

Katarzyna Jankowiak

Lexico-semantic processing in bilingual
figurative language comprehension:
Behavioral and electrophysiological
evidence

Przetwarzanie leksykalno-semantyczne w
rozumieniu języka figuratywnego w
dwujęzyczności: Wskaźniki behawioralne
oraz elektrofizjologiczne

Praca doktorska napisana
na Wydziale Anglistyki
Uniwersytetu im. Adama Mickiewicza w Poznaniu
pod kierunkiem dr hab. Joanny Pawelczyk, prof. UAM
oraz dr Karoliny Rataj

Poznań, 2018

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my two advisors, who have served as excellent examples of both successful researchers and wonderful people to look up to:

To Prof. Joanna Pawelczyk for her tremendous enthusiasm, encouragement, and advice. Thank you for always believing in me and for being there whenever I needed support. Your knowledge, experience, and worldview have taught and inspired me in many ways. I am forever grateful.

To Dr. Karolina Rataj for her continuous support, motivation, and wealth of knowledge. Your guidance helped me throughout the whole process of research and writing this thesis. Thank you for providing me with constant inspirations for further work, and for helping me to grow as a researcher.

My words of appreciation go to Prof. Ryszard Naskręcki for scientific cooperation. Without this support, it would not have been possible to conduct this research. I would also like to extend my gratitude to Justyna Jakubczak for her contribution to data collection. Thank you.

I am extremely grateful to the Kulczyk Foundation for their trust and support. Thank you for being a great source of inspiration, and for showing me that where there is a will, there is always a way.

Importantly, this thesis could not have come into being without the support of those closest to me:

My family. Thank you for your unconditional love, support, and for teaching me the value of hard work and perseverance. I could not be more grateful.

My amazing colleague and friend, Paweł Korpala. Thank you for your friendship, advice, and for giving me the necessary pep-talks whenever I started doubting myself.

My dearest friends, who have always encouraged me to pursue this PhD journey: Urszula Grzesiek, Tomasz Józefiak, Agnieszka Kalisz, Joanna Kowalska, Karol Krajewski, Michał Mańka, and Patrycja Wietrzycka. Thank you for your love, enthusiasm and never-ending support.

Finally, I am grateful to my study participants. It is literally the case that this dissertation could not have been completed without your participation. Thank you.

OŚWIADCZENIE

Ja, niżej podpisana

Katarzyna Jankowiak

przedkładam rozprawę doktorską

pt. „Lexico-semantic processing in bilingual figurative language comprehension:
Behavioral and electrophysiological evidence”
(„Przetwarzanie leksykalno-semantyczne w rozumieniu języka figuratywnego w
dwujęzyczności: Wskaźniki behawioralne oraz elektrofizjologiczne”)

**na Uniwersytecie im. Adama Mickiewicza w Poznaniu
i oświadczam,**

że napisałam ją samodzielnie.

Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałam z pomocy innych osób, a w szczególności nie zlecałam opracowania rozprawy lub jej istotnych części innym osobom, ani nie odpisywałam tej rozprawy lub jej istotnych części od innych osób.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

(miejsowość, data)

(czytelny podpis)

Table of contents

| | |
|--|------------|
| TABLE OF CONTENTS | IV |
| LIST OF TABLES | X |
| LIST OF FIGURES | XII |
| INTRODUCTION | 1 |
| CHAPTER 1 : INVESTIGATING LANGUAGE COMPREHENSION: A REVIEW OF SELECTED OFF-LINE AND ON-LINE METHODS | 5 |
| 1.1. INTRODUCTORY REMARKS | 5 |
| 1.2. OFF-LINE METHODS IN INVESTIGATING LANGUAGE COMPREHENSION: SURVEY RESEARCH | 6 |
| 1.2.1. <i>Types of survey questions.....</i> | 7 |
| 1.2.2. <i>Survey research in normative studies</i> | 8 |
| 1.2.3. <i>Strengths and limitations of survey research.....</i> | 9 |
| 1.3. ON-LINE METHODS IN INVESTIGATING LANGUAGE COMPREHENSION: REACTION TIME RESEARCH | 11 |
| 1.3.1. <i>Selected research paradigms based on the reaction time method.....</i> | 12 |
| 1.3.2. <i>Strengths and limitations of reaction time research.....</i> | 14 |
| 1.4. ON-LINE METHODS IN INVESTIGATING LANGUAGE COMPREHENSION: ELECTROPHYSIOLOGICAL RESEARCH | 15 |
| 1.4.1. <i>The origins of the EEG signal.....</i> | 15 |
| 1.4.2. <i>Principles of EEG recording</i> | 17 |
| 1.4.3. <i>Event-related potentials (ERPs)</i> | 20 |
| 1.4.3.1. <i>The N400 and lexico-semantic access</i> | 21 |
| 1.4.3.2. <i>The late positive complex (LPC) and meaning integration</i> | 27 |
| 1.4.4. <i>Strengths and limitations of electrophysiological research.....</i> | 29 |
| 1.5. RESEARCH METHODS EMPLOYED IN THE PRESENT DISSERTATION | 30 |
| 1.6. CONCLUDING REMARKS | 32 |

