing German neighbourhood size and a decreasing English neighbourhood size. Although not central to the research focus of this study, this finding could be explained by assuming that word identification processes were largely dominated by the bilinguals' L1. As noted by van Heuven and colleagues, however, more exhaustive empirical research is needed to determine the precise nature and combination of variables that affect the size and direction of bilingual neighbourhood effects, particularly in different tasks and language combinations (1998:480).

None of the main predictors were returned significant in the error analysis, nor were there any interactions between the predictors.

4.3.3.2 Cognate conditions

To examine the cross-language activation patterns of orthographic and phonological cognate representations for speakers with a reversed L1-L2 language background, an analysis was performed on responses to just the cognate items (i.e. excluding the controls). The statistical models fitted to the data included Orthographic Similarity and Phonological Similarity as two of the main predictors, as well as Language and

Proficiency. Other model specifications were identical to the previous analysis. Only variables that reached significance as predictors were kept in the model and are reported below, first outlining the results from the latency analysis, followed by the results of the error analysis.

Table 9.

Mean recognition latencies for correct responses (in milliseconds) and percent error rates[78] (in parentheses) for the four cognate conditions in L1 (English) and L2 (German).

Cognate condition	German (L1)	English (L2)
+O+P	1862 (0.47%)	1878 (1.4%)
+O-P	1986 (2.33%)	1927 (1.86%)
-O+P	1867 (1.55%)	1980 (2.56%)
-O-P	1813 (1.4%)	1879 (1.47%)

One of the significant simple effects revealed in the latency analysis was Phonological Similarity (MCMCmean = 0.057, HPD95lower = 0.009, HPD95upper = 0.102, p = 0.015; see Table G11, Appendix D), with faster responses to cognates that do not share the same phonological representation across two languages. More importantly, however, Phonological Similarity entered an interaction with Orthographic Similarity (MCMCmean = -0.075, HPD95lower = -0.144, HPD95upper = -0.008, p = 0.034). As can be seen in Figure 99, the interaction is mainly carried by the difference between identification times for the two sets of cognates with different orthographic representations across two languages, i.e. between those with a high degree of phonological similarity (e.g. *mouse*/*Maus*, a -O+P pair) and those with more distinct pronunciations (e.g. *book*/*Buch*, -O-P). That is, cognate identification is considerably slowed down when two fairly distinct orthographic representations mapped onto a highly similar phonological representation. This phenomenon has been previously interpreted as evidence for the *feedback effect*, i.e. a backward activation from phonology to orthography across languages (cf. Schwartz et al. 2007, and Figure 110 in Appendix D).

[78] Given that the nature of the task does not permit many errors (i.e. requiring a full identification of the gradually appearing word), the overall error rate is very low. The few instances where a response has been classified as an error reflect this fact, including merely letter substitutions, deletions or incorrect letter insertions. None of these errors are believed to have been a result of a typing mistake, since participants were explicitly informed that they can take as much time as they need for typing.

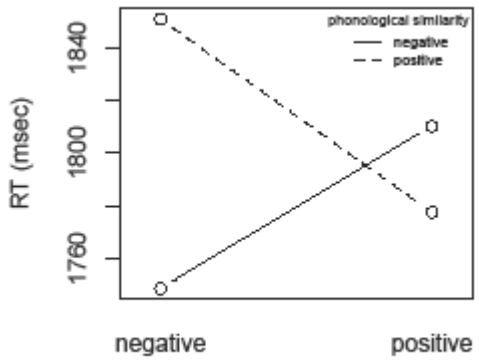


Figure 99. Mean response times for the four O-P cognate conditions (*lmer* model estimates), by their orthographic and phonological similarity (both positive vs. negative).

It could also be argued that there is some evidence for the opposite, i.e. the feedforward effect, even though noticeably weaker.[79] That is, cognate identification seems somewhat delayed when a highly similar orthographic representation maps onto two, more distinct phonological representations (e.g. *nation*/*Nation*, a +O-P pair) as opposed to highly similar ones (e.g. *finger*/*Finger*, +O+P). This finding not only confirms that phonology is automatically activated, even when it is not explicitly required, but it also suggests that the activated phonological codes compete with each other and delay identification.

Interestingly, this effect was not further defined by an interaction with Language (as seen in Experiment Three), which might suggest that German L1-English L2 bilinguals apply similar reading strategies for both languages. Similarly, there is no clear indication that the L2 proficiency level of a bilingual speaker has a major impact on the competitive processing dynamics between orthography and phonology.

Separate analyses of the +O/-O distinction shown in Figure 99, confirmed the general pattern of findings. That is, the analysis performed on a data set including only responses to highly dissimilar (-O) cognate material revealed Phonological Similarity as a simple effect (MCMCmean = 0.063, HPD95lower = 0.022, HPD95upper = 0.104, $p < 0.001$; Figure 100), as well as an interaction between Proficiency and Language (MCMCmean = 0.145, HPD95lower = 0.071, HPD95upper = 0.212, $p < 0.001$; Figure 101).

[79] Note that the amount of noise in the data - particularly in the -P condition - cannot be explained by a large range of similarity scores within that group, since it did not differ considerably from the distribution of similarity scores in the +P condition (1.33 versus 1.38, respectively).

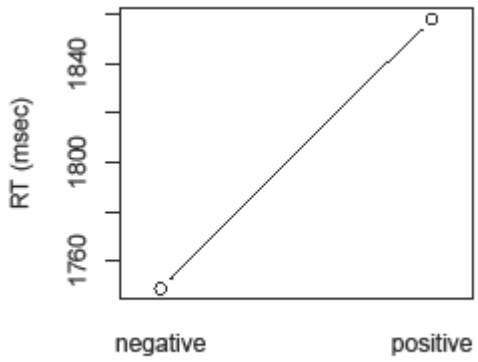


Figure 100. Mean response times for the –O cognate material (*lmer* model estimates), by their phonological similarity (both positive vs. negative).

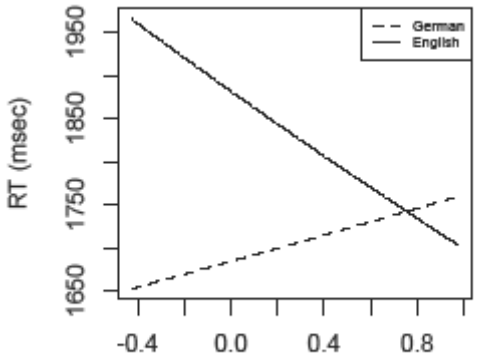


Figure 101. Mean response times for the -O cognate material (*lmer* model estimates), by task language and with increasing L2 proficiency.

An analysis of data reduced to orthographically highly similar (+O) cognate material, on the other hand, revealed no such effects. In fact, even though there was a marginally significant simple effect of Phonological Similarity (MCMCmean = -0.073, HPD95lower = -0.146, HPD95upper = 0.001, $p = 0.055$) in the statistical model, the effect disappeared following model criticism. The latter may explain why Figure 99 above displays a small difference in response times to cognates, conditioned by their phonological overlap across languages (for an interpretation of this finding, see the discussion section).

Returning to further effects obtained within the full (cognate) data set, where the predictors Language and Proficiency do play a role is in the speed with which cognates can be identified in one of the bilinguals' two languages. This is reflected in the simple effect of Language (MCMCmean = -0.111, HPD95lower = -0.176, HPD95upper = -0.048, $p < 0.001$), which also enters an interaction with Proficiency (MCMCmean = 0.135, HPD95lower = 0.064, HPD95upper = 0.206, $p < 0.001$, Figure 102). As seen in the previous experiment, the interaction is mainly carried by the contrasting speed of responses for lower proficiency bilinguals in their respective languages, i.e. faster in L1 (German) and slower in L2 (English). In addition, the data demonstrates yet again that less proficient bilinguals are faster in identifying cognates in their L1, compared to their more proficient counterparts. As discussed in the previous experiment, this finding can be interpreted as further evidence for the language non-selective access hypothesis, with higher global activation levels and increased cross-linguistic competition being the cause of response delays with increasing L2 proficiency.

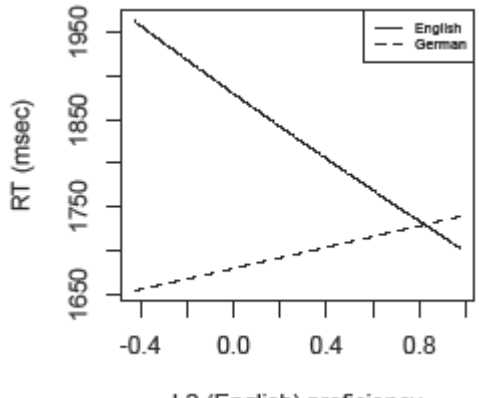


Figure 102. Mean response times for cognates in relation to task language and increasing L2 proficiency.

The same two predictors (i.e. Language and Proficiency) also enter a three-way interaction with Orthographic Similarity (MCMCmean = -0.112, HPD95lower = -0.212, HPD95upper = -0.016, p = 0.023, Figure 103).

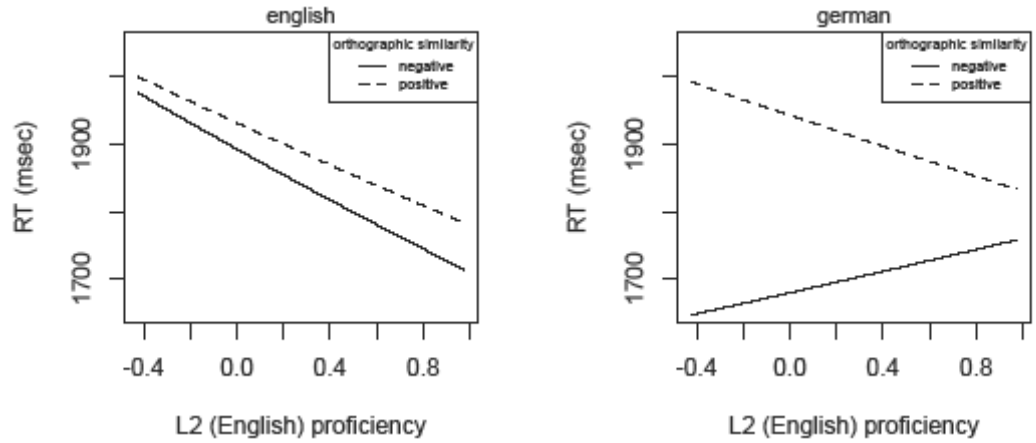


Figure 103. Mean response times for cognates, by orthographic similarity (positive vs. negative), participants' L2 proficiency and task language (L2 English, in the left panel, and L1 German, in the right panel).

This three-way interaction is carried by the difference in the speed with which low proficiency L2 speakers can identify a cognate in their L1 (German) that does not share the same orthographic representation with the L2. That is, low proficiency bilingual speakers are much faster at identifying cognates such as *Buch* or *Maus* (meaning 'book' and 'mouse', respectively), as opposed to *Finger* or *Hand*. Given that stimulus characteristics were matched across the different cognate conditions as closely as possible, it is less likely that the obtained latency difference derives from inherent lexical characteristics. A more likely explanation is that for lower proficiency speakers, there is less cross-linguistic activation of orthographically related but relatively dissimilar word entries, when processing in their L1. This reduced global activation leads then to less strong competitive dynamics, hence enabling faster processing and responses. In

contrast, while identifying orthographically highly similar words across two languages, or when processing is done in an L2, less proficient bilinguals may experience more cross-language activation and interference; hence, responses are slowed down. These conjectures seem largely supported by the findings of an additional analysis by Language. That is, only in the German (i.e. L1) subset of data was Orthographic Similarity revealed as a simple effect (MCMCmean = 0.110, HPD95lower = 0.046, HPD95upper = 0.172, $p < 0.001$), together with its interaction with Proficiency (MCMCmean = -0.101, HPD95lower = -0.166, HPD95upper = -0.032, $p < 0.001$). Interestingly, the additional analysis identified a further factor which may modulate cognate responses in an L2, namely frequency. Although reflecting merely a trend towards significance, Frequency entered an interaction with Orthographic Similarity in the English data subset (MCMCmean = -0.081, HPD95lower = -0.167, HPD95upper = 0.003, $p = 0.063$; see Figure 104). That is, while the identification of orthographically less similar cognates was not clearly affected by their frequency, orthographically highly similar cognates were more difficult to identify in the L2 when their frequency was low. A possible interpretation of this finding is that, as may be expected, something that has a similar form across languages but is not very frequent in the L2 is more difficult to identify, given a high competition from its L1 equivalent; hence, its recognition is delayed. With increased frequency of L2 items, the amount of interference and competition from the L1 representations may be less sustainable and/or detrimental on L2 processing. Notably, an equivalent prediction was recently made by Dijkstra and colleagues who hypothesised that "form overlap between the two readings of the cognates leads to a frequency-dependent parallel activation of these readings" (2010:300).

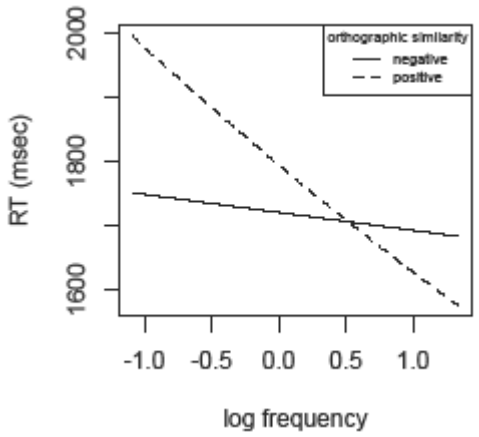


Figure 104. Mean response times for cognates in the L2 task, in relation to their orthographic similarity and increasing log frequency.

The earlier notion of an increased global activation causing greater cross-linguistic interference, seems further supported in the (full) cognate data set by the interaction of Orthographic Similarity and Language (MCMCmean = 0.116, HPD95lower = 0.038, HPD95upper = 0.202, $p < 0.001$), where responses to orthographically more similar cognates (e.g. *finger-Finger*) are slowed down when compared to more distinct orthographic representations (e.g. *cat-Katze*). The same effect is observed for both languages (although less pronounced in the L2), which suggests that very similar competitive dynamics influence processing in the two languages.
Finally, with increasing proficiency in an L2 (see Figures 102 and 103), items seem to be processed in a similar manner, irrespective of their degree of orthographic overlap or the task language. In addition, more advanced bilingual speakers seem better able to process cognates, i.e. resulting in an overall faster identification. The latter finding may be indicative of different cognitive processes (if not certain benefits) that come with increasing proficiency (e.g. Midgley et al. 2009; see also Mitchell & Myles 2004, for an overview of relevant aspects and work).
The last effect to be reported for the latency analysis was found among the secondary predictors, namely the German Neighbourhood Density (MCMCmean = -0.007, HPD95lower = -0.011, HPD95upper = -0.003, $p < 0.001$). In this context, an increase in the number of German neighbours seems to facilitate identification and speed up responses.

With regards to the error analysis, the logit model fitted to the data revealed a result pattern that differed from the results reported above (see Table G12, Appendix D). That is, Word Class was returned as a significant simple effect ($\beta = -0.974$, SE = 0.369, $z = -2.642$, $p = 0.008$) and the German neighbourhood density showed a strong tendency towards significance ($\beta = 0.111$, SE = 0.061, $z = 1.837$, $p = 0.066$). There were no significant interactions.

4.3.3.3 Discussion

The general pattern of responses obtained in the present experiment confirms that a major determinant of bilingual cognate processing is their frequency. On the one hand, a clear facilitation effect was only obtained for low-frequency cognates relative to their matched controls. This suggests that in a task which requires the precise identification of an item, the availability and co-activation of an equivalent non-target representation is particularly beneficial for lower frequency cognates relative to other words. The other side of the coin, however, is that the co-activation of a non-target representation may not always be beneficial. That is, the present data revealed that (relative to other cognates) the lower the frequency of a form-identical cognate, the stronger the negative effects of cross-linguistic activation and interference. While the former finding is in line with Experiment Three and previous research (e.g. Dijkstra et al. 2010), both results provide further support for the hypothesis that bilingual lexical access is fundamentally non-selective (cf. van Heuven et al. 2008).

Cognate identification is also shaped by the language of processing and by the L2 proficiency of a bilingual speaker, although only a more detailed cognate analysis was able to reveal these effects. The appearance of such effects in the present data reflects the fact that bilingual performance stands in direct relation to global activation levels, with reduced global activation leading to improved performance (cf. Midgley et al. 2009). Accordingly, the identification of cognates in the L1 seems to induce less cross-linguistic activation and competition, which in turn facilitates responses, particularly for less proficient bilinguals. Because we can expect generally higher global activation levels with increasing L2 proficiency, no such task language advantage is observed with more proficient bilinguals, whose performance in the L1 is somewhat inhibited in comparison to their less proficient counterparts. Instead, more proficient bilingual readers perform equally well in both languages. As suggested in the previous experiment, this pattern of responses could also be interpreted as indicative of the developmental nature of bilingual lexical organisation, with the two languages becoming increasingly combined.

More central to the main research interest of this study is the extent to which bilingual cognate identification can be modulated by sub-lexical processes in form of cross-linguistic feedforward and feedback activation. Once again, the present results are able to show that evidence for the sub-lexical interactive dynamics between the orthographic

and phonological representations can be obtained in a task which does not require the explicit activation of phonology.

In line with my prediction, the data revealed a clear feedback effect, with a delayed identification of cognates in the -O+P condition (e.g. *mouse*/*Maus*) in comparison to cognates in the -O-P condition (e.g. *book*/*Buch*). A possible explanation for this finding is that the orthographic input encountered by the bilinguals in the present study was able to activate a fairly detailed phonological representation. Where the activated phonological code was shared to a great extent by the bilinguals' two languages, it would have allowed sufficient information to flow back to the orthographic level and across languages, giving rise to inhibitory competition from a non-target representation. Evidence for cross-language feedforward effects, on the other hand, was not as strong in the present experiment. Even though there was a trend indicating faster cognate identification in the +O+P condition (e.g. *finger*) as opposed to the +O-P condition (e.g. *nation*), the differences could not be statistically confirmed. While it would not be surprising to find the lack of such an effect in the L1 (cf. Haigh & Jared 2007, and the previous experiment; but see Schwartz et al. 2007:120, for a contrasting finding and argumentation), the finding that no obvious inhibitory cross-language activation seems to have taken place when cognates were identified in the L2 is less expected, particularly in light of its appearance in the previous experiment.

The most likely explanation of the absence of an effect may be linked to a relatively higher L2 proficiency of the bilinguals compared to those participating in the previous experiment. To test this conjecture, I performed Welch's two-sample *t*-test, which compared the L2 proficiency of the two bilingual groups participating in this study. The test showed that the groups varied significantly [$t(70) = -8.96$, $p < 0.001$], with German-English bilinguals being more proficient than their English-German counterparts from Experiment Three.[80]

This finding could imply that bilingual cognate identification in the present experiment may have involved very similar processes in both the L1 and in the L2. Accordingly, we can assume that in the present experiment lexical representations in both the participants' L1 and L2 were similarly robust and that participants were processing cognates in a comparably automatic fashion. As a result, the overall amount of

[80] Even though the proficiency tests were conducted in different languages, I strongly believe that they resemble each other and that this comparison is a fair one. Unfortunately, there is no independent evaluation of the equivalence of the two tests currently available.

(feedforward) activation and interference from a non-target phonological representation may have been considerably reduced, irrespective of the language of processing.

This interpretation also explains why the effects of feedback and feedforward activation are not further defined by the language of the task. That is, if we assume that increased L2 proficiency follows an increasingly merging processing mechanism (cf. Midgley et al. 2009:296, and Study One in this dissertation), then we may interpret the present results in terms of a system that is fairly advanced in its integrating development. Interestingly, the main implication of this assumption is that a joined, rather than two separate GPC system(s) is formed (van Wijnendaele & Brysbaert 2002); possibly with higher resemblance to the L1 processing mechanism, than vice versa. This is an exciting finding, and needs a more comprehensive investigation in future research.