| | |
|---|-----------|
| CHAPTER 2 : METAPHORIC MEANING COMPREHENSION..... | 34 |
| 2.1. INTRODUCTORY REMARKS | 34 |
| 2.2. FIGURATIVE LANGUAGE..... | 34 |
| 2.2.1. <i>From literal to figurative language</i> | 35 |
| 2.2.2. <i>A traditional perspective on figurative language comprehension: The Standard Pragmatic View</i> | 37 |
| 2.3. SELECTED THEORIES AND MODELS OF METAPHORIC MEANING COMPREHENSION .. | 39 |
| 2.3.1. <i>The Relevance Theory</i> | 40 |
| 2.3.2. <i>The Constraint Satisfaction View</i> | 42 |
| 2.3.3. <i>The Underspecification View</i> | 43 |
| 2.3.4. <i>The Graded Salience Hypothesis</i> | 43 |
| 2.3.5. <i>The Conceptual Metaphor Theory</i> | 45 |
| 2.3.6. <i>The Career of Metaphor Model</i> | 48 |
| 2.3.7. <i>Computational models of metaphor comprehension</i> | 50 |
| 2.4. FROM METAPHOR TO SIMILE COMPREHENSION | 54 |
| 2.4.1. <i>Comparison and categorization views on metaphor and simile processing.</i> | 54 |
| 2.4.2. <i>Structure mapping in analogy and similarity</i> | 55 |
| 2.4.3. <i>The role of topic-vehicle relationship in metaphor and simile comprehension</i> | 57 |
| 2.4.4. <i>A Relevance Theory perspective on metaphor and simile comprehension</i> ... | 58 |
| 2.5. FACTORS MODULATING METAPHOR COMPREHENSION..... | 58 |
| 2.6. METAPHORIC MEANING COMPREHENSION: EMPIRICAL EVIDENCE..... | 60 |
| 2.6.1. <i>How context modulates metaphoric language comprehension</i> | 62 |
| 2.6.2. <i>How metaphor conventionality modulates metaphoric language comprehension</i> | 64 |
| 2.6.3. <i>How comparison mechanisms modulate metaphor and simile comprehension</i> | 72 |
| 2.7. CONCLUDING REMARKS | 74 |
| CHAPTER 3 : BILINGUAL LEXICO-SEMANTIC PROCESSING | 75 |
| 3.1. INTRODUCTORY REMARKS | 75 |
| 3.2. SELECTED THEORIES AND MODELS OF BILINGUAL LEXICO-SEMANTIC PROCESSING | 76 |
| 3.2.1. <i>Structure-oriented perspectives on bilingual language processing</i> | 76 |

| | |
|---|-----|
| 3.2.1.1. Compound, coordinate, and subordinate bilingualism | 76 |
| 3.2.1.2. Bilingual Dual-Coding Theory | 78 |
| 3.2.1.3. Distributed Feature Model | 81 |
| 3.2.2. <i>Process-oriented perspectives on bilingual language processing</i> | 82 |
| 3.2.2.1. Hierarchical Model | 82 |
| 3.2.2.2. Bilingual Interactive Activation Models..... | 86 |
| 3.3. BILINGUAL LEXICO-SEMANTIC PROCESSING: EMPIRICAL EVIDENCE..... | 93 |
| 3.3.1. <i>The N400 and lexico-semantic access in bilingualism</i> | 93 |
| 3.3.1.1. The N400 effect in studies on translation priming: Testing the Revised Hierarchical Model | 94 |
| 3.3.1.2. The N400 effect in studies on language non-selectivity: Testing the Bilingual Interactive Activation Models..... | 98 |
| 3.3.1.3. The N400 sensitivity to L2 proficiency level, language dominance, and age of L2 acquisition..... | 102 |
| 3.3.2. <i>The LPC and semantic integration in bilingualism</i> | 106 |
| 3.4. BILINGUAL FIGURATIVE LANGUAGE COMPREHENSION | 107 |
| 3.4.1. <i>Behavioral and electrophysiological research into bilingual nonliteral meaning comprehension</i> | 109 |
| 3.5. CONCLUDING REMARKS | 116 |

CHAPTER 4 : EXPERIMENT 1: NOVEL AND CONVENTIONAL METAPHOR COMPREHENSION IN BILINGUALISM 117

| | |
|---|-----|
| 4.1. INTRODUCTORY REMARKS | 117 |
| 4.2. AIMS, RESEARCH QUESTIONS AND HYPOTHESES..... | 117 |
| 4.3. METHOD | 123 |
| 4.3.1. <i>Participants</i> | 123 |
| 4.3.2. <i>Materials</i> | 124 |
| 4.3.2.1. Normative studies: Critical words | 126 |
| 4.3.2.2. Normative studies: Word dyads..... | 126 |
| 4.3.3. <i>Procedure</i> | 130 |
| 4.3.3.1. Electrophysiological recording | 132 |
| 4.3.3.2. Overview of statistical analyses..... | 133 |
| 4.4. RESULTS | 134 |

| | |
|--|------------|
| 4.4.1. Behavioral results | 134 |
| 4.4.1.1. Accuracy rates..... | 134 |
| 4.4.1.2. Reaction times..... | 136 |
| 4.4.1.3. Correlation analyses..... | 138 |
| 4.4.2. Electrophysiological results..... | 140 |
| 4.4.2.1. P200 (150-250 ms)..... | 140 |
| 4.4.2.2. Early N400 (300-400 ms) | 142 |
| 4.4.2.3. Late N400 (400-500 ms)..... | 144 |
| 4.4.2.4. Late positivity (500-800 ms)..... | 147 |
| 4.5. DISCUSSION AND CONCLUSION..... | 151 |
| | |
| CHAPTER 5 : EXPERIMENT 2: NOVEL NOMINAL METAPHOR AND NOVEL SIMILE COMPREHENSION IN BILINGUALISM..... | 154 |
| 5.1. INTRODUCTORY REMARKS | 154 |
| 5.2. AIMS, RESEARCH QUESTIONS AND HYPOTHESES..... | 155 |
| 5.3. METHOD | 158 |
| 5.3.1. Participants..... | 158 |
| 5.3.2. Materials..... | 158 |
| 5.3.2.1. Normative studies: Critical words | 160 |
| 5.3.2.2. Normative studies: Sentences | 161 |
| 5.3.3. Procedure..... | 165 |
| 5.3.3.1. Overview of statistical analyses..... | 166 |
| 5.4. RESULTS | 167 |
| 5.4.1. Accuracy rates | 167 |
| 5.4.2. Reaction times..... | 169 |
| 5.4.3. Correlation analyses..... | 170 |
| 5.5. DISCUSSION AND CONCLUSION..... | 170 |
| | |
| CHAPTER 6 : LEXICO-SEMANTIC PROCESSING IN BILINGUAL FIGURATIVE LANGUAGE COMPREHENSION: GENERAL DISCUSSION | 172 |
| 6.1. INTRODUCTORY REMARKS | 172 |
| 6.2. EXPERIMENT 1: NOVEL AND CONVENTIONAL METAPHOR COMPREHENSION IN BILINGUALISM..... | 173 |