4.4 General Discussion

The primary aim of the present study was to further explore the previously reported feedforward and feedback effects in bilingual lexical processing. One objective was to investigate whether such effects could be obtained in an exclusively visual task. The other two objectives tested to what extent the characteristics of a particular language and/or L2 proficiency can modify cross-linguistic activation dynamics. To do this, the study employed cognates, which were manipulated in the degree of their orthographic and/or phonological overlap (cf. Schwartz et al. 2007). Because a recent study identified word frequency and semantic overlap as major determinants of the recognition speed of items in progressive demasking (the task chosen for the present study; cf. Dijkstra et al. 2010), both variables were matched across items and languages as closely as possible. This would ensure that the appearance (or absence) of the respective effects could be confidently ascribed to cross-linguistic interactions at the orthographic and/or phonological levels.

The results of the two experiments not only show that competitive activation dynamics between the orthographic and phonological level across languages are in place during an exclusively visual task, but also that language-dependent sub-lexical decoding processes can modulate the strength of these effects. The experiments also show how proficiency in the L2 is reflected in the cross-language dynamics. In the following, it is first discussed how the cross-linguistic consistency effects obtained in both experiments can be interpreted in terms of the current models of bilingual processing. Next, it will be considered how the orthographic depth of the bilinguals' two languages had an impact on the processing dynamics. And finally, a brief account of the role of L2 proficiency will be given.

Feedforward and feedback effects

The present results confirm prior findings by showing that bilingual lexical processing is influenced by the consistency of the orthographic to phonological and phonological to orthographic mappings across languages (e.g. van Assche et al. 2011; Schwartz et al. 2007). The former is shown in the present data in the form of feedforward effects. That is, the identification of cognates which are orthographically highly similar across languages (+O condition) was slower when the corresponding phonological codes varied considerably as opposed to when the corresponding phonological codes were highly

similar. Feedback effects (i.e. evidence for the consistency effect of phonological to orthographic mappings across languages; -O condition), on the other hand, were present in the identification of cognates whose orthographic representations are fairly distinct across languages. More specifically, the identification of such cognates was delayed when there was a high degree of phonological overlap relative to when the corresponding phonological representations were more distinct.

Both effects can be explained within the theoretical framework of the BIA+ model, which has been most successful in accounting for a range of bilingual visual processing phenomena (for a recent statement of the advantages of this over other theoretical frameworks, see Dijkstra & van Heuven 2002a; Dijkstra et al. 2010). The model incorporates a very elaborate interactive network between orthography and phonology (first presented in an earlier model called SOPHIA, van Heuven & Dijkstra 2001; for a recent appraisal and significance of the model, see de Groot 2011; Thomas & van Heuven 2005), which forms the basis for the observed effects. As outlined in section 2.3.2, upon the presentation of visual input a corresponding orthographic representation is activated first. The flow of information is modelled to be bi-directional (upwards to the next higher level(s) and downwards from higher levels above) and lateral (between the orthographic and phonological representations). That is, the activated orthographic code gives rise to (cross-linguistic) activation at both a higher orthographic level (e.g. orthographic clusters, then syllables, then complete words), and the corresponding, laterally connected phonological representations. Orthographically and/or phonologically ambiguous input causes lexical competition, which needs resolving before a word can be fully identified. What this means in the context of the present findings is the following: in the +O condition, an orthographically highly similar cognate representation across languages is assumed to activate a phonological representation from both the target and the non-target language. Where the activated phonological representations are most dissimilar from each other, lexical competition will be the greatest; hence a delay in cognate identification.

In contrast, a cognate input with relatively little resemblance to its equivalent in the non-target language (i.e. -O condition) is less likely to activate a competitor phonological representation. In this case, it is presumably the amount of cross-linguistic similarity at the phonological level that leads to subsequent lexical competition. Accordingly, the higher the cross-linguistic similarity of an activated phonological representation, the

stronger the co-activation of a non-target orthographic representation (due to lateral and downward transmission of information). As previously, resolving the created cross-linguistic competition is what delays cognate identification.

While the present results are largely in line with the findings obtained by Schwartz and colleagues (2007) in their naming task, they extend the previous work by showing that similar effects can be observed in a task which does not explicitly require the activation of phonology. The obtained results also go beyond previous bilingual research by providing first explicit evidence for feedback effects. This finding is consistent with earlier monolingual research, which collected evidence for feedback effects in lexical decision (e.g. Lacruz & Folk 2004; Perry 2003; Pexman et al. 2001; Stone et al. 1997; Ziegler et al. 1997). Notably, there have also been a number of failures to replicate the effect, and more recently, neural network simulations have cast doubt on the actual existence of such effects (see Ziegler et al. 2008, for a comprehensive overview of previous studies, arguments and the relevant simulation work).
The reader will recall that even though evidence for both effects was collected in the current study, their presence and magnitude was defined by further experimental factors. One such factor was the orthographic depth of a language, having arguably an impact on the occurrence of feedback effects. In the following, I discuss this possibility in relation to the interrelated aspects of the phonological decoding process and the language of processing. It will be argued that the findings of the present research may postulate a tentative explanation as to why evidence for feedback effects has been found so inconsistently.

Cross-linguistic consistency effects and orthographic depth

Perhaps the most intriguing finding of the current study is the fact the bilingual speakers with reverse language dominance seem to employ different reading strategies from one another, which in turn has an impact on the kind of processes that seem to take place in bilingual visual word recognition. This is primarily evident in the present data in the dissimilar occurrence of the feedback effect: while a robust effect was obtained for German-English bilinguals, only a trend in the predicted direction was evident in the

response latencies for English-German bilinguals, and only when processing was in the L2 (i.e. in German).[81]

To explain the absence of feedback effects, researchers have typically argued that these effects may be more difficult to obtain in tasks in which stimuli are presented visually. This is partly because

> feedback consistency in the visual modality could be considered second-order feedback: Orthography is activated by the visual input, and then orthography has to feed forward to phonology (O → P) before the computed phonology can feed back to orthography (P → O). (Ziegler et al. 2008:651)

However, the present results indicate strongly that this may not be the only reason for the absence of such effects. Another possibility put forward by the proponents of the bi-directional consistency effect in the monolingual domain (e.g. Frost & Ziegler 2007; Grainger & Ziegler 2007; van Orden & Goldinger 1994) is an "underspecified and coarse computation of phonology" (Ziegler et al. 2008:656). Accordingly,

> [i]f the initial computation of phonology is coarse or underspecified (e.g., vowels might not be assembled at first), then feedback effects are difficult to show because these effects require phonology to be fully activated. (op.cit.)

This interpretation is more in line with my previous discussion. That is, given that feedback activation was only truly evident in connection to the more shallow of the two languages – i.e. for German L1 bilinguals, and then with a similar trend when the English L1 bilinguals were identifying cognates in their L2 (German) – the present findings could be interpreted as direct evidence that orthographic depth plays a crucial role in the sub-lexical processing of word items. More specifically, if there is substance in the proposal that lexical access in shallow orthographies is based on a relatively detailed phonological representation as opposed to an impoverished one for deep orthographies (cf. Frost 2005:258), then shallow orthographies may better allow for the emergence of feedback effects, relative to deep orthographies. Consequently, a fully activated phonological code in a shallow language would be better able to co-activate a non-target orthographic representation, leading to competitive dynamics at the orthographic level and response inhibition. In contrast, the activation of an impoverished phonological representation for speakers of deep orthographies may not accumulate signals that are

[81] As mentioned in the relevant discussion section, this latency trend could not be statistically verified. However, given that Schwartz and colleagues (2007) obtained a similar trend in their data, it seems that there is at least some validity in arguing this point.

strong or complete enough to be fed back to the orthographic level; or, alternatively this information may not be complete enough to allow for sufficient cross-linguistic activation. Notably, there is currently no mechanism in place in the BIA+ model to simulate such strategic "grain-size effects" (Green 2002:206–207).

These interpretations may explain why the present study provides such robust evidence for feedback effects for German-English bilinguals. In addition, at present I am not aware of monolingual research on bi-directional consistency effects that has been conducted in other than deep orthographies (for useful studies in a relatively shallow language and a closely related area, i.e. investigating the phonological consistency effect in Dutch, see Martensen, Dijkstra, & Maris 2005; Martensen, Maris, & Dijkstra 2003). Accordingly, it could be argued that this is the reason why feedback effects in the monolingual domain have been found so sporadically, and why the interpretation presented in the current study has not been entertained to date.

Another interesting addition to the current knowledge of bilingual phonological activation and processing is the fact that bilinguals seem to gradually acquire (if not master) an L2 GPC system, or alternatively allow for phonological decoding processes other than those typical of their L1 to take place (for a similar proposal, see van Wijnendaele & Brysbaert 2002). Although the reliability of the predicted effect ideally could have been more robust, the picture emerging from the present data seems to reflect the changing dynamics. That is, when identifying cognates in their L2 the English-German bilinguals' response latencies appear in the direction predicted by the feedback effects.

Finally, even though the development of 'new' reading strategies may be closely related to the L2 proficiency of a bilingual speaker (hence, the occurrence of a response trend rather than a conventionally significant effect), it seems that ultimately a phonological decoding system is developed which is capable of processing either language. This interpretation may explain the fact that the (overall more proficient) German-English bilinguals employed similar decoding strategies, irrespective of the language of processing.

While the construal of the involvement of L2 proficiency in feedback effects is rather tentative, its impact on the feedforward effects is much clearer. This will be discussed in the following section.

Cross-linguistic consistency effects and proficiency

Perhaps somewhat surprisingly, clear evidence for the feedforward effects was only truly obtained on one occasion, namely during the English-German bilinguals' L2 cognate identification. Given that feedforward effects are argued to reflect the amount of activation and subsequent interference from a non-target phonological representation (whether within a single language or cross-linguistically, e.g. Schwartz et al. 2007; Stone et al. 1997), the present results provide evidence that under certain circumstances such an effect may be largely limited to processing in the bilinguals' weaker language. This finding is in line with the emerging pattern from current bilingual research. That is, depending on the experimental setup (including task and/or stimulus material), researchers have either obtained or failed to obtain traces of cross-language activation of phonological codes (see section 4.1.1 above, for relevant references). The main argument put forward to explain the occurrence of feedforward effects in the L2 is compatible with both the non-selective access hypothesis (for an overview, see Dijkstra & van Heuven 2002a; Kroll & Dijkstra 2002) and the strong phonological model of visual word recognition (Frost 1998). That is, upon the encounter of an orthographic input a mandatory, pre-lexical activation of a phonological code (probably highly imprecise for deep orthographies, see discussion above) takes place, which is subjected to effects from both the L1 and the L2 (cf. Brysbaert et al. 1999). In contrast, the absence of feedforward effects during processing in the L1 has been typically explained by the fact that "bilinguals nonselectively activate representations from both languages but that first language target representations quickly inhibit second language representations" (Haigh & Jared 2007:638).

In a similar fashion, Lemhöfer and colleagues argued that "L2 processing (regardless of specific cross-language influences) is fundamentally different from word processing in L1" (2008:27), which may explain the usually distinct results obtained in the L1 as opposed to the L2. In this context, Study Two presents an extension of that research by showing that the obtained differences in processing may be partly a result of an interaction of orthographic depth differences and the language of processing.

More recently, however, the findings of an ERP study were taken to show that the L2 proficiency of a bilingual speaker may be a further defining factor in the observed differences between L1 and L2 processing. The research argued that "word recognition in L2 involves distinct mechanisms compared with the first language, at least in the relatively early phases of L2 acquisition in late learners of L2" (Midgley et al. 2009:296).

This is compatible with the findings of the current study, showing contrasting processing behaviour for the less proficient English-German bilinguals (i.e. evidence for feedforward effects in the L2, no effects in the L1), relative to the German-English bilingual group. Accordingly, the absence of feedforward effects in both languages for German-English bilinguals may indicate their relative higher L2 proficiency, and hence more similar processing mechanism(s).
Clearly, either interpretation calls for new and exciting future research directions.

To conclude, the present study has led to several conclusions about the competitive activation dynamics from orthography to phonology across languages. First, cross-linguistic consistency effects were obtained in a study which did not explicitly require the activation of phonology. Second, L2 proficiency and/or language dominance were shown to have a relatively minor but significant impact on the co-activation of a competing phonological representation. This was reflected in the limited occurrence of feedforward effects. Most importantly, however, the present findings indicate that cross-linguistic consistency effects seem to be modulated by the orthographic depth of a bilingual's two languages. This was mainly reflected in the amount or strength of cross-linguistic feedback activation. Although more research is needed to further confirm the conjectures postulated in the present research, these initial results may be informative with respect to processing in orthographically distinct languages and to the developing phonological decoding system in bilinguals.

Chapter 5
CONCLUSION

This thesis explored the processes involved during bilingual visual word processing, focussing on sub-lexical processing which has received relatively little attention to date. The first study concerned mainly orthographic aspects of sub-lexical recognition, and the second study investigated interactions between orthography and phonology. To provide a comprehensive view of these aspects of bilingual visual word recognition, both studies were conducted with English-German and German-English bilingual groups, both of whom had varying proficiency in their L2. The paragraphs below will briefly summarise the main conclusions of each of the studies and assess how well the BIA+ (the most suitable theoretical model of bilingual visual recognition for the present purposes, see section 2.3.2) can account for the present findings. These paragraphs will be followed by a brief account of potential limitations of the studies, together with suggestions for future research directions.

Chapter 3 examined the extent to which bilingual speakers make use of sub-lexical language-specific information to determine the language membership of an item. The present data suggest that this kind of information is particularly useful when the lexical selection process has not yet become automatic. More specifically, the manipulation of sub-lexical cues had a measurable impact on real word responses only for less proficient bilingual speakers. When English-German bilinguals were conducting the task in their L2, this was reflected in a facilitative effect for their responses to both existing German words and interlingual homographs when the target words carried an onset capital letter. For the German-English bilingual group, however (i.e. conducting the task in their L1), facilitation was only evident in responses to interlingual homographs. In contrast, when responses were made to nonwords all bilinguals were affected in a similar way; this time, the sub-lexical cue had an inhibitory impact on participants' responses. These findings are consistent with the idea that the bilingual word recognition system includes an early language discrimination mechanism. The findings further extend this proposal by showing that this mechanism is responsive to signals from the earliest processing stages. The present findings also go beyond previous research by demonstrating that the effectiveness of such a mechanism is restricted to situations where processing is less automatic.

In Chapter 4, the focus was on sub-lexical processes at the phonological level and on the extent to which activated phonological information transmits signals back to the orthographic level and across languages. The findings of this study give further support to the argument that cross-language competition can arise at the phonological level in purely visual tasks. The novel contribution of the current research to this view is the proposal that the magnitude of such effects depends on the L2 proficiency of a bilingual speaker and the language of processing (i.e. L1 or L2). The present results also provide the first evidence of the need to consider the orthographic depth of a bilingual's two languages. This is reflected in the present study in the amount of detectable interaction between the activated phonological and orthographic representations across languages, which are subject to the orthographic transparency of a bilingual's two languages and to the associated phonological decoding strategies.

As seen in the previous chapters, out of the three most prominent models of bilingual word recognition introduced in Chapter 2, the model that could, in principle, best capture the current results is the BIA+ (Dijkstra & van Heuven 2002a). Perhaps the most straightforward modification of the BIA+ to include the present results relates to the obtained proficiency effects. As discussed earlier, on the assumption that the degree of L2 proficiency is associated with the resting levels of activation of L2 words relative to L1 words, this could be done through an adjustment of these levels. More specifically, to model the finding of higher levels of L1 interference in an L2 task for less proficient participants, the resting level of activation for L2 words should be set lower for such participants compared to more advanced bilingual speakers (cf. Schulpen 2003:203). Technically, the interactive network of the BIA+ also allows us to simulate the observed interactions between orthography and phonology across languages. However, there is currently no mechanism in place to simulate certain strategic effects, such as the 'grain-size effects' identified in Chapter 4 (Green 2002:207). In fact, Dijkstra and van Heuven admit that the architectural structure of the BIA+ in its current form may be "too simple" to capture certain subtle cross-linguistic effects, including those at the sub-lexical level (2002b:221–222). On the assumption that the cross-linguistic effects observed in the current studies appeared at the sub-lexical level, it remains to be seen how future adjustments of the BIA+ will be able to simulate such subtle effects.

Finally, it is also not entirely clear how the BIA+ can account for the effects of sub-lexical language-specific cues. One possibility is that language-specific orthographic

information is directly linked to the language nodes, which can send signals back to the orthographic level, in a top-down way. Views on this matter are quite divided. While some researchers argue that language membership cannot be determined in this way, because information about the language of processing becomes available rather late in the recognition process (Dijkstra & van Heuven 2002a; e.g. Dijkstra et al. 2000; Schulpen 2003), others argue that this pathway is a legitimate option (e.g. Green 2002; Libben & Titone 2009; Schwartz & Kroll 2006). Alternatively, if the identified early language discrimination mechanism operates bottom-up, this "could in principle be implemented in the BIA+ framework" (Dijkstra & van Heuven 2002b:222), although no further specifications are given.

It remains to be seen how future adaptations of the BIA+ model can incorporate the empirical data presented in this thesis. In the meantime, further research can be collected providing more insight into various aspects of sub-lexical processing. Some suggestions for further investigation are outlined below.

Limitations and future research directions

Limitations of the research in this thesis simultaneously provide areas for future research directions. These complement the suggestions for further research discussed in previous chapters and will be outlined in the following sections.

With regard to the first study (Chapter 3), there are at least four areas worth further consideration. First, given the key role that has been identified for sub-lexical language-specific information, it would be interesting to further study this feature. One aspect is its grammatical function (i.e. in German all nouns receive an initial capital letter, hence this cue signalled a word's syntactic category), which was revealed in more advanced bilinguals' responses. Although researchers acknowledge the fact that grammatical information must be stored in some form with a lexical representation, currently little is known about how bilinguals make use of such information and what effects it might have on bilingual lexical processing (e.g. Baten et al. 2011; Green 1998b). In the context of the present study, this would call for a more rigid selection of bilingual material, with matching syntactic categories across languages along with the (mis)matching presentation formats included in the current research.

Second, a related aspect which was mentioned in earlier discussion is the fact that the current study employed a stimulus presentation format that was conceivably biased

towards one or the other language (i.e. with Title presentation denoting a German item and lower case suggesting an item may be English). To reinforce the suggestions made in that earlier discussion, it would be useful to conduct a parallel experiment including a more neutral presentation format, namely all upper case (e.g. GIFT). This would provide a baseline for responses - unbiased towards a particular language - against which the responses to items with the presentation formats used in this thesis could be tested. Such a comparison could clarify the magnitude of the impact of language-specific sub-lexical cues.

Third, another argument put forward earlier related to the fact that the sentences provided in the current study seemed to have limited impact on participants' lexical processing. Instead, I argued that sub-lexical language-specific information may play a more prominent role in bilingual word recognition, over and above the role of the language of the prime sentence. Given that the effects of sentence context are still under debate (e.g. Elston-Güttler, Gunter, et al. 2005; van Heuven et al. 2008; Libben & Titone 2009; Schwartz & Kroll 2006; Titone et al. 2011), it would be useful to further clarify this claim. This could be done in various ways; for instance, by conducting a study in which the sentences are biased towards the meaning of an interlingual homograph which is incongruent with the language of a target item represented in its presentation format (e.g. sentence bias towards the English meaning of an IHG, with a Title case presentation of target items). Alternatively, another study could employ sentences in which the critical stimulus material is embedded in a sentence rather than appearing in a sentence-final position. (This study would also call for more advanced measuring techniques, such as eye-tracking.) Yet another possibility would be to compare the current results with a study based solely on isolated word recognition (i.e. a study with no sentence context). The results of a series of studies designed to explore these features would be more conclusive in terms of the exact role (i.e. primary or secondary) of sub-lexical language-specific information.

Finally, one aspect that remains to be explored further is the exact point at which sub-lexical cues influence bilingual language selection. While the previous discussion has suggested that this may take place at a very early stage of the word recognition process, i.e. via top-down inhibition executed by means of the language nodes (or language tags, cf. Green 2002), there is also the possibility that this influence develops fully at some later decision stage and is bottom-up rather than top-down (cf. Dijkstra & van Heuven 2002a). Although some research has tackled this and similar issues before (e.g. von

Studnitz & Green 2002a, 2002b), the current evidence may be too limited to arrive at definite conclusions.