| | |
|--|------------|
| 6.2.1. Normative studies' results in light of previous research on bilingual lexico-semantic processing..... | 173 |
| 6.2.2. Behavioral results..... | 175 |
| 6.2.3. Electrophysiological results..... | 178 |
| 6.2.4. Limitations of the study..... | 183 |
| 6.3. EXPERIMENT 2: NOVEL NOMINAL METAPHOR AND NOVEL SIMILE COMPREHENSION IN BILINGUALISM | 185 |
| 6.3.1. Normative studies' results in light of previous research on bilingual lexico-semantic processing..... | 186 |
| 6.3.2. Behavioral results..... | 187 |
| 6.3.3. Limitations of the study..... | 189 |
| 6.4. FIGURATIVE LANGUAGE COMPREHENSION IN BILINGUALISM: INSIGHTS FROM EXPERIMENT 1 AND EXPERIMENT 2 | 191 |
| 6.4.1. Novel and conventional metaphor comprehension in bilingualism..... | 191 |
| 6.4.2. Lexico-semantic access and meaning integration in bilingualism..... | 194 |
| 6.5. THE IMPORTANCE OF METHODOLOGY TRIANGULATION..... | 196 |
| 6.6. FUTURE RESEARCH DIRECTIONS | 198 |
| 6.7. CONCLUDING REMARKS | 200 |
| CONCLUSION | 201 |
| SUMMARY | 204 |
| STRESZCZENIE..... | 206 |
| REFERENCES..... | 208 |
| APPENDIX A: ELECTROPHYSIOLOGICAL EXPERIMENTS INTO THE N400 AND LPC RESPONSES IN MONOLINGUAL METAPHORIC LANGUAGE COMPREHENSION | 239 |
| APPENDIX B: ELECTROPHYSIOLOGICAL EXPERIMENTS INTO THE N400 AND LPC RESPONSES IN BILINGUAL LANGUAGE PROCESSING (BASED ON A REVIEW PAPER BY JANKOWIAK AND RATAJ 2017: 141-151)..... | 247 |
| APPENDIX C: EXPERIMENT 1: POLISH EXPERIMENTAL STIMULI | 257 |
| APPENDIX D: EXPERIMENT 1: ENGLISH EXPERIMENTAL STIMULI | 260 |

| | |
|--|------------|
| APPENDIX E: EXPERIMENT 2: POLISH EXPERIMENTAL STIMULI | 263 |
| APPENDIX F: EXPERIMENT 2: ENGLISH EXPERIMENTAL STIMULI | 266 |
| APPENDIX G: EXPERIMENT 1: INSTRUCTIONS IN THE NORMATIVE STUDIES ON POLISH EXPERIMENTAL WORD DYADS..... | 269 |
| APPENDIX H: EXPERIMENT 1: INSTRUCTIONS IN THE NORMATIVE STUDIES ON ENGLISH EXPERIMENTAL WORD DYADS..... | 272 |
| APPENDIX I: EXPERIMENT 2: INSTRUCTIONS IN THE NORMATIVE STUDIES ON POLISH EXPERIMENTAL SENTENCES..... | 275 |
| APPENDIX J: EXPERIMENT 2: INSTRUCTIONS IN THE NORMATIVE STUDIES ON ENGLISH EXPERIMENTAL SENTENCES..... | 278 |
| APPENDIX K: AN INFORMED CONSENT FORM (IN POLISH)..... | 281 |

List of tables

| | |
|---|-----|
| Table 1. Polish and English critical words' characteristics, including their frequency per million, number of syllables, and number of letters (adapted from Jankowiak et al. 2017: 6). | 124 |
| Table 2. Demographic information regarding the participants of the five normative studies in each language, including the number of raters, their gender, and mean age (adapted from Jankowiak et al. 2017: 7). | 125 |
| Table 3. Results of the normative tests on the stimuli included in the Experiment 1 (adapted from Jankowiak et al. 2017: 8). | 130 |
| Table 4. An interaction between language and laterality over central, centro-parietal, and parietal electrode positions within the 150-250 ms time window (adapted from Jankowiak et al. 2017: 14). | 141 |
| Table 5. Between-language differences within the 150-250 ms time window, with mean amplitudes for Polish and English utterances over left and right central, centro-parietal, and parietal electrode positions (adapted from Jankowiak et al. 2017: 14). | 141 |
| Table 6. Main effect of language within the 300-400 ms time window, with mean amplitudes for Polish and English utterances over central, centro-parietal, and parietal electrode positions (adapted from Jankowiak et al. 2017: 17). | 143 |
| Table 7. Between-language differences within the 300-400 ms time window, with mean amplitudes for Polish and English utterances over left and right electrode positions (adapted from Jankowiak et al. 2017: 17). | 143 |
| Table 8. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous (ANO) word pairs over midline and right electrode positions within the 400-500 ms time window (adapted from Jankowiak et al. 2017: 19). | 146 |
| Table 9. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous (ANO) word pairs over fronto-central and central electrode positions within the 500-800 ms time window (adapted from Jankowiak et al. 2017: 23). | 147 |

| | |
|--|-----|
| Table 10. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous word pairs (ANO) in Polish and English within the 500-800 ms time window (adapted from Jankowiak et al. 2017: 23)..... | 150 |
| Table 11. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous (ANO) word pairs within the 500-800 ms time window (adapted from Jankowiak et al. 2017: 24)..... | 150 |
| Table 12. Polish and English critical words' characteristics, including their frequency per million, number of syllables, and number of letters..... | 159 |
| Table 13. Mean sentence length of Polish and English experimental materials..... | 159 |
| Table 14. Demographic information regarding the participants of the five normative studies in each language, including the number of raters, their gender, and mean age. | 160 |
| Table 15. Results of the normative tests on the stimuli included in the Experiment 2. | 165 |