With respect to the second study (Chapter 4), perhaps the most intriguing question concerns the effects found for increasing proficiency in the L2. One suggestion made in this context was that GPC rules for L2 may be cumulatively added to those for L1 (for a similar proposition, see van Wijnendaele & Brysbaert 2002). This should be further tested against proficiency, together with an investigation of whether the L1 GPC system would remain the more dominant one.
A further research direction should also address the question of whether similar feedforward and feedback effects to those obtained in the present study would also be observed with a continuous distribution of the orthographic and phonological similarity measurements. That is, both these variables (i.e. orthographic similarity and phonological similarity) were dichotomised for the purposes of the present study, even though previous research has pointed out potential problems with dichotomisation (for references, see section 3.2.2.1). The main reason to use dichotomised variables in the present study was its occurrence in Schwartz et al.'s (2007) research, whose findings I aimed to partially replicate. While deferred analyses using the actual (i.e. continuous) values for the two variables failed to fully replicate the reported effects (e.g. with inconsistent effects of phonological similarity), I believe that this outcome primarily reflects the limited number of items in a specific condition and a wide dispersion of item values. Interestingly, while the results of some related studies suggest that the effects of phonology may be lost when using a continuous measurement of phonological similarity across languages (e.g. Dijkstra et al. 2010), other studies have obtained a clear phonological similarity effect (e.g. van Assche et al. 2011) despite using this measuring technique.

To conclude, while the studies reported in this thesis present interesting new insights into the processes underlying bilingual visual word recognition, particularly in relation to changing effects with increasing L2 proficiency, additional work is needed to support the various findings. I have proposed a number of ways in which some of the aspects could be addressed. These may provide exciting new avenues for future investigation, leading to a more comprehensive view of a developing bilingual lexicon.

REFERENCES

Albrecht, U., Dane, D., Fandrych, C., Grüβhaber, G., Henningsen, U., Kilimann, A., & Schäfer, T. 2005. *Passwort Deutsch: Kursbuch 1*, Stuttgart: Ernst Klett International.

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. 1996. The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition* 24(4): 477–492.

Andrews, S. 1982. Phonological recoding: Is the regularity effect consistent? *Memory & Cognition* 10(6): 565–575.

Andrews, S. 1989. Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(5): 802–814.

Andrews, S. 1996. Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language* 35(6): 775–800.

Andrews, S. 1997. The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review* 4(4): 439–461.

Andrews, S. 2006. Preface. In Sally Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*, xix–xxix. New York, NY: Psychology Press.

de Angelis, G. 2007. *Third or additional language acquisition*, Clevedon; Buffalo: Multilingual Matters.

van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. 2011. The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language* 64(1): 88–107.

van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. 2009. Does bilingualism change native-language reading? *Psychological Science* 20(8): 923 –927.

Baayen, R. H. 2008. *Analyzing linguistic data. A practical introduction to statistics using R*, Cambridge: Cambridge University Press.

Baayen, R. H. 2010. A real experiment is a factorial experiment? *The Mental Lexicon* 5(1): 149–157.

Baayen, R. H., Davidson, D. J., & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4): 390–412.

Baayen, R. H., Feldman, L. B., & Schreuder, R. 2006. Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* 55(2): 290 – 313.

Baayen, R. H., & Milin, P. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2): 12–28.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. 1993. *The CELEX lexical database [CD-ROM]*, Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Balling, L. W. 2008. A brief introduction to regression designs and mixed-effects modelling by a recent convert. *Copenhagen Studies in Language* 36: 175–192.

Balling, L. W., & Baayen, R. H. 2008. Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes* 23(7-8): 1159–1190.

Balota, D. A. 1994. Visual word recognition: The journey from features to meaning. In Morton Ann Gernsbacher (Ed.), *Handbook of psycholinguistics*, 303–358. Amsterdam: Academic Press.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. 2004. Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General* 133(2): 283–316.

Balota, D. A., Yap, M. J., & Cortese, M. J. 2006. Visual word recognition: The journey from features to meaning (a travel update). In Matthew J. Traxler & Morton Ann

Gernsbacher (Eds.), *Handbook of psycholinguistics*, 285–375. Amsterdam: Academic Press.

Baron, J., & Strawson, C. 1976. Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance* 2(3): 386–393.

Baten, K., Hofman, F., & Loeys, T. 2011. Cross-linguistic activation in bilingual sentence processing: The role of word class meaning. *Bilingualism: Language and Cognition* 14(3): 351–359.

Bates, D. M., Maechler, M., & Dai, B. 2007. *lme4: Linear mixed-effects models using S4 classes.* Retrieved from http://lme4.r-forge.r-project.org/ on 23 April 2010.

Beauvillain, C., & Grainger, J. 1987. Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language* 26(6): 658–672.

Besner, D., & Smith, M. C. 1992. Basic processes in reading: Is the orthographic depth hypothesis sinking? In Ram Frost & Leonard Katz (Eds.), *Orthography, phonology, morphology, and meaning*, 45–66. Amsterdam: North-Holland.

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. 1997. Masked orthographic priming in bilingual word recognition. *Memory & Cognition* 25(4): 447–457.

Borgwaldt, S. R., Hellwig, F. M., & de Groot, A. M. B. 2005. Onset entropy matters – Letter-to-phoneme mappings in seven languages. *Reading and Writing* 18(3): 211–229.

Boulinguez, P., Barthélémy, S., & Debu, B. 2000. Influence of the movement parameter to be controlled on manual RT asymmetries in right-handers. *Brain and Cognition* 44(3): 653–661.

Broersma, M., & Cutler, A. 2008. Phantom word activation in L2. *System* 36(1): 22–34.

de Bruijn, E. R. A., Dijkstra, T., Chwilla, D. J., & Schriefers, H. 2001. Language context effects on interlingual homograph recognition: Evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition. Special Issue: The cognitive neuroscience of bilingualism* 4(2): 155–168.

Brysbaert, M., & Duyck, W. 2010. Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition* 13(3): 359–371.

Brysbaert, M., van Dyck, G., & van de Poel, M. 1999. Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance* 25(1): 137–148.

Caramazza, A., & Brones, I. 1979. Lexical access in bilinguals. *Bulletin of the Psychonomic Society* 13(4): 212–214.

Carr, T. H., & Pollatsek, A. 1985. Recognizing printed words: A look at current models. In Derek Besner, T. Gary Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice*, 1–82. San Diego, CA: Academic Press.

Carreiras, M., Alvarez, C. J., & Devega, M. 1993. Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language* 32(6): 766–780.

Cattell, J. M. 1886. The time it takes to see and name objects. *Mind* 11(41): 63–65.

Chambers, C. G., & Cooke, H. 2009. Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35(4): 1029–1040.

Chauncey, K., Holcomb, P. J., & Grainger, J. 2008. Effects of stimulus font and size on masked repetition priming: An event-related potentials (ERP) investigation. *Language and Cognitive Processes* 23(1): 183–200.

Cheung, H., & Chen, H.-C. 1998. Lexical and conceptual processing in Chinese-English bilinguals: Further evidence for asymmetry. *Memory & Cognition* 26(5): 1002–1013.

Cohen, J. 1983. The cost of dichotomization. *Applied Psychological Measurement* 7(3): 249 – 253.

Coltheart, M. 1980. Reading, phonological recoding, and deep dyslexia. In Max Coltheart, John C. Marshall, & Karalyn E. Patterson (Eds.), *Deep dyslexia*, 197–226. London: Routledge & Kegan Paul.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. 1977. Access to the internal lexicon. In Stanislav Dornič (Ed.), *Attention and performance VI: Proceedings of the sixth international symposium on attention and performance*, 535–555. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. 2001. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 108(1): 204–256.

Coltheart, V., Patterson, K., & Leahy, J. 1994. When a ROWS is a ROSE: Phonological effects in written word comprehension. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 47(4): 917–955.

Conrad, M., Grainger, J., & Jacobs, A. M. 2007. Phonology as the source of syllable frequency effects in visual word recognition: Evidence from French. *Memory & Cognition* 35(5): 974–983.

Cutler, A. 1981. Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990? *Cognition* 10(1-3): 65 – 70.

Daneman, M., & Reingold, E. 1993. What eye fixations tell us about phonological recoding during reading. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale* 47(2): 153–178.

Daneman, M., & Stainton, M. 1991. Phonological recoding in silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17(4): 618–632.

Davis, C. J. 1999. *The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition*. Unpublished doctoral dissertation. Sydney, Australia: University of New South Wales.

Davis, C. J., Perea, M., & Acha, J. 2009. Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology: Human Perception and Performance* 35(5): 1550–1570.

Degani, T., & Tokowicz, N. 2010. Semantic ambiguity within and across languages: An integrative review. *The Quarterly Journal of Experimental Psychology* 63(7): 1266–1303.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. 2005. The neural code for written words: A proposal. *Trends in Cognitive Sciences* 9(7): 335–341.

Desmet, T., & Duyck, W. 2007. Bilingual language processing. *Language and Linguistics Compass* 1(3): 168–194.

Devadiga, D. N., & Bhat, J. S. 2011. *An investigation into bilingual cognitive processing*. Paper presented at the Annual International Conference on Languages, Literature and Linguistics, 2011, Singapore.

Dijkstra, T. 2003. Lexical processing in bilinguals and multilinguals. In Jasone Cenoz, Britta Hufeisen, & Ulrike Jessner (Eds.), *The multilingual lexicon*, 11–26. Dordrecht: Kluwer.

Dijkstra, T. 2005. Bilingual visual word recognition and lexical access. In Judith F. Kroll & Annette M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 179–201. Oxford: Oxford University Press.

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(4): 496–518.

Dijkstra, T., & van Heuven, W. J. B. 1998. The BIA model and bilingual word recognition. In Jonathan Grainger & Arthur M. Jacobs (Eds.), *Localist connectionist approaches to human cognition*, 189–225. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Dijkstra, T., & van Heuven, W. J. B. 2002b. Modeling bilingual word recognition: Past, present and future: Reply. *Bilingualism: Language and Cognition* 5(3): 219–224.

Dijkstra, T., van Jaarsveld, H., & ten Brinke, S. 1998. Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition* 1(1): 51–66.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, R. H. 2010. How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language* 62(3): 284–301.

Dijkstra, T., del Prado Martín, F. M., Schulpen, B. J. H., Schreuder, R., & Baayen, R. H. 2005. A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes* 20(1): 7–41.

Dijkstra, T., & De Smedt, K. (Eds.) 1996. *Computational psycholinguistics: AI and connectionist models of human language processing*, London, UK: Taylor & Francis.

Dijkstra, T., Timmermans, M., & Schriefers, H. 2000. On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language* 42(4): 445–464.

Doctor, E. A., & Klein, D. 1992. Phonological processing in bilingual word recognition. In Richard Jackson Harris (Ed.), *Cognitive processing in bilinguals*, 237–252. Amsterdam: North-Holland.

Dufau, S., Stevens, M., & Grainger, J. 2008. Windows executable software for the progressive demasking task (PDM). *Behavior Research Methods, Instruments, & Computers* 40(1): 33–37.

Duffy, S. A., Kambe, G., & Rayner, K. 2001. The effect of prior disambiguating context on the comprehension of ambiguous words: Evidence from eye movements. In David S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity*, 27–43. Washington, DC: American Psychological Association.

Dufour, R., & Kroll, J. F. 1995. Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition* 23(2): 166–180.

Duyck, W. 2005. Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(6): 1340–1359.

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(4): 663–679.

Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. 2004. WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments & Computers. Special Issue: Web-based archive of norms, stimuli, and data: Part 1* 36(3): 488–499.

Duyck, W., Diependaele, K., Drieghe, D., & Brysbaert, M. 2004. The size of the cross-lingual masked phonological priming effect does not depend on second language proficiency. *Experimental Psychology* 51(2): 116–124.

Elston-Güttler, K. E., & Friederici, A. D. 2005. Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language* 52(2): 256–283.

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(1): 57–70.

Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. 2005. Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience* 17(10): 1593–1610.

Ferrand, L., & Grainger, J. 1992. Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 45(3): 353–372.

Forster, K. I. 1976. Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*, 257–287. Amsterdam: North-Holland.

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. 1987. Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 39(2): 211–251.

Forster, K. I., Witzel, N., & Qiao, X. 2010. *Text orientation and transposed letter priming*. Paper presented at the 7th International Conference on the Mental Lexicon, 2010, Windsor, Canada.

Frenck-Mestre, C., & Prince, P. 1997. Second language autonomy. *Journal of Memory and Language* 37(4): 481–501.

Frost, R. 1998. Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin* 123(1): 71–99.

Frost, R. 2005. Orthographic systems and skilled word recognition processes in reading. In Margaret J. Snowling & Charles Hulme (Eds.), *The science of reading: A handbook*, 272–295. Oxford: Wiley-Blackwell.

Frost, R., & Ziegler, J. C. 2007. Speech and spelling interaction: The interdependence of visual and auditory word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*, 107–118. Oxford: Oxford University Press.

Gerard, L. D., & Scarborough, D. L. 1989. Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(2): 305–315.

Gernsbacher, M. A. 1993. Less skilled readers have less efficient suppression mechanisms. *Psychological Science* 4(5): 294–298.

Gernsbacher, M. A. 1997. Attentuating interference during comprehension: The role of suppression. In Douglas L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory*, 85–104. San Diego, CA: Academic Press.

Gernsbacher, M. A., & St. John, M. F. 2001. Modeling suppression in lexical access. In David S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Decade of behavior*, 47–65. Washington, DC: American Psychological Association.

Glushko, R. J. 1979. The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance* 5(4): 674–691.

Gollan, T. H., Forster, K. I., & Frost, R. 1997. Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(5): 1122–1139.

Goswami, U., & Ziegler, J. C. 2006. A developmental perspective on the neural code for written words. *Trends in Cognitive Sciences* 10(4): 142–143.

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. 2001. Pseudohomophone effects and phonological recoding procedures in reading development in English and German. *Journal of Memory and Language* 45(4): 648–664.

Grainger, J. 1993. Visual word recognition in bilinguals. In Robert Schreuder & Bert Weltens (Eds.), *The bilingual lexicon*, 11–25. Amsterdam: John Benjamins Publishing Company.

Grainger, J. 2008. Cracking the orthographic code: An introduction. *Language and Cognitive Processes* 23(1): 1–35.

Grainger, J., & Beauvillain, C. 1987. Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 39(2): 295–319.

Grainger, J., & Dijkstra, T. 1992. On the representation and use of language information in bilinguals. In Richard Jackson Harris (Ed.), *Cognitive processing in bilinguals*, 207–220. Amsterdam: North-Holland.

Grainger, J., & Dijkstra, T. 1996. Visual word recognition: Models and experiments. In Ton Dijkstra & Koenraad De Smedt (Eds.), *Computational psycholinguistics: AI and connectionist models of human language processing*, 139–165. London: Taylor & Francis.

Grainger, J., & van Heuven, W. J. B. 2003. Modeling letter position coding in printed word perception. In Patrick Bonin (Ed.), *Mental lexicon: Some words to talk about words*, 1–24. New York: Nova Science Publishers.

Grainger, J., & Jacobs, A. M. 1996. Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review* 103(3): 518–565.

Grainger, J., Kiyonaga, K., & Holcomb, P. J. 2006. The time course of orthographic and phonological code activation. *Psychological Science* 17(12): 1021 –1026.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. 2005. Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 58(6): 981–998.

Grainger, J., & Segui, J. 1990. Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics* 47(2): 191–198.

Grainger, J., & Ziegler, J. C. 2007. Cross-code consistency in a functional architecture for word recognition. In Elena L. Grigorenko & Adam J. Naples (Eds.), *Single-word reading: Biological and behavioral perspectives*, 129–157. Hillsdale, NJ: Lawrence Erlbaum Associates.

Green, D. W. 1998a. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1(2): 67–81.

Green, D. W. 1998b. Schemas, tags and inhibition. *Bilingualism: Language and Cognition* 1(2): 100–104.

Green, D. W. 2002. The bilingual as an adaptive system: Comment. *Bilingualism: Language and Cognition* 5(3): 206–208.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24(1): 187–203.

de Groot, A. M. B. 1992. Bilingual lexical representation: A closer look at conceptual representations. In Ram Frost & Leonard Katz (Eds.), *Orthography, phonology, morphology, and meaning*, 389–412. Amsterdam: North-Holland.

de Groot, A. M. B. 2011. *Language and cognition in bilinguals and multilinguals: An introduction*, New York: Psychology Press.

de Groot, A. M. B., Delmaar, P., & Lupker, S. J. 2000. The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 53(2): 397–428.

Grosjean, F. 1992. Another view of bilingualism. In Richard Jackson Harris (Ed.), *Cognitive processing in bilinguals*, 51–61. Amsterdam: North-Holland.

Grosjean, F. 1997. Processing mixed language: Issues, findings and models. In Annette M. B. de Groot & Judith F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, 225–254. Hillsdale, NJ: Lawrence Erlbaum Associates.

Grosjean, F. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1(2): 131–149.

Grosjean, F. 2001. The bilingual's language modes. In Janet L. Nicol (Ed.), *One mind, two languages: Bilingual language processing*, 1–22. Malden, MA: Blackwell.

Haigh, C. A., & Jared, D. 2007. The activation of phonological representations by bilinguals while reading silently: Evidence from interlingual homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(4): 623–644.

Hay, J., Drager, K., & Warren, P. 2009. Careful who you talk to: An effect of experimenter identity on the production of the NEAR/SQUARE merger in New Zealand English. *Australian Journal of Linguistics* 29(2): 269–285.

van Hell, J. G., & Dijkstra, T. 2002. Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review* 9(4): 780–789.

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 Heuven, W. J. B., & Dijkstra, T. 2001. *The semantic, orthographic, and phonological interactive activation model*. Paper presented at the 12th Conference of the European Society for Cognitive Psychology, 2001, Edinburgh, UK.

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. 1998. Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language* 39(3): 458–483.

van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. 2008. Language conflict in the bilingual brain. *Cerebral Cortex* 18(11): 2706–2716.

Holcomb, P. J., & Grainger, J. 2006. On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience* 18(10): 1631–1643.

Holmes, V. M., & Ng, E. 1993. Word-specific knowledge, word-recognition strategies, and spelling ability. *Journal of Memory and Language* 32(2): 230–257.

Humphreys, G. W., Evett, L. J., & Quinlan, P. T. 1990. Orthographic processing in visual word identification. *Cognitive Psychology* 22(4): 517–560.

Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. 1998. MROM-p: An interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition. In Jonathan Grainger & Arthur M. Jacobs (Eds.), *Localist connectionist approaches to human cognition*, 147–188. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jaeger, T. F. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4): 434–446.

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(1): 2–31.

Jared, D., Levy, B. A., & Rayner, K. 1999. The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General* 128(3): 219–264.

Jared, D., & Seidenberg, M. S. 1991. Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General* 120(4): 358–394.

Jared, D., & Szucs, C. 2002. Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition* 5(3): 225–239.

Katz, L., & Feldman, L. B. 1983. Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9(1): 157–166.

Katz, L., & Frost, R. 1992. The reading process is different for different orthographies: The orthographic depth hypothesis. In Ram Frost & Leonard Katz (Eds.), *Orthography, phonology, morphology, and meaning*, 67–84. Amsterdam: North-Holland.

van Kesteren, R., Dijkstra, T., & de Smedt, K. 2012. Markedness effects in Norwegian–English bilinguals: Task-dependent use of language-specific letters and bigrams. *The Quarterly Journal of Experimental Psychology* Online First Publication: 1–26.

Kim, J., & Davis, C. 2003. Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language* 49(4): 484–499.

Kolers, P. A. 1966a. Reading and talking bilingually. *The American Journal of Psychology* 79(3): 357–376.

Kolers, P. A. 1966b. Interlingual facilitation of short-term memory. *Journal of Verbal Learning and Verbal Behavior* 5(3): 314–319.

Kondrak, G. 2002. *Algorithms for language reconstruction*. Unpublished doctoral dissertation. Canada: University of Toronto.

Kotz, S. A., & Elston-Güttler, K. E. 2004. The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics* 17(2-3): 215–235.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. 2006. Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition* 9(02): 119–135.

Kroll, J. F., & Dijkstra, T. 2002. The bilingual lexicon. In Robert B. Kaplan (Ed.), *The Oxford handbook of applied linguistics*, 301–321. Oxford: Oxford University Press.

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. 2010. The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition* 13(3): 373–381.

Kroll, J. F., & Stewart, E. 1994. Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language* 33(2): 149–174.

Kroll, J. F., & Tokowicz, N. 2005. Models of bilingual representation and processing: Looking back and to the future. In Judith F. Kroll & Annette M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 531–553. Oxford: Oxford University Press.

Kuperman, V., Bertram, R., & Baayen, R. H. 2008. Morphological dynamics in compound processing. *Language and Cognitive Processes* 23(7-8): 1089–1132.

Lacruz, I., & Folk, J. R. 2004. Feedforward and feedback consistency effects for high- and low-frequency words in lexical decision and naming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 57(7): 1261–1284.

Legge, G. E., & Bigelow, C. A. 2011. Does print size matter for reading? A review of findings from vision science and typography. *Journal of Vision* 11(5): 1–22.

Lemhöfer, K., & Broersma, M. 2011. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* 44(2): 325–343.

Lemhöfer, K., & Dijkstra, T. 2004. Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition* 32(4): 533–550.

Lemhöfer, K., Dijkstra, T., & Michel, M. C. 2004. Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and Cognitive Processes* 19(5): 585–611.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. 2008. Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(1): 12–31.

Lemhöfer, K., & Radach, R. 2009. Task context effects in bilingual nonword processing. *Experimental Psychology* 56(1): 41–47.

Lété, B., & Pynte, J. 2003. Word-shape and word-lexical-frequency effects in lexical-decision and naming tasks. *Visual Cognition* 10(8): 913–948.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. 2009. Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences* 106(50): 21086–21090.

Li, P., Sepanski, S., & Zhao, X. 2006. Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods* 38(2): 202–210.

Li, P., & Yip, M. C. 1998. Context effects and the processing of spoken homophones. *Reading and Writing: An Interdisciplinary Journal* 10(3-5): 223–243.

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(2): 381–390.

Los, S. A. 1999. Identifying stimuli of different perceptual categories in mixed blocks of trials: Evidence for cost in switching between computational processes. *Journal of Experimental Psychology: Human Perception and Performance* 25(1): 3–23.

Luce, R. D. 1986. *Response times: Their role in inferring elementary mental organization*, Oxford: Oxford University Press.

Lukatela, G., & Turvey, M. T. 1994a. Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General* 123(4): 331–353.

Lukatela, G., & Turvey, M. T. 1994b. Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General* 123(2): 107–128.

Lupker, S. J. 2005. Visual word recognition. In Margaret J. Snowling & Charles Hulme (Eds.), *The science of reading: A handbook*, 39–60. Oxford: Blackwell.

Lupker, S. J., Kinoshita, S., Coltheart, M., & Taylor, T. E. 2003. Mixing costs and mixing benefits in naming words, pictures, and sums. *Journal of Memory and Language* 49(4): 556–575.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. 2002. On the practice of dichotomization of quantitative variables. *Psychological Methods* 7(1): 19–40.

Macizo, P., Bajo, T., & Cruz Martín, M. 2010. Inhibitory processes in bilingual language comprehension: Evidence from Spanish-English interlexical homographs. *Journal of Memory and Language* 63(2): 232–244.

Macnamara, J. 1967. The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior* 6(5): 729–736.

Martensen, H., Dijkstra, T., & Maris, E. 2005. A werd is not quite a word: On the role of sublexical phonological information in visual lexical decision. *Language and Cognitive Processes* 20(4): 513–552.

Martensen, H., Maris, E., & Dijkstra, T. 2003. Phonological ambiguity and context sensitivity: On sublexical clustering in visual word recognition. *Journal of Memory and Language* 49(3): 375–395.

Mason, M. 1975. Reading ability and letter search time: Effects of orthographic structure defined by single-letter positional frequency. *Journal of Experimental Psychology: General* 104(2): 146–166.

Mathey, S., & Zagar, D. 2000. The neighborhood distribution effect in visual word recognition: Words with single and twin neighbors. *Journal of Experimental Psychology: Human Perception and Performance* 26(1): 184–205.

McClelland, J. L. 1976. Preliminary letter identification in the perception of words and nonwords. *Journal of Experimental Psychology: Human Perception and Performance* 2(1): 80–91.

McClelland, J. L., & Johnston, J. C. 1977. The role of familiar units in perception of words and nonwords. *Perception & Psychophysics* 22(3): 249–261.

McClelland, J. L., & Rumelhart, D. E. 1981. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88(5): 375–407.

Meara, P. M. 1996. *English vocabulary tests: 10k*. Swansea, UK: Center for Applied Language Studies.

van der Meij, M., Cuetos, F., Carreiras, M., & Barber, H. A. 2011. Electrophysiological correlates of language switching in second language learners. *Psychophysiology* 48(1): 44–54.

Meuter, R. F. I. 2005. Language selection in bilinguals: Mechanisms and processes. In Judith F. Kroll & Annette M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 349–370. Oxford: Oxford University Press.

Midgley, K. J., Holcomb, P. J., & Grainger, J. 2009. Language effects in second language learners and proficient bilinguals investigated with event-related potentials. *Journal of Neurolinguistics* 22(3): 281–300.

Midgley, K. J., Holcomb, P. J., & Grainger, J. 2010. Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience* 23(7): 1634–1647.

Mitchell, R., & Myles, F. 2004. *Second language learning theories* 2nd ed., London: Arnold.

Morton, J., & Patterson, K. E. 1980. A new attempt at an interpretation, or, an attempt at a new interpretation. In Max Coltheart, John C. Marshall, & Karalyn E. Patterson (Eds.), *Deep dyslexia*, 91–118. London: Routledge & Kegan Paul.

Mulloni, A. 2007. *Automatic prediction of cognate orthography using support vector machines*. Paper presented at the 45th Annual Meeting of the Association for Computational Linguistics, 2007, Prague, The Czech Republic.

Nas, G. 1983. Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning & Verbal Behavior* 22(5): 526–534.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. 2006. Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin & Review* 13(1): 45–52.

Norris, D., McQueen, J. M., & Cutler, A. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23(03): 299–325.

van Orden, G. C. 1987. A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition* 15(3): 181–198.

van Orden, G. C., & Goldinger, S. D. 1994. Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance* 20(6): 1269–1291.

van Orden, G. C., & Kloos, H. 2005. The question of phonology and reading. In Margaret J. Snowling & Charles Hulme (Eds.), *The science of reading: A handbook*, 61–78. Oxford: Blackwell.

van Orden, G. C., Pennington, B. F., & Stone, G. O. 1990. Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review* 97(4): 488–522.

Orfanidou, E., & Sumner, P. 2005. Language switching and the effects of orthographic specificity and response repetition. *Memory & Cognition* 33(2): 355–369.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. 2009. The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition* 111(2): 263–269.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. 2010. Is a FAN always FUN? Phonological and orthographic effects in bilingual visual word recognition. *Language and Speech* 53(3): 383–403.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. 1982. An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review* 89(5): 573–594.

Paap, K. R., Newsome, S. L., & Noel, R. W. 1984. Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance* 10(3): 413–428.

Paap, K. R., & Noel, R. W. 1991. Dual-route models of print to sound: Still a good horse race. *Psychological Research/Psychologische Forschung* 53(1): 13–24.

Patterson, K. E., & Coltheart, V. 1987. Phonological processes in reading: A tutorial review. In Max Coltheart (Ed.), *Attention and performance XII: The psychology of reading*, 421–447. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pennington, B. F., Lefly, D. L., van Orden, G. C., Bookman, M. O., & Smith, S. D. 1987. Is phonology bypassed in normal or dyslexic development? *Annals of Dyslexia* 37(1): 62–89.

Perea, M., & Lupker, S. J. 2003. Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition* 31(6): 829–841.

Perea, M., & Rosa, E. 2002. Does 'whole-word shape' play a role in visual word recognition? *Perception & Psychophysics* 64(5): 785–794.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. 1988. Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language* 27(1): 59–70.

Perlmann-Balme, M., & Kiefer, P. 2004. *Start Deutsch. Deutschprüfungen für Erwachsene. A1-A2. Prüfungsziele, Testbeschreibung*, München, Frankfurt: Goethe-Institut und WBT.

Perry, C. 2003. A phoneme-grapheme feedback consistency effect. *Psychonomic Bulletin & Review* 10(2): 392–397.

Perry, C., & Ziegler, J. C. 2002. Cross-language computational investigation of the length effect in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance* 28(4): 990–1001.

Petit, J.-P., Midgley, K. J., Holcomb, P. J., & Grainger, J. 2006. On the time course of letter perception: A masked priming ERP investigation. *Psychonomic Bulletin & Review* 13(4): 674–681.

Pexman, P. M., Lupker, S. J., & Jared, D. 2001. Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(1): 139–156.

Pinheiro, J. C., & Bates, D. M. 2000. *Mixed effects models in S and S-Plus*, New York: Springer.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. 1996. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 103(1): 56–115.

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. 2012. Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance* Online First Publication: 1–21.

Procter, P. (Ed.) 1999. *Cambridge international dictionary of English*, New York: Cambridge University Press.

R Development Core Team 2006. *R: A language and environment for statistical computing*, Vienna: R Foundation for Statistical Computing.

Rapp, B., Folk, J. R., & Tainturier, M.-J. 2001. Word reading. In Brenda Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind*, 233–262. Philadelphia, PA, USA: Psychology Press.

Rastle, K., Davis, M. H., & New, B. 2004. The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review* 11(6): 1090–1098.

Ratcliff, R. 1993. Methods for dealing with reaction time outliers. *Psychological Bulletin* 114(3): 510–532.

Raz, A., Moreno-Iñiguez, M., Martin, L., & Zhu, H. 2007. Suggestion overrides the Stroop effect in highly hypnotizable individuals. *Consciousness and Cognition* 16(2): 331–338.

Reicher, G. M. 1969. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology* 81(2): 275–280.

Rey, A., Ziegler, J. C., & Jacobs, A. M. 2000. Graphemes are perceptual reading units. *Cognition* 75(1): B1–B12.

Rodriguez-Fornells, A., Rotte, M., Heinze, H.-J., Nösselt, T., & Münte, T. F. 2002. Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature* 415(6875): 1026–1029.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. 1971. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning & Verbal Behavior* 10(6): 645–657.

Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., & Orban, G. A. 2005. Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *The Journal of Neuroscience* 25(17): 4294–4306.

Scarborough, D. L., Gerard, L. D., & Cortese, C. 1984. Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior* 23(1): 84–99.

Schepens, J., Dijkstra, T., & Grootjen, F. 2012. Distributions of cognates in Europe as based on Levenshtein Distance. *Bilingualism: Language and Cognition* 15(1): 157–166.

Schneider, W., & Chein, J. M. 2003. Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science* 27(3): 525–559.

Schneider, W., Eschman, A., & Zuccolotto, A. 2002a. *E-Prime reference guide*, Pittsburgh: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. 2002b. *E-Prime user's guide*, Pittsburgh: Psychology Software Tools Inc.

Schulpen, B. J. H. 2003. *Explorations in bilingual word recognition: Cross-modal, cross-sectional, and cross-language effects*. Unpublished doctoral dissertation. The Netherlands: University of Nijmegen.

Schulpen, B. J. H., Dijkstra, T., Schriefers, H., & Hasper, M. 2003. Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 29(6): 1155–1178.

Schwartz, A. I., & Kroll, J. F. 2006. Bilingual lexical activation in sentence context. *Journal of Memory and Language* 55(2): 197–212.

Schwartz, A. I., Kroll, J. F., & Diaz, M. 2000. *Reading Spanish words with English word bodies: Activation of spelling-to-sound correspondences across languages*. Paper

presented at the 2nd International Conference on the Mental Lexicon, 2000, Montreal, Canada.

Schwartz, A. I., Kroll, J. F., & Diaz, M. 2007. Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes* 22(1): 106–129.

Segalowitz, N. S., & Segalowitz, S. J. 1993. Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics* 14(3): 369–385.

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. 1998. Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics* 19(1): 53–67.

Seidenberg, M. S. 1985. The time course of phonological code activation in two writing systems. *Cognition* 19(1): 1–30.

Seidenberg, M. S. 1987. Sublexical structures in visual word recognition: Access units or orthographic redundancy? In Max Coltheart (Ed.), *Attention and performance XII: The psychology of reading*, 245–263. Hillsdale, NJ: Lawrence Erlbaum Associates.

Seidenberg, M. S., & McClelland, J. L. 1989. A distributed, developmental model of word recognition and naming. *Psychological Review* 96(4): 523–568.

Share, D. L. 2008. On the anglocentricities of current reading research and practice: The perils of overreliance on an 'outlier' orthography. *Psychological Bulletin* 134(4): 584–615.

Siakaluk, P. D., Sears, C. R., & Lupker, S. J. 2002. Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception and Performance* 28(3): 661–681.

Smith, F. 1969. Familiarity of configuration vs. discriminability of features in the visual identification of words. *Psychonomic Science* 14(6): 261–263.

Smits, E., Martensen, H., Dijkstra, T., & Sandra, D. 2006. Naming interlingual homographs: Variable competition and the role of the decision system. *Bilingualism: Language and Cognition* 9(3): 281–297.

Smits, E., Sandra, D., Martensen, H., & Dijkstra, T. 2009. Phonological inconsistency in word naming: Determinants of the interference effect between languages. *Bilingualism: Language and Cognition* 12(1): 23–39.

Soares, C., & Grosjean, F. 1984. Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory & Cognition* 12(4): 380–386.

Stone, G. O., Vanhoy, M., & van Orden, G. C. 1997. Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language* 36(3): 337–359.

Stroop, J. R. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18(6): 643–662.

von Studnitz, R. E., & Green, D. W. 1997. Lexical decision and language switching. *International Journal of Bilingualism* 1(1): 3 –24.

von Studnitz, R. E., & Green, D. W. 2002a. Interlingual homograph interference in German-English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition* 5(1): 1–23.

von Studnitz, R. E., & Green, D. W. 2002b. The cost of switching language in a semantic categorization task. *Bilingualism: Language and Cognition* 5(3): 241–251.

Sunderman, G., & Kroll, J. F. 2006. First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition* 28(3): 387–422.

Sunderman, G., & Schwartz, A. I. 2008. Using cognates to investigate cross-language competition in second language processing. *TESOL Quarterly* 42(3): 527–536.

Taft, M. 1979. Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior* 18(1): 21–39.

Taft, M., & Forster, K. I. 1976. Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior* 15(6): 607–620.

Thomas, M. S. C., & Allport, A. 2000. Language switching costs in bilingual visual word recognition. *Journal of Memory and Language* 43(1): 44–66.

Thomas, M. S. C., & van Heuven, W. J. B. 2005. Computational models of bilingual comprehension. In Judith F. Kroll & Annette M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 202–225. Oxford: Oxford University Press.

Tinker, M. A. 1963. *Legibility of print*, Ames, IA: Iowa State University Press.

Titone, D. A., Libben, M. R., Mercier, J., Whitford, V., & Pivneva, I. 2011. Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(6): 1412–1431.

Tzeglov, J. 1999. Automaticity and processing without awareness. *Psyche* 5. Retrieved from http://www.theassc.org/vol_5_1999 on 22 November 2011.

Tzeglov, J., & Kadosh, R. C. 2009. From automaticity to control in bilinguals. *Trends in Cognitive Sciences* 13(11): 455.

Tzelgov, J., Henik, A., & Leiser, D. 1990. Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16(5): 760–771.

de Vaan, L., Schreuder, R., & Baayen, R. H. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon* 2(1): 1–24.

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. 2010. Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology* 102(4): 827–842.

Vaid, J., & Frenck-Mestre, C. 2002. Do orthographic cues aid language recognition? A laterality study with French-English bilinguals. *Brain and Language* 82(1): 47–53.

Vandeberg, L., Guadalupe, T., & Zwaan, R. A. 2011. How verbs can activate things: Cross-language activation across word classes. *Acta Psychologica* 138(1): 68–73.

Weber, R.-M. 1970. A linguistic analysis of first-grade reading errors. *Reading Research Quarterly* 5(3): 427–451.

Weiner, I. B., Freedheim, D. K., Schinka, J. A., & Velicer, W. F. 2003. *Handbook of psychology: Research methods in psychology*, New York: John Wiley and Sons.

Wheeler, D. D. 1970. Processes in word recognition. *Cognitive Psychology* 1(1): 59–85.

Whitney, C. 2001. How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review* 8(2): 221–243.

van Wijnendaele, I., & Brysbaert, M. 2002. Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance* 28(3): 616–627.

Wu, Y. J., & Thierry, G. 2010. Investigating bilingual processing: The neglected role of language processing contexts. *Frontiers in Psychology* 1: Article 178.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. 2010. Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science: A Journal of the American Psychological Society* 21(4): 551–559.

Ziegler, J. C., & Goswami, U. 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin* 131(1): 3–29.

Ziegler, J. C., & Goswami, U. 2006. Becoming literate in different languages: Similar problems, different solutions. *Developmental Science* 9(5): 429–453.

Ziegler, J. C., Grainger, J., & Brysbaert, M. 2010. Modelling word recognition and reading aloud. *European Journal of Cognitive Psychology* 22(5): 641.

Ziegler, J. C., Montant, M., & Jacobs, A. M. 1997. The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language* 37(4): 533–554.

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. 2001. Identical words are read differently in different languages. *Psychological Science* 12(5): 379–384.

Ziegler, J. C., Petrova, A., & Ferrand, L. 2008. Feedback consistency effects in visual and auditory word recognition: Where do we stand after more than a decade? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34(3): 643–661.

Zorzi, M., Houghton, G., & Butterworth, B. 1998. Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance* 24(4): 1131–1161.

Appendix A

MODELS OF BILINGUAL VISUAL WORD RECOGNITION

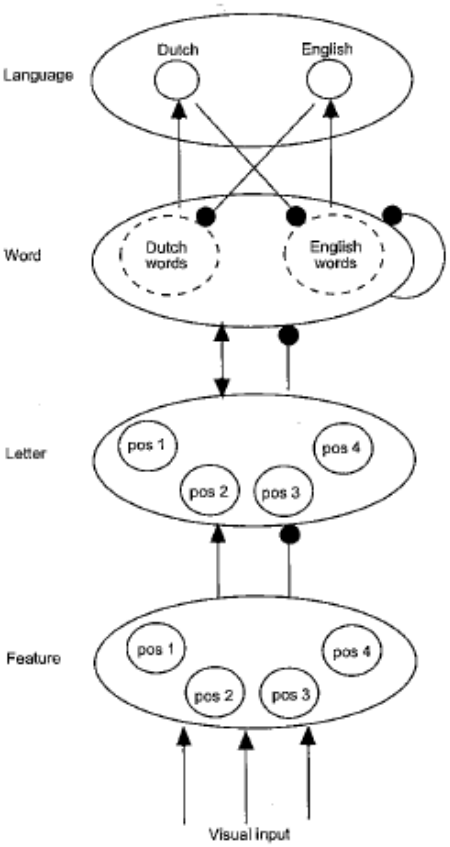


Figure 105. Bilingual Interactive Activation Model (Dijkstra & van Heuven 1998). Excitatory connections are indicated by arrows (with arrowheads pointing to the direction of activation spread), inhibitory connections by lines with closed circles. Pos, position

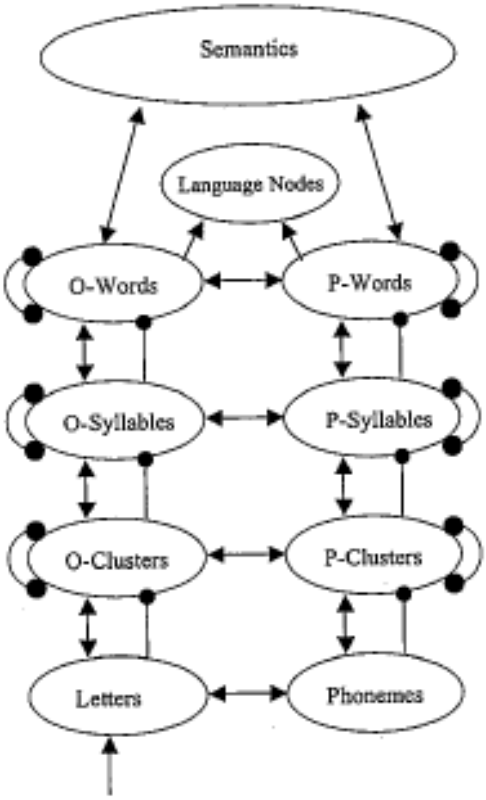


Figure 106. The Semantic, Orthographic, Phonological Interactive Activation Model (van Heuven & Dijkstra 2001). O, Orthography; P, Phonology

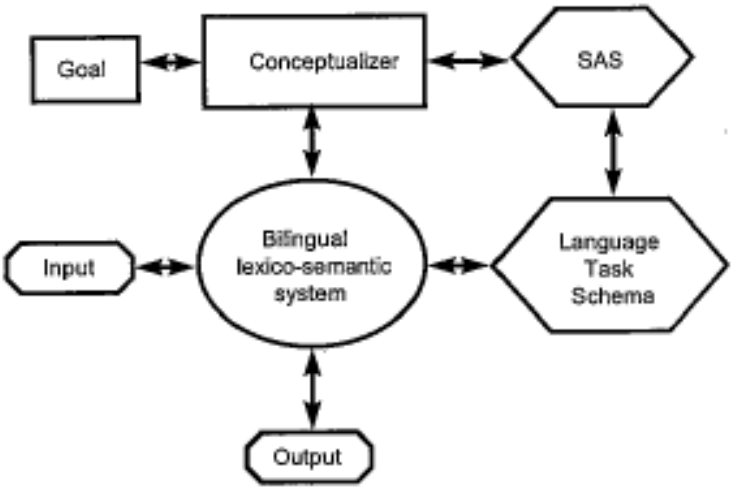


Figure 107 The Inhibitory Control Model (Green 1998a).
SAS, Supervisory Attentional System.