List of figures

- Fig. 1. The structure of a neuron and the direction of a nerve impulse generated in response to a stimulus (after Sanei and Chambers 2007: 8) 16
- Fig. 2. Excitatory and inhibitory presynaptic and postsynaptic activity (after Sanei and Chambers 2007: 5). 17
- Fig. 3. Electrode montage with 64 channels placed at the standard extended 10-20 position (after Brain Products 2012). 18
- Fig. 4. Grand average ERPs in response to sentence-final critical words, which were expected items (i.e., *palms*), unexpected and implausible representatives of the same semantic category (i.e., *pinés*), or unexpected and implausible representatives of a different semantic category (i.e., *tulips*). The presented N400 response varies as a function of plausibility, with the most pronounced N400 amplitudes for *tulips*, followed by *pinés*, and *palms*. The effect was observed over the right-medial central electrode position (after Kutas and Federmeier 2000: 466). 23
- Fig. 5. Grand average ERPs in response to sentence-final critical words, which were either grammatical or ungrammatical. The presented P600/LPC response varies as a function of grammatical correctness, with more pronounced P600/LPC amplitudes for grammatically incorrect compared to grammatically correct words. The effect was observed over the Pz electrode (after Osterhout and Nicol 1999: 299). 28
- Fig. 6. The Conceptual Metaphor Theory. In line with the assumptions postulated within this theory, a metaphor *Our marriage was a rollercoaster ride* is comprehended through the preactivation of the conceptual metaphor LOVE IS A JOURNEY. Arrows going from JOURNEY to LOVE present the direction of the cross-domain mapping from the conceptual metaphor source JOURNEY to the conceptual metaphor target LOVE (after McGlone 1996: 548, adapted from Lakoff and Johnson 1980). 46
- Fig. 7. The distinction between novel and conventional metaphors according to the Career of Metaphor Model. While a conventional metaphoric vehicle (i.e., base) is claimed to have both a literal and metaphoric reference, a novel metaphoric vehicle is assumed to have only a literal reference (after Bowdle and Gentner 1999: 92). 49

| | |
|---|-----|
| Fig. 8. Neighborhood cliques of the verb <i>drink</i> according to the distributional models of word co-occurrence (after Andrews et al. 2009: 477)..... | 51 |
| Fig. 9. The network representing the metaphor <i>Happiness is gold</i> . The network consists of an argument (<i>happiness</i>), a predicate (<i>gold</i>), and three close neighbors of the predicate (<i>precious</i> , <i>nuggets</i> , <i>currencies</i>). The solid lines provide cosines between the items in the LSA space, while the dashed lines indicate the activation between the three neighbors of the predicate within the network (after Kintsch and Bowles 2002: 261)..... | 52 |
| Fig. 10. Three bilingual storage representations (after Heredia and Cieřlicka 2014: 13, adapted from Weinreich 1953)..... | 77 |
| Fig. 11. The Bilingual Dual-Coding Theory (after Paivio and Desrochers 1980: 391).. | 80 |
| Fig. 12. The Distributed Feature Model (after Heredia and Brown 2006: 239, adapted from de Groot 1992)..... | 81 |
| Fig. 13. Word Association, Concept Mediation, and the Revised Hierarchical Model of lexical and conceptual representation in bilingual memory (after Kroll and Stewart 1994: 158) | 84 |
| Fig. 14. The Interactive Activation Model of monolingual visual word recognition (after McClelland and Rumelhart 1981: 380)..... | 87 |
| Fig. 15. The Bilingual Interactive Activation Model (BIA) of bilingual word recognition (after Dijkstra and van Heuven 2002: 177, adapted from Dijkstra and van Heuven 1998)..... | 88 |
| Fig. 16. The Bilingual Interactive Activation Plus Model (BIA+) of bilingual word recognition (after Dijkstra and van Heuven 2002: 182)..... | 89 |
| Fig. 17. The extended BIA+ model of bilingual word recognition (after Van Kesteren et al. 2012: 2148)..... | 92 |
| Fig. 18. Meaningfulness ratings for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, literal, and anomalous word pairs (1 = totally meaningless, 7 = totally meaningful). | 127 |
| Fig. 19. Familiarity ratings for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, and literal word pairs (1 = very rarely, 7 = very frequently). | 128 |

| | |
|--|-----|
| Fig. 20. Metaphoricity ratings for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, and literal word pairs (1 = very literal, 7 = very metaphorical)..... | 129 |
| Fig. 21. Time sequence of stimulus presentation (after Jankowiak et al. 2017: 9)..... | 131 |
| Fig. 22. Accuracy rates (%) for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, literal, and anomalous word dyads (adapted from Jankowiak et al. 2017: 11)..... | 136 |
| Fig. 23. Reaction times (ms) for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, literal, and anomalous word dyads (adapted from Jankowiak et al. 2017: 13)..... | 138 |
| Fig. 24. Scatterplots showing the correlation between meaningfulness ratings (y-axis) and accuracy rates (x-axis) for Polish (left-hand side) and English (right-hand side) novel metaphoric, conventional metaphoric, and literal word dyads (after Jankowiak et al. 2017: 12). | 139 |
| Fig. 25. Scatterplot showing the correlation between familiarity ratings obtained from English native speakers (y-axis) and English L2 learners (x-axis) for English novel metaphoric, conventional metaphoric, and literal word pairs. | 140 |
| Fig. 26. A: Grand averages for Polish (solid line) and English (dotted line) utterances over left central and centro-parietal electrodes, where the P200 effect was maximal (adapted from Jankowiak et al. 2017: 15). B: Topographic distribution of Polish and English word dyads in the 150-250 ms time window. Voltage maps were obtained for the averaged value of difference waves (English word pairs minus Polish word pairs; after Jankowiak et al. 2017: 16). | 142 |
| Fig. 27. A: Grand averages for Polish (solid line) and English (dotted line) utterances over right central, centro-parietal, and parietal electrodes, where the early N400 effect was maximal (adapted from Jankowiak et al. 2017: 19). B: Topographic distribution of Polish and English word dyads in the 300-400 ms time window. Voltage maps were obtained for the averaged value of difference waves (Polish word pairs minus English word pairs; after Jankowiak et al. 2017: 18). | 144 |
| Fig. 28. Topographic distribution of novel metaphoric (NM), conventional metaphoric (CM), literal (LIT), and anomalous (ANO) word dyads within the 400-500 ms time window. Voltage maps were obtained for the averaged value of difference waves (anomalous minus literal word pairs, novel metaphoric minus literal word pairs, and | |