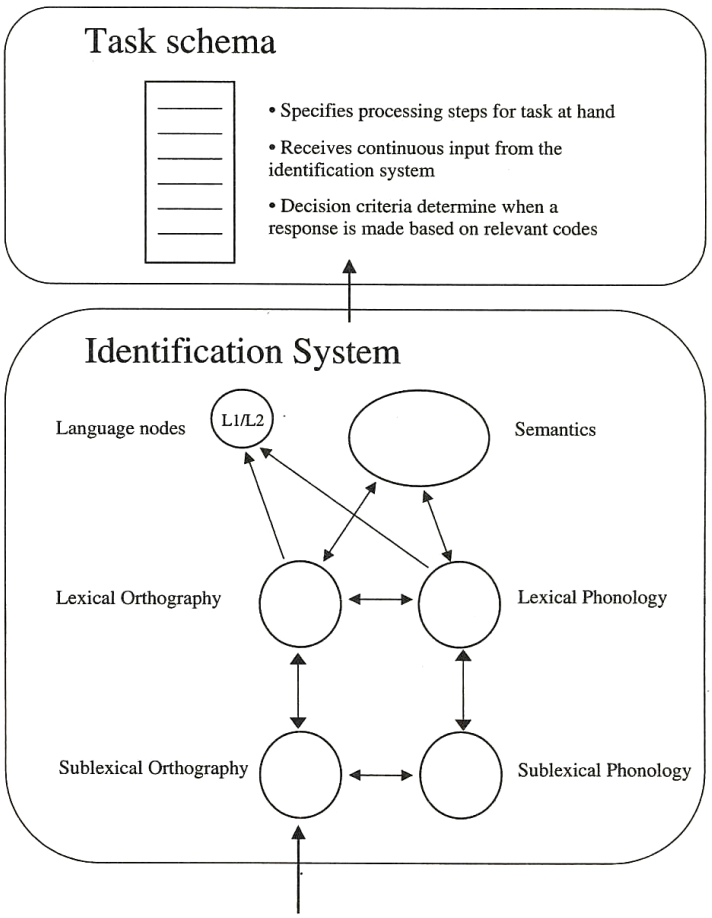


Figure 108. The architecture of the Bilingual Interactive Activation Plus (BIA+) model (Dijkstra & van Heuven 2002a).
L1, first language; L2, second language.

.

Appendix B

LANGUAGE PROFICIENCY ASSESSMENT

Language proficiency test

The following items used in the language proficiency test – in German or English, depending on a participant's first language – are adopted from Lemhöfer et al. (2004). The formula for computing Meara's Δ*M* (1996) is

$$\Delta M = \frac{(h-f)(1+h-f)}{h(1-f)} - 1 = \frac{h-f}{1-f} - \frac{f}{h}$$

where

h = percentage correctly recognised words (hit rate), and

f = percentage incorrectly accepted nonwords (false alarm rate).

	English	***German***
Words	ABLAZE	RUPPIG
	ALLIED	RASEND
	BEWITCH	SATTELN
	BREEDING	ZÜCHTUNG
	CARBOHYDRATE	DESTILLATION
	CELESTIAL	WAGHALSIG
	CENSORSHIP	SUMMIERUNG
	CLEANLINESS	SCHWACHHEIT
	CYLINDER	GEOGRAPH
	DISPATCH	ERBARMEN
	ELOQUENCE	KANNIBALE
	FESTIVITY	PENSIONAT
	FLAW	ZEHE
	FLUID	FEIGE
	FRAY	GAREN
	HASTY	MEHLIG
	HURRICANE	TURBULENZ
	INGENIOUS	SUBVERSIV
	LENGTHY	DÄMMRIG
	LISTLESS	UNSTETIG
	LOFTY	ZUGIG
	MAJESTIC	MONSTRÖS
	MOONLIT	KLAGLOS
	MUDDY	KLAMM
	NOURISHMENT	SPEICHERUNG
	PLAINTIVELY	EIMERWEISE
	RASCAL	ZIERDE
	RECIPIENT	KLEMPNER
	SAVOURY	STAKSIG
	SCHOLAR	LEUCHTE
	SCORNFUL	REUEVOLL
	SCREECH	ZAPFEN
	SHIN	MALZ
	SLAIN	FEIST
	STOUTLY	ZUOBERST
	TURMOIL	STRÄHNE
	TURTLE	FLINTE
	UNKEMPT	UNTIEF
	UPKEEP	ANPROBE
	WROUGHT	HERZIG
Nonwords	ABERGY	MALODIE
	ALBERATION	DEGERATION
	CRUMPER	TRACHTER
	DESTRIPTION	ENTSACHTUNG
	EXPRATE	AUSREBEN
	FELLICK	MACKEL
	INTERFATE	STOCKFEST
	KERMSHAW	PETURAT
	KILP	SCHEIL
	MAGRITY	SONITÄT
	MENSIBLE	WELSTBAR
	PLAUDATE	STALMEN
	PROOM	NARKE
	PUDOUR	FLISTOR
	PULSH	LUDAL
	PURRAGE	FAUNIK
	QUIRTY	DRAUNIG
	REBONDICATE	VERMASTIGEN
	SKAVE	PLANG
	SPAUNCH	FÜRREN

Language history questionnaire

The following questions were included in the English questionnaire (according to their order of presentation).

1. Age (in years):
2. Sex (please circle): Male / Female
3. Education (degree obtained or school level achieved):
4. (a). Country of origin:

   (b). Country of residence:
5. How long have you been in the country of your current residence? (in years)
6. What is your first language (if you grew up with more than one language, please specify)?
7. Do you speak other languages (please specify)?

   __YES _____
   _____
   _____

   __NO (if you answered NO, you do not need to continue filling out this form)
8. If you answered YES to question 7, please specify the age at which you started to learn these other languages in the following situations (specify language and write age below any situation that applies).

Language	At home	In school	After arriving in the country where the language is spoken
e.g. Spanish e.g. Dutch		12 yrs old	19 years old

9. List all languages you know in order of most proficient to least proficient. Rate your ability on the following aspects in each language. Please rate according to the following scale (write down the number in the table):

Very poor **Native-like**

1 2 3 4 5 6

Language	Reading proficiency	Writing proficiency	Speaking fluency	Listening ability

10. Provide the age at which you were first exposed to each language in terms of speaking, reading, and writing, and the number of years you have spent on learning each language.

Language	Age first exposed to the language			Number of years learning
	Speaking	Reading	Writing	

11. Do you have a foreign accent in the languages you speak? If so, please rate the strength of your accent according to the following scale (write down the number in the table):

No Accent **Very Strong**

1 2 3 4 5 6

Language	**Accent (circle one)**	**Strength**
	Y N	
	Y N	
	Y N	

12. How did you learn German (please circle all that apply)?

Through formal classroom instruction: Mainly Mostly Occasionally

Through interacting with people: Mainly Mostly Occasionally

A mixture of both, but More classroom More interaction Equally both

Other (specify: _____**)**

13. Thinking about your daily language use, what proportion of time do you use in each language (circle one that applies):

English: <25% 25% 50% 75% 100%
German: <25% 25% 50% 75% 100%
Other languages: <25% 25% 50% 75% 100%
(specify the languages: _____)

14. Estimate, in terms of hours per day, how often you are engaged in the following activities with all languages.

Activities	**English**	**German**	**Other Languages (specify _____)**
Listen to Radio/ Watching TV:	_____(hrs)	_____(hrs)	_____(hrs)
Reading for fun:	_____(hrs)	_____(hrs)	_____(hrs)
Reading for work:	_____(hrs)	_____(hrs)	_____(hrs)
Reading on the Internet:	_____(hrs)	_____(hrs)	_____(hrs)
Writing emails to friends:	_____(hrs)	_____(hrs)	_____(hrs)
Writing articles/papers:	_____(hrs)	_____(hrs)	_____(hrs)

15. Estimate, in terms of hours per day, how often you speak (or used to speak) English (L1) and German (L2) with the following people.

	Language	**Hours**
Father:	(L1) _____	_____ (hrs)
	(L2) _____	_____ (hrs)
Mother:	(L1) _____	_____ (hrs)
	(L2) _____	_____ (hrs)
Grandfather(s):	(L1) _____	_____ (hrs)
	(L2) _____	_____ (hrs)
Grandmother(s):	(L1) _____	_____ (hrs)
	(L2) _____	_____ (hrs)
Brother(s)/Sister(s):	(L1) _____	_____ (hrs)
	(L2) _____	_____ (hrs)
Other family members:	(L1) _____	_____ (hrs)
	(L2) _____	_____ (hrs)

16. Estimate, in terms of hours per day, how often you now speak English (L1) and German (L2) with the following people.

		Language	Hours
Spouse/partner:	(L1)	_____	_____ (hrs)
	(L2)	_____	_____ (hrs)
Friends:	(L1)	_____	_____ (hrs)
	(L2)	_____	_____ (hrs)
Classmates:	(L1)	_____	_____ (hrs)
	(L2)	_____	_____ (hrs)
Co-workers:	(L1)	_____	_____ (hrs)
	(L2)	_____	_____ (hrs)

17. Write down the name of the language in which you received instruction in school, for each schooling level:

Primary School (incl. Middle School): _____
Secondary School/College: _____
Tertiary Education/University: _____

18. In which languages do you usually:

Count, add, multiply, and do simple arithmetic? _____
Dream? _____
Express anger or affection? _____

19. When you are speaking, do you ever mix words or sentences from the two or more languages you know (if no, skip to question 21)?

20. List the languages that you mix and rate the frequency of mixing in normal conversation with the following people according to the following scale (write down the number in the table):

Rarely **Very Frequently**
1 2 3 4 5 6

Relationship	Languages mixed	Frequency of mixing
Spouse/family members		
Friends		
Co-workers		
Classmates		

21. In which language (English or German) do you feel you usually do better? Write the name of the language under each condition.

	At home	At work
Reading	_____	_____
Writing	_____	_____
Speaking	_____	_____
Understanding	_____	_____

22. Among the languages you know, which language is the one that you would prefer to use in these situations?

At home _____
At work _____
At a party _____
In general _____

23. If you have lived or travelled in other countries than your country of residence for more than three months, please indicate the name(s) of the country or countries, your length of stay, and the language(s) you learned or tried to learn.

24. If you have taken a standardized test of proficiency for languages other than your first language (e.g., TOEFL or Test of English as a Foreign Language), please indicate the scores you received for each.

Language	**Scores**	**Name of the Test**
_____	_____	_____
_____	_____	_____
_____	_____	_____

25. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.

26. Have you recently, i.e. today been in contact with the languages you know (please specify), e.g. course book, lecture, television, radio, foreign friends?

Appendix C

STUDY ONE

Stimulus material

Table 10.

Interlingual homographs used in both sessions of Experiments One and Two, including item characteristics. *Note*. Items marked with an asterisk were used in Elston-Güttler, Gunter, et al. (2005). *Note*. GF=German log frequency; EF=English log frequency; GB=German bigram frequency; EB=English bigram frequency; GN= German neighbour count; EN=English neighbour count; L=length (in letters).

Interlingual Homograph	GF	EF	GB	EB	GN	EN	L
ALTER*	1.72	1.60	29745	16228	4	4	5
BAD	2.19	2.52	2308	2346	5	20	3
BALD*	2.21	0.95	5545	5572	5	8	4
BITTEN*	2.60	2.00	31189	13407	5	4	6
BRIEF*	2.15	1.73	11063	5499	1	1	5
CHEF*	1.75	0.60	20177	4446	0	3	4
DOSE*	0.70	1.08	5907	4414	4	13	4
FASTEN*	0.00	1.04	33936	15581	7	2	6
GENIE	1.08	0.00	30729	8513	1	1	5
GIFT*	0.95	1.71	2541	1474	1	5	4
GUT*	3.14	1.00	1769	1947	3	16	3
HALL*	1.62	2.13	8575	8156	8	17	4
HANDY*	0.00	0.90	11592	8330	0	7	5
HELL	1.76	1.99	12468	6785	7	15	4
HERD*	1.23	1.34	24053	10021	5	9	4
HERB*	1.00	1.32	24590	9388	5	7	4
HOSE*	1.36	0.70	6782	5096	5	10	4
HUT*	1.52	1.53	2111	1810	4	17	3
KIND*	2.62	2.68	9400	10008	6	8	4
LIST*	1.04	2.06	15797	10372	5	11	4
MADE	0.30	3.00	6342	6157	5	13	4
MIST*	0.48	1.26	13036	8795	5	11	4
MODE*	1.30	1.30	4813	4664	5	11	4
MUSTER*	1.34	0.78	35866	18543	5	5	6
MUTTER*	2.31	1.42	29030	14470	4	6	6
NOTE	1.66	2.26	8389	6755	8	9	4
NUN*	2.92	1.04	7242	2316	3	13	3
RAT*	2.05	1.40	7280	8234	5	23	3
RATE	0.95	2.33	13964	12803	5	19	4
ROMAN*	1.77	1.69	9780	10687	0	2	5
SAGE*	0.78	0.70	8742	3414	9	15	4
SMOKING*	0.30	1.15	13027	12901	0	0	7
STERN*	1.85	1.15	33269	16499	3	1	5
STILL	1.74	2.97	15240	13746	3	6	5
STRAND*	1.40	1.20	22323	16261	1	0	6
TAG*	2.94	1.04	4567	3277	8	21	3
TASTE*	0.30	1.93	18363	12814	6	6	5
TELLER	1.04	0.30	37234	20847	5	2	6
TOLL	1.15	1.04	5365	6199	5	11	4

Table 11.

Two sets of control material used in Experiments One and Two, including item characteristics. *Note*. GF=German log frequency; GB=German bigram frequency; EB=English bigram frequency; GN= German neighbour count; EN=English neighbour count; Length is identical to that of IHGs.

Control One	GF	GB	EB	GN	EN	Control Two	GF	GB	EB	GN	EN
STEIN	1.78	27057	16190	4	3	KARTE	1.68	13515	9956	4	0
MUT	1.63	2196	1906	5	13	ARM	2.20	4304	4621	5	5
WAHR	2.15	4082	1362	5	0	DANK	2.02	6769	5399	5	10
STELLE	2.43	25318	17786	5	0	NENNEN	2.52	41532	16451	5	0
LIEBE	2.06	17011	5294	1	1	NATUR	2.08	8338	8966	1	0
KNIE	1.57	7619	2693	0	2	EWIG	1.62	5453	2131	0	2
MEHL	0.60	6705	2261	4	1	KAMM	0.90	3987	1627	4	0
KETTEN	0.30	31648	12682	7	1	ROCKEN	0.00	25024	10593	7	2
HERAB	1.26	29382	14076	1	0	BUSCH	1.11	25307	5786	1	2
EGAL	1.08	5918	5308	1	0	SEKT	1.08	5932	2584	1	3
BIS	3.24	5819	3979	3	9	LAS	3.27	3227	2755	3	8
TOTE	1.67	9099	7717	8	8	NASE	1.54	8115	5910	8	8
GRUND	2.64	13833	5985	0	3	PUNKT	2.43	9237	2986	0	0
HALS	1.64	6730	5774	7	5	HAUT	1.87	8244	3612	7	5
HEIM	1.42	14074	3231	5	2	HEER	1.26	23719	10333	5	12
KERN	1.51	23046	8818	5	4	EDEL	1.26	10174	7503	5	1
TANZ	1.45	9054	6608	5	2	TOPF	1.00	3036	2968	5	4
LOS	1.94	2938	2921	4	10	LOK	1.28	1786	2227	4	7
DORT	2.65	5229	5415	6	10	HOCH	2.90	16201	4402	6	2
FAUL	0.90	6391	2562	5	8	SOFA	0.85	2344	2279	5	3
LAMM	0.48	5708	3967	5	4	RAHM	0.00	6078	3709	5	0
OBER	0.30	24157	8665	5	3	OASE	0.30	6056	4811	5	5
KAUF	1.53	7049	857	5	0	MORD	1.62	4341	5323	5	7
BITTER	1.43	32180	16169	5	7	KOCHEN	1.20	38815	9198	5	0
FEHLEN	2.10	27993	9604	4	0	KOSTEN	2.18	33771	14209	4	0
SAND	1.49	10364	7473	8	9	BIER	1.66	25003	8932	8	4
MAI	2.54	2193	3210	3	11	UHR	2.87	1865	325	3	1
TOR	2.03	4172	5180	5	14	RAD	1.63	5156	4776	5	21
FARM	0.78	5589	5248	5	7	WEHR	1.00	6176	1093	5	2
STOFF	1.71	9099	7545	0	0	VOGEL	1.59	12778	4859	0	1
ROST	0.60	10024	7985	9	10	MAUS	1.00	9215	4893	9	3
FROSTIG	0.30	19110	14777	0	0	DRECKIG	0.00	15229	9909	0	0
FISCH	1.53	27340	7125	3	2	KUNDE	1.67	14959	7133	3	0
BRAUN	1.51	16131	6829	3	2	MIETE	1.52	16448	8822	3	0
NACHTS	1.56	21357	6205	1	0	TEUFEL	1.34	17168	7900	1	0
BAU	2.11	5814	1673	8	8	ALL	3.43	5730	6619	8	7
BEIGE	0.30	23285	4070	6	0	TANNE	0.30	11910	10578	6	1
KUCHEN	1.04	38675	8889	5	0	TRAUEN	1.18	32890	11040	5	0
ZOLL	1.30	4044	4311	5	7	MATT	1.18	6461	7591	5	8

Table 12.

Four sets of nonwords (NW) used in Experiments One and Two. Two sets were matched against the Interlingual homographs (NW one and NW two) and the other two were matched against controls (NW three and NW four). *Note*. GB=German bigram frequency; EB=English bigram frequency; GN= German neighbour count; EN=English neighbour count; Length was identical to that of IHGs.