| | |
|--|-----|
| conventional metaphoric minus literal word pairs; after Jankowiak et al. 2017: 20). | 145 |
| Fig. 29. Grand averages for anomalous (black solid line), novel metaphoric (grey solid line), conventional metaphoric (black dashed line), and literal word dyads (black dotted line) over midline and right electrode positions (adapted from Jankowiak et al. 2017: 21). | 146 |
| Fig. 30. Grand averages for anomalous (black solid line), novel metaphoric (grey solid line), conventional metaphoric (black dashed line), and literal word dyads (black dotted line) over fronto-central and central electrodes (adapted from Jankowiak et al. 2017: 22). | 148 |
| Fig. 31. Topographic distribution of novel metaphoric (NM), conventional metaphoric (CM), literal (LIT), and anomalous (ANO) word dyads in Polish (PL) and in English (ENG) within the 500-800 ms time window. Voltage maps were obtained for the averaged value of difference waves (Polish conventional metaphoric minus Polish novel metaphoric word pairs, Polish literal minus Polish novel metaphoric word pairs, English conventional metaphoric minus English novel metaphoric word pairs, and English literal minus English novel metaphoric word pairs; after Jankowiak et al. 2017: 23). | 149 |
| Fig. 32. Topographic distribution of novel metaphoric (NM), conventional metaphoric (CM), literal (LIT), and anomalous (ANO) word dyads within the 500-800 ms time window. Voltage maps were obtained for the averaged value of difference waves (literal minus anomalous word pairs, literal minus novel metaphoric word pairs, literal minus conventional metaphoric word pairs, and conventional metaphoric minus novel metaphoric word pairs; after Jankowiak et al. 2017: 24). | 151 |
| Fig. 33. Meaningfulness ratings for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, literal, and anomalous sentences (1 = totally meaningless, 7 = totally meaningful). | 162 |
| Fig. 34. Familiarity ratings for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, and literal sentences (1 = very rarely, 7 = very frequently). | 163 |
| Fig. 35. Metaphoricity ratings for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, and literal sentences (1 = very literal, 7 = very metaphorical). | 164 |

| | |
|---|-----|
| Fig. 36. Time sequence of stimulus presentation..... | 166 |
| Fig. 37. Accuracy rates (%) for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, literal, and anomalous sentences. | 168 |
| Fig. 38. Reaction times (ms) for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, literal, and anomalous sentences. | 169 |
| Fig. 39. Scatterplots showing the correlation between meaningfulness ratings (y-axis) and accuracy rates (x-axis) for Polish (left-hand side) and English (right-hand side) novel nominal metaphors, novel similes, and literal sentences. | 170 |

Introduction

Lexico-semantic processing in metaphoric language comprehension has frequently been discussed in the monolingual context. Although a traditional perspective on figurative meaning processing (the Standard Pragmatic View; Grice 1975) claims that nonliteral language comprehension is delayed relative to the literal meaning, more recent approaches emphasize that metaphor comprehension is modulated by such factors as contextual cues (Frisson and Pickering 2001; Katz and Ferretti 2001) or meaning conventionality (Gentner 1983; Sperber and Wilson 1986; Bowdle and Gentner 2005). For instance, the Career of Metaphor Model (Bowdle and Gentner 2005) claims that specific cognitive mechanisms engaged in metaphor processing are influenced by metaphor conventionality, with novel (unfamiliar) metaphors requiring comparison processes involved in meaning construction, and conventional (familiar) metaphors preferentially involving categorization mechanisms engaged in meaning retrieval. Consequently, it has been argued that novel metaphor comprehension should be easier when such metaphors are already presented in the comparison form (i.e., as similes), which has been supported in previous monolingual experiments (Bowdle and Gentner 2005; Shibata et al. 2012).

However, little attention has been devoted to examining metaphor comprehension in the context of bilingualism. Previous bilingual studies into various other types of figurative utterances (e.g., irony, proverbs, and idioms) have suggested that bilingual speakers are less sensitive to nonliteral expressions presented in their non-native tongue (Danesi 1992; Littlemore and Low 2006: 6). Experiments conducted thus far have, nonetheless, rarely focused on examining whether meaning conventionality modulates metaphor comprehension also in the non-native tongue. Consequently, specific mechanisms

engaged when processing novel metaphors (e.g., comparison mechanisms) and conventional metaphoric utterances (e.g., categorization mechanisms) in bilingualism remain under-investigated. Importantly, examining metaphor comprehension can provide valuable insights into how bilingual speakers compute meanings of different semantic complexity, as semantically simple meanings (e.g., literal) can be contrasted with meanings that are semantically complex (e.g., metaphoric). Consequently, this can show how semantic complexity modulates bilingual language processing.

The main objective of two experiments described in the thesis was to extend monolingual research into metaphor comprehension to the context of bilingualism. Experiment 1 aimed to examine behavioral as well as electrophysiological correlates of novel and conventional metaphor comprehension in the native (Polish; L1) and non-native language (English; L2). In Experiment 2, reaction times were measured in response to L1 and L2 novel nominal metaphors and novel similes in order to investigate whether a comparison form could ease novel metaphor comprehension in both languages.

This dissertation is divided into three parts: a theoretical part (Chapters 1-3), an experimental part (Chapters 4-5), and a general discussion (Chapter 6). Chapter 1 provides an overview of selected off-line and on-line methods that can be employed when studying language comprehension. The following two chapters discuss metaphoric meaning processing (Chapter 2) and bilingual lexico-semantic processing (Chapter 3). The experimental part of the thesis includes the description of the design as well as the results obtained from the two experiments (Chapter 4 and Chapter 5, respectively). Finally, a general discussion (Chapter 6) provides an interpretation of the results obtained from the studies, and relates them to the theoretical frameworks previously discussed in the theoretical chapters.

Chapter 1 discusses three quantitative research methods that can be used in studies on both figurative and bilingual language processing, pointing to their advantages as well as limitations. First, survey research is presented as an off-line measure used to investigate the perception of various types of linguistic stimuli. A special emphasis is placed on rating scales, which allow for a quantitative analysis of survey results. Next, reaction time (RT) measures are discussed as one of the most widely employed methods for studying real-time language processing. Finally, an electroencephalographic (EEG) measure is presented, with a special focus on event-related potentials (ERPs) as a type of an analysis

that can be performed based on the recorded EEG signal. In addition, two language-related ERP components that have been frequently analyzed in studies on lexico-semantic processing are discussed: the N400 and the late positive complex (LPC). Importantly, Chapter 1 provides a background on how to interpret results obtained from RT and ERP experiments, which the two following chapters extensively refer to when discussing previous studies on metaphoric (Chapter 2) and bilingual (Chapter 3) language processing. The research methods discussed in Chapter 1 were employed in the two experiments reported in this thesis (Chapter 4 and Chapter 5).