NW one	GB	EB	GN	EN	NW two	GB	EB	GN	EN
TOLEN	24779	12267	4	1	LAFER	24915	10673	4	5
OHT	2417	571	5	5	SUT	1932	2403	5	18
GOSE	5848	4024	5	9	LORM	5386	5786	5	11
BITZEN	25899	8524	5	0	GIEBEN	31029	7336	5	0
STAPS	10770	7644	1	3	PREHM	11100	6638	1	0
FENO	20496	6100	0	1	LERO	24796	14234	0	2
WAUT	6840	3028	4	6	NARK	6371	5968	4	6
SALLEN	29174	16141	7	1	LACKER	28042	14011	7	4
ENGAN	29055	12636	1	0	GRENA	27382	11720	1	1
KOOM	2445	2785	1	5	TUGS	3248	1471	1	1
KUB	1483	517	3	8	MOG	1432	1878	3	19
ZAUS	7591	2980	8	0	TAUS	10478	5099	8	3
ANARO	12204	12725	0	0	DRELD	13361	8080	0	1
REIG	19082	6297	7	1	GETT	10115	4189	7	3
PERN	21400	9735	5	9	ERRE	26125	12650	5	1
REIR	16007	6238	5	4	HEIE	18790	3139	5	2
NATT	6557	7110	5	3	LALB	7052	6597	5	1
VAD	1030	1791	4	13	MEL	8087	4282	4	3
JABE	7126	2642	6	4	GEIM	15161	2878	6	3
STED	14496	11076	5	8	ARIN	11997	14418	5	5
BOBE	6070	2604	5	8	RORN	6139	6657	5	7
WEIA	11184	2297	5	1	ZIES	10281	4950	5	0
PALZ	4704	5519	5	3	BRUM	4951	2428	5	3
ZUNKER	30544	10614	5	1	DICKEL	14863	10025	5	2
ZENNER	41682	16093	4	1	NENDEN	45215	17572	4	0
FAGE	8380	3150	8	13	SALL	7377	7510	8	14
BEL	10592	3368	3	8	UNZ	8051	7442	3	3
TEU	8198	4758	5	6	EIG	12279	1477	5	14
BAUN	12379	3659	5	2	RANZ	10216	7837	5	4
HEKEG	10359	3474	0	0	ABRIS	11119	8626	0	0
HEST	14925	9985	9	13	GAUS	7960	3577	9	2
RIVIKUN	12986	7425	0	0	IRIVESE	15176	14356	0	0
SEREE	30047	15747	3	4	DABER	26046	10419	3	1
LEKTE	12629	8849	3	0	KIGEL	15301	5578	3	0
VERSUL	24906	12640	1	1	ZOSERN	25709	11345	1	0
BAB	3284	2543	8	18	PAU	5313	2036	8	11
PADER	23962	12728	6	5	BEIHE	18564	3281	6	1
VEILER	35545	16033	5	1	WERWEN	41171	13002	5	0
ARDE	7600	7719	5	2	WALN	5505	5119	5	5

NW three	GB	EB	GN	EN	NW four	GB	EB	GN	EN
LAREN	30685	15722	4	1	VOSEN	24166	8223	4	0
ZOR	2851	3292	5	6	KUF	2370	182	5	0
WUTE	8252	6075	5	4	KOTA	5744	3555	5	3
GANKEN	27574	10922	5	0	DIEGEL	20021	7315	5	1
GIEME	10214	4594	1	0	VETIV	10343	10508	1	0
MENA	22359	8015	0	3	SERC	23216	10190	0	2
KURI	6177	5027	4	0	HEZE	6815	2414	4	2
PACHEN	39267	11627	7	0	FEILER	35175	14315	7	1
JENTE	27395	12626	1	0	FLOND	9027	10132	1	3
DIFF	2920	3346	1	7	WUMD	1988	840	1	0
PEX	1751	2569	3	15	FOL	2112	2537	3	9
BANT	9540	8847	8	11	BOST	8473	6286	8	14
ZARAR	11177	11682	0	0	TAMAB	8306	6922	0	0
ZIND	10156	9544	7	9	WEIZ	11132	1481	7	2
KEER	22102	9463	5	10	LENN	22837	9046	5	4
FEIE	16283	1911	5	1	BEID	14430	2454	5	2
BEKT	6914	1136	5	5	SARN	6246	5461	5	7
PEU	2957	2136	4	8	EUG	2079	608	4	15
SLAN	8500	7562	6	15	MARA	9615	9610	6	10
QUCH	15784	3214	5	3	VACH	16231	4368	5	2
WOSE	6121	3981	5	7	GAUT	6146	2854	5	6
LEIV	13261	5667	5	0	TIND	13591	14567	5	12
HUDE	4846	3515	5	5	FALG	4923	4752	5	2
GUCHEN	38302	9300	5	0	FOHNEN	22367	9053	5	0
WEHKEN	25556	6247	4	1	DIEHER	32565	12211	4	1
FOTE	8229	6618	8	9	MANK	8187	6767	8	16
RIE	9150	4193	3	14	NEL	8072	5899	3	6
WIS	6165	3812	5	5	VIN	5749	8014	5	16
NUST	10283	6697	5	9	MEIM	12076	3291	5	1
ATRIV	9487	11142	0	0	ABHEU	7998	3541	0	0
FAUS	8498	3425	9	1	DAND	9301	7309	9	8
LIEHMOF	13832	6101	0	0	MAKESIS	16501	12580	0	0
MITEN	30172	13289	3	0	ERNSE	26686	11433	3	1
STARA	17506	13958	3	4	WIEKE	10634	2663	3	0
TADMER	25296	12787	1	0	WISTIN	22579	20224	1	0
SEU	5745	2683	8	8	LUT	2630	2187	8	15
PANTE	15723	13779	6	4	KANGE	17970	9099	6	2
CICKEN	25617	11189	5	1	KIETER	33878	15093	5	0
FARG	5334	4933	5	4	KALN	5486	4302	5	2

Table 13

Examples of stimulus material by condition for presentation format (Title; lower), sentence language (English sentence; German sentence), and target item. The use of bold and normal font for the auditory sentence frames indicates the pairing of sentences across sublists. E.g. Sentence frame 1 for the target CHEF in English in one sublist is paired with Sentence 2 in German in the other sublist

Condition(s)			**Prime**	**Target**
Item Type	Item Format	Sentence Language	(auditory sentence frame)	(visually presented final word)
Interlingual Homograph (IHG)	Title case		*Sentence 1 and its translation, for the target CHEF (= 'boss' in German)*	Chef
		ENG	**Because his work was so outstanding he deserved the appointment as a**	
		GER	Da seine Arbeit so hervorragend war, verdiente er die Nominierung zum	
			Sentence 2 and its translation, for the target CHEF (= 'boss' in German)	Chef
		GER	**Der Markt boomte, so dass sie beschlossen, sie brauchen einen weiteren**	
		ENG	The market was booming so they decided they needed another	
	lower case		*Sentence 1 and its translation, for the target GIFT (= 'poison' in German)*	gift
		ENG	**The woman listened to a radio show about the perfect**	
		GER	Die Frau hörte im Radio eine Sendung über das perfekte	
			Sentence 2 and its translation, for the target GIFT (= 'poison' in German)	gift
		GER	**Er dachte an den Keller als das beste Versteck für das**	
		ENG	He thought of the cellar as the best hiding place for the	
German control	Title case	**ENG**	**The aunt looked in her bag for the small**	Kamm
		GER	Die Tante suchte in ihrer Handtasche nach dem kleinen	
	lower case	**GER**	**Alexander bat seinen Nachbarn um**	mehl
		ENG	Alexander asked his neighbour for	
Nonword	Title case	**ENG**	**The people loved the goofy**	Arde
		GER	Alle liebten den albernen	
	lower case	**GER**	**Der Prüfer untersuchte sorgsam die**	nark
		ENG	The examiner carefully studied the	

Regression Models for Experiment One

Table E1. Response accuracy to interlingual homographs and controls.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.6563	0.2459	6.7356	0.0000
item type=control	0.2677	0.1520	1.7609	0.0783
L2 proficiency	1.1368	0.3449	3.2963	0.0010
format=Title	0.5025	0.2849	1.7640	0.0777
sentence language=German	0.1879	0.1040	1.8064	0.0709
word frequency	0.8008	0.1158	6.9129	0.0000
word length	0.3933	0.0835	4.7077	0.0000
experimental session=2	0.3602	0.0820	4.3923	0.0000
trial number	0.0018	0.0014	1.3188	0.1872
item type by L2 proficiency	-0.2276	0.1194	-1.9058	0.0567
L2 proficiency by format	-0.3685	0.4235	-0.8702	0.3842
L2 proficiency by sentence language	0.5362	0.1890	2.8361	0.0046
format by sentence language	-0.3179	0.1566	-2.0300	0.0424
L2 proficiency by format and sentence language	-0.6453	0.2418	-2.6685	0.0076

Intercept levels: IHG, Proficiency, Lower, English, Log Frequency, Item Length, Session One, Item number.

Table E2. Response accuracy to English-German homographs.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.6826	0.2757	6.1039	0.0000
L2 proficiency	1.1021	0.4324	2.5488	0.0108
format=Title	0.3937	0.3907	1.0078	0.3136
sentence language=German	0.3397	0.1446	2.3487	0.0188
word frequency	0.6897	0.1843	3.7422	0.0002
English word class=not noun	1.0910	0.3706	2.9439	0.0032
experimental session=2	0.4912	0.0945	5.1976	0.0000
trial number	0.0032	0.0023	1.3787	0.1680
L2 proficiency by format	-0.2185	0.5335	-0.4096	0.6821
L2 proficiency by sentence language	0.4588	0.2596	1.7676	0.0771
format by sentence language	-0.4360	0.2132	-2.0450	0.0409
L2 proficiency by format and sentence language	-0.6812	0.3236	-2.1054	0.0353

Intercept levels: Proficiency, Lower, English, Log Frequency, English Nouns, Session One, Item number.

Table E3. Response accuracy to German controls.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.3822	0.2155	11.0554	0.0000
L2 proficiency	0.8766	0.3771	2.3246	0.0201
format=Title	0.4663	0.3200	1.4571	0.1451
sentence language=German	0.0168	0.1539	0.1089	0.9133
word frequency	0.7484	0.1439	5.2001	0.0000
word length	0.3042	0.1246	2.4415	0.0146
English neighbours	-0.0555	0.0257	-2.1573	0.0310
trial number	0.0008	0.0016	0.5019	0.6157
L2 proficiency by format	-0.3809	0.4771	-0.7983	0.4247
L2 proficiency by sentence language	0.6447	0.2785	2.3145	0.0206
format by sentence language	-0.2147	0.2369	-0.9064	0.3647
L2 proficiency by format and sentence language	-0.6756	0.3699	-1.8265	0.0678

Intercept levels: Proficiency, Lower, English, Log Frequency, Item Length, English neighbours, Item number.

Table E4. Response accuracy to nonsense items (both English and German).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.1344	0.2141	9.9710	0.0000
sentence language=German	0.2914	0.0706	4.1252	0.0000
L2 proficiency	1.2197	0.2582	4.7243	0.0000
German neighbours	-0.2387	0.0294	-8.1296	0.0000
English neighbours	0.0802	0.0163	4.9138	0.0000
bigram frequency	0.0000	0.0000	-3.6991	0.0002
trial number	0.0051	0.0014	3.6370	0.0003
previous trial accuracy	0.2455	0.0892	2.7515	0.0059

Intercept level: English, Proficiency, German and English neighbours, Bigram Frequency, Item number, Previous Trial Accuracy.

Table E5. Response latencies to interlingual homographs and controls.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.2987	-1.3016	-1.3680	-1.2334	0.0001
item type=control	0.0407	0.0412	0.0042	0.0748	0.0214
sentence language=German	-0.0979	-0.0973	-0.1208	-0.0739	0.0001
L2 proficiency	-0.3710	-0.3719	-0.4976	-0.2500	0.0001
format=Title	-0.0031	-0.0033	-0.0972	0.0889	0.9450
word frequency	-0.0673	-0.0667	-0.0900	-0.0448	0.0001
word length	-0.0209	-0.0207	-0.0368	-0.0029	0.0180
experimental session=2	-0.1349	-0.1314	-0.1490	-0.1138	0.0001
trial number	0.0001	0.0001	-0.0003	0.0004	0.6972
previous trial RT	0.1196	0.1213	0.0970	0.1454	0.0001
item type by sentence language	0.0361	0.0356	0.0077	0.0614	0.0106
L2 proficiency by format	0.2345	0.2361	0.0849	0.3969	0.0034
sentence language by L2 proficiency	0.0932	0.0933	0.0503	0.1343	0.0002
sentence language by format	0.1254	0.1252	0.0982	0.1552	0.0001
sentence language by L2 proficiency and format	-0.0525	-0.0533	-0.1057	-0.0045	0.0426

Intercept levels: IHG, English, Proficiency, Lower, Log Frequency, Item Length, Session One, Item number, Previous RT.

Table E6. Response latencies to English-German homographs.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.2677	-1.2686	-1.3376	-1.1931	0.0001
L2 proficiency	-0.3764	-0.3766	-0.5035	-0.2436	0.0001
format=Title	-0.0274	-0.0282	-0.1312	0.0791	0.5892
sentence language=German	-0.1092	-0.1090	-0.1361	-0.0809	0.0001
word frequency	-0.0573	-0.0573	-0.0921	-0.0206	0.0016
experimental session=2	-0.1561	-0.1560	-0.1760	-0.1361	0.0001
trial number	-0.0001	-0.0001	-0.0006	0.0004	0.7526
previous trial RT	0.1142	0.1180	0.0851	0.1526	0.0001
L2 proficiency by format	0.2337	0.2339	0.0776	0.3919	0.0040
format by sentence language	0.1462	0.1456	0.1056	0.1835	0.0001
L2 proficiency by sentence language	0.0693	0.0686	0.0358	0.1039	0.0001

Intercept levels: Proficiency, Lower, English, Log Frequency, Session One, Item number, Previous RT.

Table E7. Response latencies to German controls.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.3334	-1.3342	-1.4007	-1.2734	0.0001
L2 proficiency	-0.3533	-0.3519	-0.4780	-0.2367	0.0001
format=Title	0.0042	0.0042	-0.0898	0.0928	0.9264
sentence language=German	-0.0586	-0.0578	-0.0851	-0.0296	0.0001
word frequency	-0.0724	-0.0720	-0.0987	-0.0446	0.0001
word length	-0.0381	-0.0379	-0.0597	-0.0159	0.0010
trial number	0.0001	0.0001	-0.0003	0.0005	0.7848
previous trial RT	0.1357	0.1395	0.1126	0.1672	0.0001
L2 proficiency by format	0.2311	0.2310	0.0797	0.3735	0.0028
L2 proficiency by sentence language	0.0943	0.0942	0.0340	0.1528	0.0016
format by sentence language	0.1195	0.1192	0.0809	0.1596	0.0001
L2 proficiency by format and sentence language	-0.0838	-0.0837	-0.1518	-0.0106	0.0210

Intercept levels: Proficiency, Lower, English, Log Frequency, Item Length, Item number, Previous RT.

Table E8. Response latencies to nonsense items (both English and German).

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.2165	-1.2172	-1.2764	-1.1543	0.0001
format=Title	0.0292	0.0285	-0.0583	0.1138	0.5094
sentence language=German	-0.0381	-0.0384	-0.0549	-0.0206	0.0001
L2 proficiency	-0.2115	-0.2119	-0.2817	-0.1418	0.0001
bigram frequency	0.0000	0.0000	0.0000	0.0000	0.0001
item length	0.0656	0.0657	0.0487	0.0832	0.0001
German neighbours	0.0204	0.0204	0.0151	0.0258	0.0001
experimental session=2	-0.1449	-0.1452	-0.1663	-0.1228	0.0001
trial number	-0.0003	-0.0003	-0.0006	0.0000	0.0536
previous trial RT	0.1148	0.1172	0.0909	0.1447	0.0001
format by sentence language	0.0667	0.0671	0.0423	0.0905	0.0001
L2 proficiency by sentence language	0.0922	0.0922	0.0706	0.1134	0.0001

Intercept levels: Lower, English, Proficiency, Bigram Frequency, Item Length, German neighbours, Session One, Item number, Previous RT.

Regression Models for Experiment Two

Table G1. Proportion of incorrect responses to interlingual homographs and controls.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.7523	0.2157	17.3967	0.0000
sentence language=German	0.4570	0.1066	4.2850	0.0000
L2 proficiency	0.6409	0.1988	3.2232	0.0013
word length	0.5312	0.1672	3.1759	0.0015
German word class=not noun	-0.6536	0.3606	-1.8125	0.0699
experimental session=2	0.8289	0.1356	6.1113	0.0000

Intercept levels: English, Proficiency, Item Length, German Nouns, Session One.

Table G2. Proportion of incorrect responses to German-English homographs.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.3524	0.2624	12.7771	0.0000
L2 proficiency	0.5227	0.2195	2.3812	0.0173
sentence language=German	0.3664	0.1419	2.5829	0.0098
word length	0.8517	0.2510	3.3927	0.0007
English word class=not noun	-1.1609	0.5461	-2.1257	0.0335
experimental session=2	0.9931	0.1463	6.7881	0.0000

Intercept levels: Proficiency, English, Item Length, English Nouns, Session One.

Table G3. Proportion of incorrect responses to German control words.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.9343	0.3454	14.2856	0.0000
L2 proficiency	0.7950	0.2598	3.0596	0.0022
sentence language=German	0.5373	0.1691	3.1768	0.0015
German word class=not noun	-1.1549	0.4528	-2.5506	0.0108
English neighbours	-0.1080	0.0450	-2.3990	0.0164

Intercept levels: Proficiency, English, German Nouns, English neighbours.

Table G4. Proportion of incorrect responses to nonwords.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.2756	0.3717	11.5038	0.0000
L2 proficiency	0.6116	0.5004	1.2222	0.2216
format=Title	-0.8981	0.4241	-2.1179	0.0342
item length	-0.6467	0.1947	-3.3207	0.0009
German neighbours	-0.3524	0.0604	-5.8339	0.0000
English neighbours	0.0844	0.0307	2.7512	0.0059
trial number	0.0134	0.0028	4.8064	0.0000
L2 proficiency by format	1.5009	0.6500	2.3093	0.0209

Intercept levels: Proficiency, Lower, Item Length, German and English neighbours, Item number.

Table G5. Mean durations of responses to German-English homographs and German control words.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.5938	-1.5961	-1.6644	-1.5274	0.0001
item type=control	0.0126	0.0126	-0.0315	0.0546	0.5644
L2 proficiency	0.0122	0.0127	-0.0817	0.1061	0.7830
sentence language=German	-0.1414	-0.1410	-0.1599	-0.1219	0.0001
word frequency	-0.0505	-0.0506	-0.0756	-0.0255	0.0001
German word class=not noun	0.0865	0.0873	0.0388	0.1343	0.0004
experimental session=2	-0.1016	-0.0980	-0.1127	-0.0829	0.0001
trial number	0.0005	0.0005	0.0001	0.0009	0.0082
previous trial RT	0.1255	0.1287	0.1071	0.1520	0.0001
item type by L2 proficiency	-0.0457	-0.0459	-0.0692	-0.0215	0.0001
item type by sentence language	0.0486	0.0484	0.0275	0.0704	0.0001
L2 proficiency by sentence language	0.0306	0.0306	0.0076	0.0563	0.0118

Intercept levels: IHG, Proficiency, English, Log Frequency, German Nouns, Session One, Item number, Previous RT.

Table G6. Mean durations of responses to German-English homographs.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.5122	-1.5125	-1.612	-1.4125	0.0001
L2 proficiency	-0.1169	-0.1159	-0.2679	0.0311	0.1268
format=Title	-0.1193	-0.1189	-0.2568	0.015	0.0896
sentence language=German	-0.1279	-0.1272	-0.1441	-0.1105	0.0001
word frequency	-0.0724	-0.072	-0.1165	-0.0282	0.0018
experimental session=2	-0.1118	-0.1113	-0.1283	-0.0944	0.0001
trial number	0.0006	0.0006	-0.0001	0.0012	0.0854
previous trial RT	0.1418	0.1487	0.1216	0.1767	0.0001
L2 proficiency by format	0.2519	0.2503	0.0557	0.4502	0.0144

Intercept levels: Proficiency, Lower, English, Log Frequency, Session One, Item number, Previous RT.