Chapter 2 discusses metaphoric meaning comprehension in monolingual speakers. It presents both traditional and contemporary approaches aimed at explicating mechanisms engaged in metaphor and simile comprehension (Grice 1975; Sperber and Wilson 1986; Frisson and Pickering 2001; Katz and Ferretti 2001; Giora 2002; Lakoff and Johnson 1980; Bowdle and Gentner 2005; Kintsch 2001; Gentner 1983). Importantly, Chapter 2 identifies specific factors that have been postulated to modulate processes involved in nonliteral language processing, such as contextual cues as well as meaning conventionality. Following the sections devoted to the literature review, Chapter 2 presents both behavioral (RT) and electrophysiological (ERPs) research that has thus far been conducted with a view to elucidating mechanisms behind metaphor and simile comprehension, with a special focus on the role of metaphor conventionality in figurative meaning processing.

Chapter 3 focuses on bilingual lexico-semantic processing. It provides a review on the most prominent models of bilingual visual word processing (Weinreich 1953; Paivio and Desrochers 1980; de Groot 1992; Kroll and Stewart 1994; Dijkstra and van Heuven 1998; Dijkstra and van Heuven 2002), and presents studies that have tested the tenets proposed within each model. Chapter 3 also discusses bilingual figurative language comprehension, and identifies gaps in existing research on bilingual metaphor as well as simile comprehension, which the experiments discussed in Chapter 4 and Chapter 5 were aimed to fill in.

The experimental part of this dissertation (Chapter 4 and Chapter 5) is devoted to reporting on two experiments on lexico-semantic processing in bilingual figurative language comprehension. First, Chapter 4 presents an ERP experiment that was aimed at examining whether metaphoric meaning comprehension is modulated by metaphor conventionality also in the context of bilingualism. The experimental materials used in the

study included novel metaphoric, conventional metaphoric, literal, and anomalous word dyads, which were presented in Polish (L1) and English (L2). The word pairs were all paired pre-experimentally, and were pretested by means of employing rating scales with a view to evaluating the experimental stimuli in terms of their predictability, meaningfulness, familiarity, and metaphoricity. Based on the obtained results, it was possible to ensure that experimental materials were adequate representatives of categories ascribed to them. In the EEG experiment, late proficient unbalanced Polish-English bilinguals performed a semantic decision task in response to L1 and L2 novel metaphoric, conventional metaphoric, literal, and anomalous word pairs. Data analyses were based on behavioral results (reaction times and accuracy rates) and event-related brain potentials.

Chapter 5 presents an experiment aimed to examine novel nominal metaphor and novel simile comprehension in the context of bilingualism, and thus to investigate whether comparison mechanisms initiated when processing similes could ease novel meaning comprehension in both the native and non-native tongue. The experimental stimuli involved Polish (L1) and English (L2) novel nominal metaphors and novel similes, as well as literal and anomalous sentences, which were pretested under the same criteria as in Experiment 1 (Chapter 4). In the experiment proper, late proficient unbalanced Polish-English bilinguals performed a semantic decision task in response to L1 and L2 novel nominal metaphors, novel similes, as well as literal and anomalous sentences. Data analyses were based on reaction times as well as accuracy rates.

Finally, Chapter 6 provides a general discussion based on the results obtained from the two experiments, and reflects on the hypotheses formulated for each experiment. Furthermore, it presents the implications of the results with reference to the theoretical accounts previously discussed in the theoretical chapters, provides possible limitations of the two experiments, and offers potential future research directions.

Chapter 1: Investigating language comprehension: A review of selected off-line and on-line methods

1.1. Introductory remarks

In order to investigate how language is processed in the human mind, psycholinguists usually employ quantitative research techniques, which provide empirical evidence to a phenomenon under investigation. Although there is a number of such research techniques, each of them has certain limitations. Therefore, mechanisms engaged in language processing can be best examined by means of method triangulation, which allows for providing different perspectives, and can consequently counteract weaknesses that individual methods suffer from. The two experiments described in this dissertation aimed at examining lexico-semantic processes engaged in bilingual metaphor comprehension by means of employing three quantitative research methods; namely, survey-based, behavioral, and electrophysiological research techniques. With a view to providing a rationale behind employing these particular research methods, this chapter discusses advantages and limitations offered by each of these techniques. Additionally, the present chapter provides a relevant background of how results obtained from studies employing these methods could be interpreted. The following chapters of this dissertation frequently refer to this background when discussing previous research on metaphoric (Chapter 2) as well as bilingual language processing (Chapter 3).

1.2. Off-line methods in investigating language comprehension: Survey research

Off-line methods used in psycholinguistic research are based on post-interpretive processing of a particular linguistic stimulus, and therefore reflect a post-hoc usage of the extracted meaning (Newman et al. 2009: 2499). Survey research is one of the most frequently used types of nonexperimental off-line research methods, and aims to elicit attitudes, opinions, and beliefs by means of designing a questionnaire for a sample selected from the target population of interest (Christensen et al. 2014: 334). Therefore, the main objective of surveys is to obtain an accurate picture of attitudes shared by the individuals being studied. On the basis of this, a researcher can generalize from the sample to the target population (Gravetter and Forzano 2012: 373; Christensen et al. 2014: 359).

There are two most prevalent survey instruments that are employed in survey research: a questionnaire survey and an interview survey (Mitchell and Jolley 2010: 263). While the former involves participants providing their responses to the presented questions in a written form, the latter is based on the interaction between a participant and an interviewer, and involves participants hearing the questions and presenting their responses. The present chapter focuses on questionnaire surveys, as these were employed in the two experiments described in this dissertation.

Questionnaires can be either self-administered, where participants fill in the survey without investigators being present, or investigator-administered, in which case a researcher is present when participants are completing a survey (Mitchell and Jolley 2010: 263). With the advent of the widespread access to the Internet, researchers are now increasingly often collecting questionnaire data by means of designing an electronic, web-based survey (Christensen et al. 2014: 342). This type of a survey involves contacting a number of respondents over the Internet and asking them to complete a survey that is available on a specific website. Consequently, most of such web-based surveys are self-administered, and they do not require the presence of an investigator when participants are responding to survey questions.

1.2.1. Types of survey questions

Participants taking part in surveys might share their opinions by means of responding to a variety of survey questions. This section is devoted to discussing two most common types of survey questions; namely, open-ended and fixed-alternative questions.