Table G7. Mean durations of responses to German control words.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.6416	-1.6424	-1.7011	-1.5838	0.0001
L2 proficiency	-0.0352	-0.0351	-0.1188	0.0511	0.4126
sentence language=German	-0.0961	-0.0951	-0.1156	-0.0758	0.0001
word frequency	-0.0320	-0.0322	-0.0606	-0.0016	0.0300
German word class=not noun	0.1266	0.1269	0.0707	0.1805	0.0001
trial number	0.0004	0.0004	-0.0001	0.0008	0.0852
previous trial RT	0.1276	0.1354	0.1116	0.1605	0.0001
L2 proficiency by sentence language	0.0346	0.0348	0.0036	0.0657	0.0276

Intercept levels: Proficiency, English, Log Frequency, German Nouns, Item number, Previous RT.

Table G8. Mean durations of responses to nonwords.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	-1.5852	-1.5852	-1.648	-1.5234	0.0001
format=Title	0.0833	0.0834	0.0038	0.1596	0.0366
sentence language=German	-0.0642	-0.0643	-0.077	-0.0504	0.0001
item length	0.0709	0.0708	0.0572	0.085	0.0001
German neighbours	0.0272	0.0272	0.0215	0.0326	0.0001
experimental session=2	-0.0936	-0.0931	-0.1158	-0.0678	0.0001
trial number	-0.0003	-0.0003	-0.0006	0	0.0552
previous trial RT	0.1468	0.1493	0.1275	0.1707	0.0001
format by sentence language	-0.036	-0.0356	-0.0536	-0.0174	0.0001

Intercept levels: Lower, English, Length, German neighbours, Session One, Item number, Previous RT.

Table 14.

Mean recognition latencies for correct responses (in milliseconds) and percent error rates (in parentheses) for IHGs, controls and nonwords in L1 (English) and L2 (German), separated by the proficiency level of the English-German bilinguals (beginner vs. advanced).

BEGINNER

	English (L1)				German (L2)			
	lower		Title		lower		Title	
Item Type	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**
IHG	988	25.9	852	23.1	871	27.5	909	24.9
Control	955	19.9	930	14.1	856	25.5	981	16.3
Nonword	996	21.5	981	27.8	905	22.1	1027	23.1

ADVANCED

	English (L1)				German (L2)			
	lower		Title		lower		Title	
Item Type	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**
IHG	779	14.2	755	10.5	721	10.4	768	9.5
Control	776	11.8	789	7.4	755	10.9	795	8.8
Nonword	789	10.8	809	4.7	777	8.6	816	3.8

Table 15.

Mean recognition latencies for correct responses (in milliseconds) and percent error rates (in parentheses) for IHGs, controls and nonwords in L1 (German) and L2 (English), separated by the proficiency level of the German-English bilinguals (beginner vs. advanced).

BEGINNER

	English (L2)				German (L1)			
	lower		Title		lower		Title	
Item Type	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**
IHG	703	9.6	639	4.8	636	6.8	566	4.6
Control	689	6.8	635	5.7	631	3.3	599	4.0
Nonword	742	2.1	737	4.4	690	2.3	695	4.5

ADVANCED

	English (L2)				German (L1)			
	lower		Title		lower		Title	
Item Type	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**	**RT**	**ER**
IHG	678	5.3	723	5.0	609	4.3	655	2.8
Control	634	3.9	700	2.8	614	2.8	648	2.2
Nonword	712	0.8	789	0.6	662	1.0	706	0.8

Appendix D

STUDY TWO

Orthographic similarity measurement

Weber's (1970) measure of graphic similarity (GS) is computed as follows (cf. van Orden 1987):

$$GS = 10\left(\frac{50F+30V+10C}{A} + 5T + 27B + 18E\right)$$

where

F = number of pairs of adjacent letters in the same order shared by word pairs,

V = number of pairs of adjacent letters in reverse order shared by word pairs,

C = number of single letters shared by word pairs,

A = average number of letters in the two words,

T = ratio of number of letters in the shorter word to the number in the longer,

B = 1 if the first letter in the two words is the same; otherwise, B=0, and

E = 1 if the last letter in the two words is the same; otherwise, E=0.

Based on the GS, the index of orthographic similarity employed by van Orden (1987) is calculated as follows:

$$OS = \frac{(GS \text{ of target foil and category exemplar})}{(GS \text{ of category exemplar and itself})}$$

Stimulus material

Table 16.

Calculated orthographic and phonological similarity measures for each cognate pair in the four relevant conditions. *Note*. OS=orthographic similarity, PS=phonological similarity.

Cognate pair (English/German)	OS	PS
+O+P		
ARM/ARM	1	1.51
BALL/BALL	1	0.65
FINGER/FINGER	1	1.17
FISH/FISCH	0.81	1.51
FORM/FORM	1	1.15
GLASS/GLAS	0.9	1.52
GOLD/GOLD	1	0.94
MAN/MANN	0.87	0.25
HAND/HAND	1	0.81
HARD/HART	0.66	0.96
MUSIC/MUSIK	0.7	0.64
WIND/WIND	1	1.1
WINTER/WINTER	1	0.99
SUMMER/SOMMER	0.83	0.92
TOMATO/TOMATE	0.75	0.64
WARM/WARM	1	0.18
SALAD/SALAT	0.73	0.38
RING/RING	1	1.32
PERSON/PERSON	1	0.38
ORANGE/ORANGE	1	0.53
-O+P		
COMB/KAMM	0.07	0.61
DANCE/TANZ	0.2	0.14
HOUSE/HAUS	0.49	0.96
ICE/EIS	0.12	1.52
LOUD/LAUT	0.38	0.94
MOUSE/MAUS	0.49	0.98
NINE/NEUN	0.5	0.67
FEVER/FIEBER	0.59	0.63
FULL/VOLL	0.46	0.76
GOOD/GUT	0.34	0.83
UNCLE/ONKEL	0.17	0.79
WIDE/WEIT	0.41	0.42
WINE/WEIN	0.56	0.42
RICE/REIS	0.41	0.83
HERE/HIER	0.57	1.48
SHOE/SCHUH	0.34	1.35
SIX/SECHS	0.34	1.1
PRICE/PREIS	0.5	1.25
SOUR/SAUER	0.56	0.63
TEA/TEE	0.56	0.7

Cognate pair (English/German)	OS	PS
+O-P		
HAT/HUT	0.61	-0.46
JULY/JULI	0.66	-0.75
JUNE/JUNI	0.66	-0.79
LAMP/LAMPE	0.74	-0.54
LICHT/LIGHT	0.78	-0.64
LONG/LANG	0.72	-0.08
HALF/HALB	0.66	-0.6
MOTHER/MUTTER	0.62	-0.18
PAPER/PAPIER	0.86	-0.45
NATION/NATION	1	-0.25
NATURE/NATUR	0.76	-0.61
NIGHT/NACHT	0.66	-0.29
NOSE/NASE	0.72	-0.74
NUMBER/NUMMER	0.8	-0.12
OPEN/OFFEN	0.66	-0.48
ROUND/RUND	0.81	-0.46
RICH/REICH	0.81	-0.67
SALT/SALZ	0.66	-0.13
SON/SOHN	0.75	-0.37
WATER/WASSER	0.72	-0.77
-O-P		
BOOK/BUCH	0.34	-0.45
CAT/KATZE	0.21	-0.48
CHILD/KIND	0.27	-1.32
COLD/KALT	0.08	-0.2
COOK/KOCH	0.27	-0.56
COW/KUH	0.05	-0.48
DARK/DUNKEL	0.34	-1.19
DAY/TAG	0.09	-1.28
EAR/OHR	0.28	-0.82
EIGHT/ACHT	0.38	-0.59
HEART/HERZ	0.49	-0.4
HEAT/HITZE	0.38	-0.64
OLD/ALT	0.09	-0.35
ONE/EIN	0.12	-1.18
RAIN/REGEN	0.51	-0.97
SEAT/SITZ	0.38	-0.5
SOAP/SEIFE	0.34	-1.41
TEN/ZEHN	0.29	-0.25
THREE/DREI	0.19	-1.05
YEAR/JAHR	0.29	-0.25

Table 17.

Stimulus material used in Experiments Three and Four, including item characteristics.

Note. Lang(uage): E=English, G=German; EF/GF=English log frequency/German log frequency (depending on the language of the item); EN=English neighbour count; GN= German neighbour count; EB=English bigram frequency; GB=German bigram frequency; L=length (in letters).

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
+O+P	ARM	English	2.36	5	5	4621	4304	3
+O+P	GLASS	English	2.18	3	1	7933	8937	5
+O+P	SALAD	English	1.32	0	1	8669	8866	5
+O+P	TOMATO	English	1.18	0	1	12163	8621	6
+O+P	WIND	English	2.16	12	7	10099	10271	4
+O+P	WARM	English	2.08	10	7	5574	5745	4
+O+P	MAN	English	3.24	19	4	6380	6914	3
+O+P	ORANGE	English	1.54	1	0	15724	23693	6
+O+P	HAND	English	2.92	9	10	8119	11562	4
+O+P	FORM	English	2.64	12	3	4717	4420	4
+O+P	BALL	English	2.07	18	5	7719	7309	4
+O+P	FISH	English	2.31	4	0	5875	5864	4
+O+P	HARD	English	2.49	13	7	6440	6834	4
+O+P	MUSIC	English	2.11	0	1	7700	9445	5
+O+P	FINGER	English	2.13	6	2	20017	36120	6
+O+P	GOLD	English	1.96	10	4	2631	2376	4
+O+P	WINTER	English	1.94	3	4	23036	34743	6
+O+P	PERSON	English	2.41	1	2	16217	25694	6
+O+P	RING	English	2.18	10	3	13564	14654	4
+O+P	SUMMER	English	2.11	3	5	11303	24763	6
+O+P control	AIR	English	2.42	5	4	2213	893	3
+O+P control	CRASH	English	1.57	5	0	8017	6445	5
+O+P control	VALID	English	1.18	1	0	8750	8590	5
+O+P control	TIMING	English	1.04	1	0	17896	17961	6
+O+P control	WIFE	English	2.41	9	0	1875	4150	4
+O+P control	HARM	English	1.62	8	7	6158	7149	4
+O+P control	FAR	English	2.84	18	7	4612	4471	3
+O+P control	ORPHAN	English	0.84	0	0	10002	11377	6
+O+P control	TALK	English	2.71	9	3	6529	6966	4
+O+P control	FOOD	English	2.49	10	1	2788	1035	4
+O+P control	TALE	English	1.51	18	4	10547	10856	4
+O+P control	BIRD	English	2.03	4	2	2826	2382	4
+O+P control	HOLE	English	1.99	14	6	7477	6935	4
+O+P control	PUPIL	English	1.67	1	0	4572	3234	5
+O+P control	SINGLE	English	2.17	5	0	17055	18345	6
+O+P control	BOLD	English	1.18	10	6	3245	2604	4
+O+P control	WORKER	English	2.33	3	2	12318	25940	6
+O+P control	REASON	English	2.53	1	0	15302	12431	6
+O+P control	WING	English	1.8	10	4	10933	12309	4
+O+P control	SUFFER	English	2.09	4	1	9656	23652	6

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
+O+P	ARM	German	2.2	5	5	4304	4621	3
+O+P	GLAS	German	1.86	1	2	5663	4942	4
+O+P	SALAT	German	0.78	1	2	10990	12127	5
+O+P	TOMATE	German	0.7	0	1	13851	14783	6
+O+P	WIND	German	1.79	7	12	10271	10099	4
+O+P	WARM	German	1.73	7	10	5745	5574	4
+O+P	MANN	German	2.87	4	4	7954	6795	4
+O+P	ORANGE	German	0	0	1	23693	15724	6
+O+P	HAND	German	2.68	10	9	11562	8119	4
+O+P	FORM	German	2.36	3	12	4420	4717	4
+O+P	BALL	German	1.79	5	18	7309	7719	4
+O+P	FISCH	German	1.53	3	2	27340	7125	5
+O+P	HART	German	2.1	13	15	8254	6788	4
+O+P	MUSIK	German	2.08	0	1	7145	4659	5
+O+P	FINGER	German	1.78	2	6	36120	20017	6
+O+P	GOLD	German	1.68	4	10	2376	2631	4
+O+P	WINTER	German	1.68	4	3	34743	23036	6
+O+P	PERSON	German	2.18	2	1	25694	16217	6
+O+P	RING	German	1.66	3	10	14654	13564	4
+O+P	SOMMER	German	1.84	1	2	24132	11967	6
+O+P control	AMT	German	2.07	7	8	1517	1089	3
+O+P control	GROB	German	1.36	2	5	3336	4102	4
+O+P control	SELIG	German	0.84	1	0	17614	9034	5
+O+P control	ROMANE	German	0.7	0	1	12473	14264	6
+O+P control	WIRT	German	1.23	6	6	4148	3007	4
+O+P control	ZAHL	German	2.3	6	1	3091	437	4
+O+P control	GANZ	German	3.12	3	1	6536	5086	4
+O+P control	BRAUCH	German	1	0	1	25004	7357	6
+O+P control	LAGE	German	2.47	9	15	10238	4999	4
+O+P control	WOHL	German	2.58	3	2	2729	709	4
+O+P control	BAHN	German	1.83	5	1	3602	1316	4
+O+P control	TISCH	German	2	3	0	31186	11259	5
+O+P control	HUND	German	1.78	10	5	11317	4591	4
+O+P control	RUHIG	German	2.15	0	0	8052	3384	5
+O+P control	LIEBER	German	1.64	2	2	35555	12497	6
+O+P control	DORF	German	1.98	4	3	3672	4289	4
+O+P control	WANDEL	German	1.23	4	1	19148	12673	6
+O+P control	FERTIG	German	1.85	2	0	31286	15409	6
+O+P control	LIED	German	1.69	3	4	10993	6345	4
+O+P control	KOFFER	German	1.32	1	1	23278	9384	6

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
+O-P	WATER	English	2.69	7	4	17429	29522	5
+O-P	JUNE	English	1.78	8	5	5774	9413	4
+O-P	RICH	English	2.16	2	4	8414	21184	4
+O-P	NOSE	English	1.93	11	7	4483	6363	4
+O-P	SON	English	2.33	20	6	6283	3553	3
+O-P	JULY	English	1.75	3	2	5195	1107	4
+O-P	PAPER	English	2.36	6	2	11604	21794	5
+O-P	NUMBER	English	2.6	2	1	9897	26200	6
+O-P	HALF	English	2.52	5	6	5834	6467	4
+O-P	NIGHT	English	2.68	9	1	4065	7532	5
+O-P	LONG	English	3.02	8	2	10403	10155	4
+O-P	NATURE	English	2.29	1	0	13786	15141	6
+O-P	ROUND	English	2.55	7	0	10405	12500	5
+O-P	LIGHT	English	2.63	9	1	5520	9958	5
+O-P	OPEN	English	2.7	2	4	7393	19633	4
+O-P	NATION	English	2.09	2	2	19611	13691	6
+O-P	LAMP	English	1.57	11	6	4494	5243	4
+O-P	SALT	English	1.7	4	8	5347	6164	4
+O-P	HAT	English	1.85	20	7	6241	5698	3
+O-P	MOTHER	English	2.72	2	0	13292	25742	6
+O-P control	WATCH	English	2.48	8	1	7973	18803	5
+O-P control	HUNT	English	1.64	7	2	5756	10561	4
+O-P control	CITY	English	2.46	2	0	4914	4556	4
+O-P control	HOPE	English	2.48	12	4	4552	3243	4
+O-P control	SIP	English	1.36	20	3	2560	2547	3
+O-P control	JURY	English	1.52	3	0	2924	1809	4
+O-P control	PAINT	English	2.01	6	1	13199	9561	5
+O-P control	SIMPLY	English	2.36	3	0	8344	4455	6
+O-P control	CALM	English	1.7	5	5	6672	3593	4
+O-P control	RIVER	English	2.13	9	0	14024	25196	5
+O-P control	HOUR	English	2.58	8	1	6600	2916	4
+O-P control	FUTURE	English	2.3	1	0	9595	11988	6
+O-P control	SOUND	English	2.46	7	1	8494	11283	5
+O-P control	FIGHT	English	2.27	9	2	3436	6428	5
+O-P control	ONLY	English	3.25	1	0	9109	3307	4
+O-P control	NOTICE	English	2.26	1	0	12433	9450	6
+O-P control	DAMP	English	1.51	9	5	2745	2684	4
+O-P control	PALM	English	1.51	4	4	5643	4656	4
+O-P control	LAW	English	2.33	19	2	2866	3175	3
+O-P control	BOTTLE	English	2.11	3	0	7985	7885	6

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
+O-P	WASSER	German	2.23	2	2	29276	15443	6
+O-P	JUNI	German	2.42	2	4	8521	3803	4
+O-P	REICH	German	2.16	3	2	32921	10280	5
+O-P	NASE	German	1.54	8	8	8115	5910	4
+O-P	SOHN	German	2.14	4	2	2328	1115	4
+O-P	JULI	German	2.5	2	2	5253	4634	4
+O-P	PAPIER	German	1.88	1	1	26750	11653	6
+O-P	NUMMER	German	1.66	4	4	24672	10732	6
+O-P	HALB	German	2.24	6	5	6754	5658	4
+O-P	NACHT	German	2.21	8	1	19925	5950	5
+O-P	LANG	German	2.74	8	13	13898	9892	4
+O-P	NATUR	German	2.08	1	0	8338	8966	5
+O-P	RUND	German	2.45	11	8	12905	5140	4
+O-P	LICHT	German	2.04	6	2	23740	8714	5
+O-P	OFFEN	German	2.27	0	2	20769	6567	5
+O-P	NATION	German	2.05	2	2	13691	19611	6
+O-P	LAMPE	German	1.23	1	0	6686	6441	5
+O-P	SALZ	German	1.04	3	2	5273	4947	4
+O-P	HUT	German	1.52	4	17	2111	1810	3
+O-P	MUTTER	German	2.31	4	6	29030	14470	6
+O-P control	WOHNEN	German	1.9	2	2	22285	8688	6
+O-P control	RUHE	German	1.87	4	4	7471	2762	4
+O-P control	RASCH	German	1.94	3	2	27914	8311	5
+O-P control	NASS	German	1.3	10	6	7158	6407	4
+O-P control	LOHN	German	1.7	4	5	3087	2104	4
+O-P control	AUGE	German	2.48	3	4	2876	2145	4
+O-P control	FAHRER	German	1.76	1	1	29273	13059	6
+O-P control	KUMMER	German	1.04	4	4	24973	10436	6
+O-P control	HALS	German	1.64	7	5	6730	5774	4
+O-P control	SACHE	German	2.42	7	1	22283	6614	5
+O-P control	LEID	German	1.94	5	3	13461	5598	4
+O-P control	DAHER	German	2.27	1	0	24840	9919	5
+O-P control	BUND	German	2	10	8	10998	4865	4
+O-P control	DICHT	German	1.75	6	0	20956	7707	5
+O-P control	OPFER	German	2.05	0	1	22325	8894	5
+O-P control	NUTZEN	German	2.05	3	0	22394	6751	6
+O-P control	WAAGE	German	1	1	1	8747	3485	5
+O-P control	SATZ	German	2.21	4	1	5648	5616	4
+O-P control	MUT	German	1.63	5	13	2196	1906	3
+O-P control	MITTEL	German	2.31	3	1	19604	11962	6