Open-ended questions are designed with a view to allowing participants to respond to the provided questions in their own words. Consequently, they usually resemble fill-in-the-blank, short-answer or essay-like questions (Mitchell and Jolley 2010: 275; Gravetter and Forzano 2012: 374). Due to the fact that such questions impose only a few restrictions, participants are more likely to reveal their actual opinions and beliefs. Additionally, respondents have a larger flexibility in choosing how to answer, as a result of which open-ended questions often reveal attitudes that could be surprising even to investigators themselves. Importantly, advantageous though such flexibility might seem to be, it also poses possible weaknesses, as the provided answers are often impossible to compare and summarize. As a result, answers obtained from open-ended questions tend to be difficult to analyze with conventional statistical tests, which makes the obtained data difficult to corroborate using additional data collection methods (Gravetter and Forzano 2012: 375).

In addition to open-ended questions, questionnaires often involve fixed-alternative questions, in which participants respond to a question by means of choosing between two or more answers (Mitchell and Jolley 2010: 272). When discussing fixed-alternative questions, two variants can be differentiated: dichotomous items and rating scales. Dichotomous items require participants to provide an answer that fits a given category. For instance, respondents might categorize themselves according to their sex, race, or religion (Mitchell and Jolley 2010: 272). Alternatively, dichotomous questions might require participants to respond to yes/no questions (e.g., *Are you married?*). Even though answers to such questions usually provide reliable and valid data, nominal-dichotomous items may sometimes provide not enough options to choose from, which may artificially limit survey participants to only few alternatives that not necessarily reflect their actual answers and opinions (Mitchell and Jolley 2010: 273).

On the other hand, in rating scales, respondents rate multiple items by selecting an appropriate numerical value on a predetermined scale, and thus such scales provide them with more options to select from (Gravetter and Forzano 2012: 376). Rating scales

are adopted from Likert scales (Likert 1932), which originally aimed to investigate the extent to which participants agree with a provided statement. For example, respondents might reflect on how much they agree or disagree with a sentence *Women deserve equal pay*. A rating scale often takes the following form:

- (1) strongly disagree
 - (2) disagree
 - (3) neither agree or disagree
 - (4) agree
 - (5) strongly agree
- (Likert 1932)

Consequently, while dichotomous items might require respondents to either agree or disagree, rating scales provide participants with more variety when choosing the answer that reflects their attitudes towards the subject under investigation (Mitchell and Jolley 2010: 274). Optimally, a rating scale should provide an equal distance between different response options, so that the distance between “strongly agree” and “agree” as a 1-point difference ought to be equivalent to every other 1-point difference on the scale. Most commonly, researchers design rating scales with 5 to 9 numerical values, with two reasons behind that. First of all, most participants try to avoid extreme categories (e.g., 1 or 5), and thus a scale with fewer than 5 categories would not provide an appropriate range of possible responses. Secondly, participants usually find it difficult to discriminate among more than 9 categories, and consequently they would likely blend the categories when provided with more than 9 options (Gravetter and Forzano 2012: 377).

1.2.2. Survey research in normative studies

As survey research aims at obtaining respondents’ attitudes and opinions towards certain topics, it can be effectively implemented in language studies in order to investigate views and judgments about almost any aspect of language use. Many psycholinguists, for example, adopt rating scales in order to assess the perception of linguistic stimuli by a given population. This is of special importance when conducting normative studies, which are

carried out before the experiment proper, and whose aim is to ensure that the experimental materials are perceived as intended (Keating and Jegerski 2015).

In a normative study, respondents are usually recruited from the same population as participants who are going to take part in the main study, and are often asked to rate words or sentences on a rating scale. The numerical values obtained from the ratings are averaged, as a result of which researchers can decide whether the items are suitable for inclusion in the experiment proper (Keating and Jegerski 2015). For instance, normative studies with rating scales have previously been conducted in order to examine the strength of association between different words (Phillips et al. 2004; Martin et al. 2012), the semantic plausibility of sentences (Roberts and Felser 2011; Jouravlev and Jared 2014), the similarity of translation equivalents (Lijewska and Chmiel 2015), the degree of words' concreteness (Choi et al. 2014), and the emotional intensity of phrases (Caldwell-Harris and Ayçiçeği-Dinn 2009).

Although in normative studies, rating scales are the most commonly adopted types of questions, many researchers also use a cloze probability test, in which respondents are presented with sentences that are truncated before the sentence-final critical word (e.g., *Tomorrow I'm going to visit my _____*). Participants are instructed to provide the first word that comes to their mind so that the sentence would be semantically and syntactically correct (Bambini et al. 2013). The cloze probability test therefore allows for examining how much a context sentence suggests a to-be-inferred concept, and can be used to validate whether the critical words are embedded in low or highly constraining contexts (Monzó and Calvo 2002).

1.2.3. Strengths and limitations of survey research

Survey research offers a great number of strengths. First of all, it allows for relatively easy, fast, and efficient gathering of a large amount of information (Gravetter and Forzano 2012: 373). Surveys can provide information regarding a wide variety of different independent variables, and can be employed by linguists to investigate respondents' perception of a number of aspects, such as plausibility or meaningfulness of the provided items. Secondly, conducting a web-based survey is relatively cost-free, as it does not require postage or printing costs. In addition, web-based surveys enable fast access to a wide

audience, irrespectively of their geographical location. Next, rating scales are advantageous, as they tap into two key dimensions by measuring both direction (positive or negative) and strength of attitude (Christensen et al. 2014: 346). Additionally, participants usually find rating scales easy to understand and answer, which makes it possible to collect a host of data on various topics within a single survey (Gravetter and Forzano 2012: 378). Finally, unlike open-ended questions, responses obtained from rating scales are in the form of numerical values, thus allowing for running descriptive and inferential statistical analyses (Christensen et al. 2014: 342).