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
-O+P	WIDE	English	2.14	14	3	4547	5549	4
-O+P	HOUSE	English	2.85	6	0	9600	8494	5
-O+P	NINE	English	1.85	14	3	12342	9737	4
-O+P	PRICE	English	2.21	5	1	10003	8132	5
-O+P	SOUR	English	1.11	12	4	5930	2544	4
-O+P	FEVER	English	1.46	5	4	10775	23495	5
-O+P	FULL	English	2.45	14	4	4550	3888	4
-O+P	COMB	English	1.15	6	0	4908	1133	4
-O+P	ICE	English	1.76	4	2	5136	3202	3
-O+P	LOUD	English	1.71	5	1	5591	1960	4
-O+P	SHOE	English	1.92	6	1	3485	2980	4
-O+P	SIX	English	2.31	12	5	2000	2312	3
-O+P	MOUSE	English	1.28	8	0	9174	8151	5
-O+P	TEA	English	1.96	9	1	6966	6973	3
-O+P	UNCLE	English	1.85	0	0	8473	10946	5
-O+P	GOOD	English	3.15	9	1	2466	680	4
-O+P	HERE	English	3	9	8	13923	30053	4
-O+P	DANCE	English	2.02	2	0	8591	5677	5
-O+P	WINE	English	1.91	21	3	11379	9159	4
-O+P	RICE	English	1.45	14	0	8410	6770	4
-O+P control	WAVE	English	2.08	14	3	3811	4130	4
-O+P control	HAPPY	English	2.18	5	0	3123	4182	5
-O+P control	FILE	English	1.78	14	3	7097	6627	4
-O+P control	STICK	English	2.12	5	5	13872	16123	5
-O+P control	DOLL	English	1.43	11	6	5168	4199	4
-O+P control	FAINT	English	1.6	5	1	12363	9768	5
-O+P control	FAIL	English	2.18	15	5	3672	3162	4
-O+P control	COIN	English	1.32	8	0	10578	5357	4
-O+P control	ILL	English	1.83	3	1	4493	4069	3
-O+P control	JOKE	English	1.79	5	0	1509	3371	4
-O+P control	SHED	English	1.46	5	1	5994	6115	4
-O+P control	LIE	English	2.56	12	5	3980	9809	3
-O+P control	COUGH	English	1.38	8	0	7444	765	5
-O+P control	JEW	English	1.95	10	1	485	562	3
-O+P control	ANKLE	English	1.32	1	0	9013	11260	5
-O+P control	SOME	English	3.27	8	3	4356	4357	4
-O+P control	MEAN	English	2.81	9	1	8642	7909	4
-O+P control	DIRTY	English	1.66	1	0	5438	44323	5
-O+P control	DINE	English	1.52	21	2	12790	9379	4
-O+P control	WIRE	English	1.71	14	5	6561	8728	4

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
-O+P	WEIT	German	3.02	9	7	15222	3753	4
-O+P	HAUS	German	2.53	10	1	10058	4335	4
-O+P	NEUN	German	1.88	3	2	10772	5752	4
-O+P	PREIS	German	2.42	2	1	21609	9902	5
-O+P	SAUER	German	1.04	5	1	28550	9156	5
-O+P	FIEBER	German	0.85	2	1	32025	10413	6
-O+P	VOLL	German	2.45	7	9	4905	4536	4
-O+P	KAMM	German	0.9	4	0	3987	1627	4
-O+P	EIS	German	1	3	2	13444	3489	3
-O+P	LAUT	German	2	5	4	8542	4551	4
-O+P	SCHUH	German	1.52	2	0	22878	3292	5
-O+P	SECHS	German	2.3	0	0	21172	5899	5
-O+P	MAUS	German	1	9	3	9215	4893	4
-O+P	TEE	German	1.11	3	17	7153	5799	3
-O+P	ONKEL	German	1.38	1	0	12503	9092	5
-O+P	GUT	German	3.14	3	16	1769	1947	3
-O+P	HIER	German	3	8	3	25636	9487	4
-O+P	TANZ	German	1.45	5	2	9054	6608	4
-O+P	WEIN	German	1.58	14	6	15996	8117	4
-O+P	REIS	German	1.91	5	1	20247	8309	4
-O+P control	KEIN	German	3.1	13	3	16834	8509	4
-O+P control	WAHL	German	2.22	8	3	4154	1208	4
-O+P control	GERN	German	2.13	3	3	25750	9335	4
-O+P control	KRAFT	German	2.55	2	3	7475	4133	5
-O+P control	FADEN	German	1.2	3	1	23178	9130	5
-O+P control	FEUCHT	German	1.04	2	0	21048	3888	6
-O+P control	HOCH	German	2.89	6	2	16201	4402	4
-O+P control	SAMT	German	0.95	4	2	3164	1980	4
-O+P control	EID	German	0.7	3	8	9217	1408	3
-O+P control	GABE	German	1	6	9	7665	3268	4
-O+P control	STUFE	German	1.72	2	1	11975	5891	5
-O+P control	SORGE	German	2.04	1	5	10165	6029	5
-O+P control	RAUS	German	1.3	10	1	11640	6328	4
-O+P control	REH	German	0.7	4	7	9093	4955	3
-O+P control	INNEN	German	1.4	0	1	26529	15592	5
-O+P control	WUT	German	1.11	3	11	1568	1506	3
-O+P control	TIER	German	2.11	9	2	28669	13233	4
-O+P control	DACH	German	1.53	6	4	16514	4550	4
-O+P control	HEIL	German	1.36	10	6	14894	4150	4
-O+P control	REDE	German	2.24	5	4	11598	10001	4

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
-O-P	SOAP	English	1.36	0	6	1559	2436	4
-O-P	DARK	English	2.32	5	11	4857	5081	4
-O-P	EAR	English	1.96	5	11	3475	6382	3
-O-P	CHILD	English	3.04	0	2	18277	5691	5
-O-P	HEART	English	2.23	0	0	10404	9548	5
-O-P	HEAT	English	2.15	2	11	7848	9001	4
-O-P	COLD	English	2.3	3	10	2168	5058	4
-O-P	COW	English	1.64	0	25	72	4031	3
-O-P	DAY	English	2.96	0	17	602	1374	3
-O-P	BOOK	English	2.63	3	12	762	2714	4
-O-P	YEAR	English	3.14	0	12	3486	6599	4
-O-P	COOK	English	1.97	0	12	326	4527	4
-O-P	EIGHT	English	2	0	9	14233	2779	5
-O-P	CAT	English	1.85	2	22	2868	7196	3
-O-P	ONE	English	3.54	1	3	5455	8930	3
-O-P	TEN	English	2.36	7	20	24237	9010	3
-O-P	RAIN	English	1.96	4	10	9861	11804	4
-O-P	OLD	English	2.95	0	1	2147	2103	3
-O-P	THREE	English	2.8	0	0	9618	8186	5
-O-P	SEAT	English	2.12	6	16	7373	9595	4
-O-P control	MOAN	English	1.23	1	6	5771	5974	4
-O-P control	FAST	English	2.16	9	11	9699	6571	4
-O-P control	TAX	English	2.14	8	12	3328	2398	3
-O-P control	CHASE	English	1.42	1	4	23324	7813	5
-O-P control	HABIT	English	1.81	1	0	9896	6585	5
-O-P control	MEAT	English	1.9	3	13	5850	9061	4
-O-P control	MOVE	English	2.7	2	11	3482	4226	4
-O-P control	COP	English	1.26	1	24	658	3960	3
-O-P control	DOG	English	2.09	2	14	900	1622	3
-O-P control	LOOK	English	3.17	0	10	1855	3491	4
-O-P control	HEAD	English	2.77	3	12	5724	5543	4
-O-P control	CURE	English	1.58	2	11	8435	7303	4
-O-P control	TIGHT	English	1.72	0	9	10274	7570	5
-O-P control	BAT	English	1.34	8	22	4432	5804	3
-O-P control	ANY	English	3.11	0	2	4930	4496	3
-O-P control	TIN	English	1.65	7	20	9786	12270	3
-O-P control	JAIL	English	1.54	0	11	2085	3198	4
-O-P control	OWN	English	2.99	0	2	52	1425	3
-O-P control	EARLY	English	2.68	0	0	4313	10372	5
-O-P control	TEAR	English	2.11	3	16	10159	10951	4

Condition	Target	Language	EF/GF	EN	GN	EB	GB	L
-O-P	KALT	German	1.93	5	6	4592	5939	4
-O-P	DUNKEL	German	1.98	0	0	6175	17176	6
-O-P	BUCH	German	2.26	5	5	3096	16066	4
-O-P	KIND	German	2.62	8	6	10008	9400	4
-O-P	TAG	German	2.94	21	8	3277	4567	3
-O-P	HITZE	German	1.23	0	3	4638	8979	5
-O-P	OHR	German	1.69	4	3	378	2155	3
-O-P	KUH	German	1.6	1	1	39	1208	3
-O-P	SEIFE	German	0.85	2	4	4046	15660	5
-O-P	HERZ	German	2.18	6	5	9110	23906	4
-O-P	JAHR	German	3.34	0	2	562	2849	4
-O-P	EIN	German	4.22	11	4	7479	13745	3
-O-P	ACHT	German	2.16	1	2	4336	17827	4
-O-P	KATZE	German	1.26	0	3	5318	7275	5
-O-P	REGEN	German	1.83	0	12	11281	31973	5
-O-P	KOCH	German	1.04	1	6	2857	16556	4
-O-P	ZEHN	German	2.35	0	2	719	5159	4
-O-P	ALT	German	2.86	10	13	4456	4517	3
-O-P	DREI	German	2.8	0	3	5662	15865	4
-O-P	SITZ	German	1.72	1	3	4727	7898	4
-O-P control	HAUT	German	1.69	5	7	3612	8244	4
-O-P control	ZUFALL	German	1.52	0	0	7402	9480	6
-O-P control	LUFT	German	2.04	6	3	1085	3979	4
-O-P control	KAUM	German	2.43	0	7	1540	7169	4
-O-P control	MAL	German	2.69	13	6	6151	5349	3
-O-P control	BITTE	German	1.34	1	3	8966	13636	5
-O-P control	ORT	German	2.05	6	3	4497	4941	3
-O-P control	KUR	German	1.08	3	4	1753	2609	3
-O-P control	STIER	German	0.95	2	4	17375	35297	5
-O-P control	GELD	German	2.3	5	5	4285	11791	4
-O-P control	SATZ	German	2.21	1	4	5616	5648	4
-O-P control	SIE	German	3.98	18	4	2830	7863	3
-O-P control	ECHT	German	1.94	1	2	3892	18298	4
-O-P control	TAUBE	German	1.11	0	3	4403	13218	5
-O-P control	WEGEN	German	2.49	0	12	7099	27421	5
-O-P control	TOPF	German	1	4	5	2968	3036	4
-O-P control	ZEIT	German	3.02	1	3	3572	14823	4
-O-P control	ALS	German	3.68	5	9	4237	3885	3
-O-P control	FRAU	German	2.85	1	2	4627	9428	4
-O-P control	WITZ	German	1.2	1	3	3459	6840	4

Regression Models for Experiment Three

Table E9. Response durations to cognates and controls.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.4769	6.4843	6.3093	6.6661	0.0001
status = control	0.0564	0.0557	0.0277	0.0829	0.0001
L2 proficiency	0.0641	0.0632	-0.0428	0.1738	0.2428
language = German	0.1087	0.1121	0.0879	0.1369	0.0001
word frequency	-0.0500	-0.0524	-0.0874	-0.0194	0.0022
previous RT	0.1361	0.1349	0.1112	0.1574	0.0001
trial number	-0.0006	-0.0006	-0.0012	0.0001	0.0734
status by L2 proficiency	-0.0594	-0.0590	-0.0867	-0.0287	0.0004
L2 proficiency by language	-0.1390	-0.1371	-0.1669	-0.1082	0.0001
status by word frequency	-0.0539	-0.0528	-0.0950	-0.0111	0.0148

Intercept levels: Cognates, L2 Proficiency, English Task, Log Frequency, Previous RT, Item Number.

Table E10. Response accuracy to cognates and controls.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.8877	0.2991	19.6840	0.0000
language = German	-0.6426	0.3346	-1.9206	0.0548

Intercept level: English Task.

Table E11. Main response durations to cognates.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.3495	6.3221	6.0581	6.5643	0.0001
L2 proficiency	0.0528	0.0528	-0.0487	0.1559	0.3118
language = German	0.0902	0.0894	0.0174	0.1600	0.0120
orth.sim = positive	-0.0203	-0.0160	-0.0867	0.0532	0.6606
phon.sim = positive	-0.0527	-0.0530	-0.1227	0.0210	0.1540
word frequency	-0.0659	-0.0665	-0.1014	-0.0352	0.0001
previous RT	0.1556	0.1591	0.1272	0.1915	0.0001
trial number	-0.0007	-0.0008	-0.0016	0.0001	0.0732
L2 proficiency by language	-0.1089	-0.1073	-0.1483	-0.0688	0.0001
orth.sim by phon.sim	0.0538	0.0512	-0.0459	0.1507	0.3028
language by orth.sim	0.0361	0.0296	-0.0715	0.1219	0.5510
language by phon.sim	0.0611	0.0629	-0.0404	0.1618	0.2198
language by orth.sim and phon.sim	-0.1332	-0.1284	-0.2525	-0.0019	0.0428

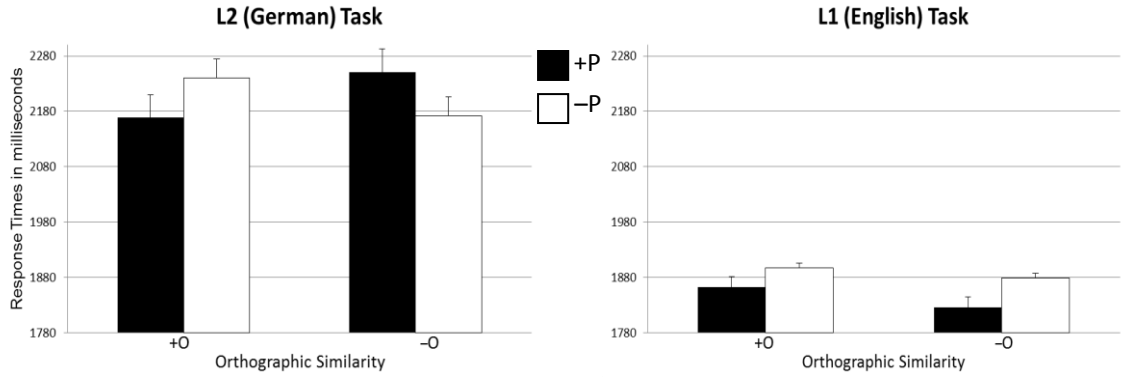
Intercept levels: L2 Proficiency, English Task, Negative Orthographic Similarity Rating, Negative Phonological Similarity Rating, Log Frequency, Previous RT, Item Number.

Table E12. Response accuracy to cognates.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	6.0081	0.4555	13.1887	1.02E-39
langauge = German	-0.9227	0.5463	-1.6888	0.0913
English neighbours	-0.0853	0.0488	-1.7488	0.0803

Intercept levels: English Task and English Neighbourhood Density.

Figure 109. Mean durations of responses to the four O-P cognate conditions, by orthographic similarity (positive vs. negative), phonological similarity (positive vs. negative) and task language (L2 German, in the left panel, and L1 English, in the right panel). To provide a visual comparison of the present findings with Schwartz et al.'s (2007) study, the following two figures are bar-charts of means and standard errors in the relevant conditions for the duration results. Technical limitations did not allow me to construct these graphs based on *lmer* model estimates. Instead, these figures are based on fairly noisy data (i.e. while it includes outliers, incorrect responses were excluded), and they do not reflect the complexity that is inherent to a linear mixed-effects model (including the absence of various effect-defining variables). Nevertheless, the figures reflect the general result trends and provide a useful comparison to Schwartz et al.'s study.



Regression Models for Experiment Four

Table G9. Response durations to cognates and controls.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.7868	6.7957	6.5859	7.0020	0.0001
L2 proficiency	-0.0710	-0.0589	-0.2394	0.1332	0.5320
language = German	-0.0769	-0.0775	-0.1083	-0.0453	0.0001
status = control	0.0692	0.0686	0.0440	0.0937	0.0001
word frequency	-0.0347	-0.0360	-0.0639	-0.0080	0.0148
word class = not noun	0.0386	0.0414	0.0182	0.0655	0.0004
word length	0.0199	0.0198	0.0033	0.0371	0.0234
German neighbours	-0.0041	-0.0042	-0.0079	-0.0007	0.0252
previous RT	0.0988	0.0963	0.0727	0.1205	0.0001
trial number	-0.0009	-0.0009	-0.0016	-0.0003	0.0038
L2 proficiency by language	0.0760	0.0754	0.0406	0.1105	0.0001
status by word frequency	-0.0454	-0.0449	-0.0811	-0.0079	0.0132

Intercept levels: L2 Proficiency, English Task, cognates, Log Frequency, Nouns, Item Length, German Neighbourhood Density, Previous RT, Item Number.

Table G10. Response accuracy to cognates and controls.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.0244	0.2139	23.4867	6E-122
word class = not noun	-0.5878	0.2323	-2.5306	0.0114
German neighbours	0.0826	0.0410	2.0158	0.0438
English neighbours	-0.0509	0.0219	-2.3234	0.0202

Intercept levels: Nouns, German and English Neighbourhood Densities.

Table G11. Main response durations to cognates.

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC
(Intercept)	6.6181	6.6191	6.3443	6.8915	0.0001
orth.sim = positive	0.0317	0.0316	-0.0380	0.0995	0.3668
phon.sim = positive	0.0571	0.0571	0.0094	0.1019	0.0148
L2 proficiency	-0.1016	-0.1009	-0.2956	0.0871	0.2820
language = German	-0.1130	-0.1111	-0.1757	-0.0485	0.0004
German neighbours	-0.0070	-0.0071	-0.0110	-0.0033	0.0004
previous RT	0.1202	0.1197	0.0868	0.1527	0.0001
trial number	-0.0014	-0.0014	-0.0022	-0.0006	0.0004
orth.sim by phon.sim	-0.0754	-0.0751	-0.1445	-0.0077	0.0336
orth.sim by prof.score	0.0042	0.0032	-0.0663	0.0710	0.9242
orth.sim by language	0.1156	0.1159	0.0377	0.2025	0.0060
L2 proficiency by language	0.1375	0.1348	0.0643	0.2059	0.0002
orth.sim by L2 proficiency and language	-0.1151	-0.1123	-0.2117	-0.0164	0.0232

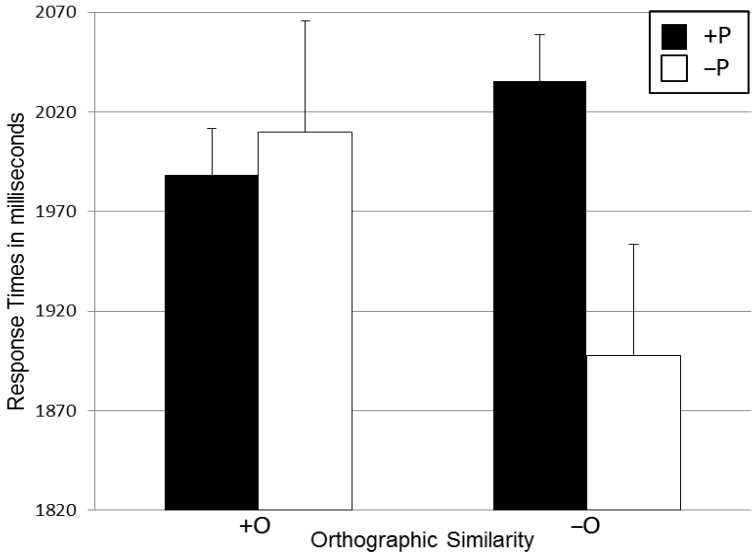
Intercept levels: Negative Orthographic Similarity Rating, Negative Phonological Similarity Rating, L2 Proficiency, English Task, German Neighbourhood Density, Previous RT, Item Number.

Table G12. Response accuracy to cognates and controls.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.3303	0.3040	17.5329	8E-69
word class = not noun	-0.9740	0.3686	-2.6421	0.0082
German neighbours	0.1114	0.0606	1.8370	0.0662

Intercept levels: Nouns and German Neighbourhood Density.

Figure 110. Mean durations of responses to the four O-P cognate conditions (*lmer* model estimates), by their orthographic and phonological similarity (both positive vs. negative). As in Experiment Three – for purposes of comparison with Schwartz et al.'s (2007) study – this figure presents bar-charts of means and standard errors in the relevant conditions for the duration results, rather than *lmer* model estimates based on transformed data.



Appendix E

ETHICS APPROVAL

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

MEMORANDUM

TO	Anna Piasecki
COPY TO	Associate Professor Paul Warren, Supervisor
FROM	Dr Allison Kirkman, Convener, Human Ethics Committee
DATE	March 10, 2009
PAGES	1
SUBJECT	**Ethics Approval: No 16246, Lexical Access in L2 Speakers**

Thank you for your application for ethical approval, which has now been considered by the Standing Committee of the Human Ethics Committee.

Your application has been approved from the above date and this approval continues until 28 February 2011. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Allison Kirkman
Convener