Advantageous as survey research might seem to be, it also exhibits some weaknesses. Its major limitation pertains to the fact that it is based on self-report, and thus respondents' answers may not necessarily reflect their actual opinions (Mitchell and Jolley 2010: 260). Namely, researchers have little control over how honest their participants are when completing a survey, as a result of which it is likely that some respondents might distort or conceal information. Additionally, web-based surveys are usually based on a volunteer sampling, which is a non-random sampling method, where participants decide themselves whether to be a part of the sample (Christensen et al. 2014: 342). Consequently, instead of providing representative samples, a volunteer sampling might provide samples that differ from the target population. Furthermore, web-based surveys are prone to a nonresponse bias, which results from the fact that such surveys are usually responded to by participants interested in the survey topic, who might not be representative of the entire target population (Gravetter and Forzano 2012: 380). Another disadvantage pertains to rating scales, which may not always provide an accurate reflection of how participants really feel. For instance, respondents who do not know an answer to the question may feel forced to choose any option, even though it does not reflect their actual position. In addition, rating scales suffer from the response set tendency, which pertains to participants choosing the same response in order to answer most of the questions (Gravetter and Forzano 2012: 377). This problem can be minimized by means of including items that are representatives of both positive and negative categories, so that respondents are more encouraged to use the opposite sides of the scale, and are thus less prone to fall into a single response set.

Importantly, survey research is an off-line measure of participants' attitudes and opinions, and does not allow for recording ongoing real-time processes taking place during language comprehension. Consequently, even though surveys provide information

regarding a final product of language comprehension, they do not capture the temporal dynamics of language processing. Results obtained from survey research ought to be therefore corroborated by means of employing experimental methods of data collection, such as, for example, behavioral or electrophysiological measures.

1.3. On-line methods in investigating language comprehension: Reaction time research

Unlike survey research, on-line methods provide real-time information regarding various aspects of language processing, and are of great significance in psycholinguistic research, as language processing is a rapid and an extremely transient mechanism (Kaiser 2013: 135). For example, many aspects of language processing are below the threshold of human conscious perception. This can be tapped into when using a masked priming paradigm, in which a prime word is presented for a very short time (usually around 50 ms), is preceded by a forward mask and might be additionally followed by a backward mask, both of which are usually presented as a string of hash marks (i.e., #####). The short presentation duration along with the masking technique results in the prime word being not consciously perceptible. Consequently, a masked priming paradigm can index automatic language-related processes that are strategy-free. Interestingly, even under such conditions, the semantic priming effect occurs, as target words are recognized faster when preceded by semantically related prime words. Such results indicate a high automaticity of language processing, which would not be detected without on-line experimental paradigms. Therefore, on-line methods can be employed in order to examine aspects of language processing which are often unnoticeable by language users, and they can be used to test competing theories of how language is processed in the human mind (Kaiser 2013: 137).

Measuring reaction times (RTs) is now one of the most widely used methods for studying real-time language processing (Kaiser 2013: 137). Reaction times pertain to how fast participants perform various types of linguistic tasks; for instance, how quickly they read sentences or decide whether a presented utterance is meaningful. Thus, reaction times reflect the time between the presentation of a stimulus and the response given, as they are defined as “the elapsed time between the presentation of a sensory stimulus and

a subsequent behavioral response” (Shelton and Kumar 2010: 30). In a reaction time experiment, once the information about the presentation of an external stimulus reaches the visual cortex, the brain processes it, after which the motor cortex becomes activated in order to command the hand muscles to make a decision by means of pressing an appropriate button that corresponds to either meaningful or meaningless stimuli.

Importantly, reaction times are argued to be an indicator of linguistic complexity, with longer reaction times reflecting increased processing difficulty (Kaiser 2013: 137). For instance, previous research has found longer RTs for low-frequency compared to high-frequency words, thus indicating that retrieving low-frequency words from the semantic memory network requires an increased cognitive load (Whaley 1978).

1.3.1. Selected research paradigms based on the reaction time method

A great variety of paradigms can be employed when using a reaction time method, all of which are aimed at investigating the speed of language comprehension. The most common paradigms include lexical decision, self-paced reading, and semantic decision paradigms (Kaiser 2013: 138). In a lexical decision paradigm, participants are presented with words, pseudowords (i.e., pronounceable yet non-existing words), and nonwords (i.e., unpronounceable and non-existing words), and decide whether the presented string of letters is a real (i.e., existing) word in a given language (Meyer and Schvaneveldt 1971). In a self-paced reading paradigm, participants read a text and press a button in order to uncover its subsequent part (Just et al. 1982). Finally, in a semantic decision paradigm, participants judge the semantic properties of a given item. For instance, they might indicate the semantic category of a word (e.g., animal vs. object; Spruyt et al. 2009), or decide whether the presented item (e.g., a sentence) is meaningful or meaningless (e.g., De Grauwe et al. 2010). The following section focuses on the semantic decision paradigm, as a semantic decision task was employed in both experiments reported in this dissertation.

A semantic decision paradigm requires participants to attend more closely to the semantic dimensions of a presented stimulus. The paradigm can involve three different tasks; namely, a go/no-go, two-choice (binary), or multiple-choice task, all of which differ in terms of cognitive demands as well as mental processes that they engage (Gomez et al.

2007: 2). In a go/no-go task, participants respond to one type of stimuli (the *go* trials; e.g., meaningful items) and withhold their response to the other type of stimuli (the *no-go* trials; e.g., meaningless items). Thus, a go/no-go task requires response inhibition during the *no-go* trials (Gomez et al. 2007: 22). A go/no-go task is, nonetheless, relatively less frequently used in studies on language comprehension when compared to a two-choice or multiple-choice task, both of which are based on the *go* trials only, thus requiring participants to respond to any type of a stimulus.

In a semantic decision paradigm adopting a two-choice task, participants respond to a stimulus by means of pressing one of the two response keys (Gomez et al. 2007: 1). For instance, participants might be instructed to press one key if they assume that a sentence is meaningful, and the other key when it is meaningless. Such a binary decision is therefore postulated to involve two mechanisms: the semantic decision itself and the response execution (Gordon and Caramazza 1982; Gomez et al. 2007: 2). Although a two-choice task is the most commonly used measure employed in a semantic decision paradigm, it requires participants to make a binary decision, which does not necessarily reflect their actual opinion regarding the stimulus. For instance, when presented with sentences of different degree of semantic complexity (e.g., literal, anomalous, and metaphoric sentences), participants might likely find some sentences easy to classify as either meaningful (i.e., literal) or meaningless (i.e., anomalous), while other utterances (i.e., metaphoric) might be relatively more difficult to categorize as clearly meaningful or meaningless. Consequently, researchers examining semantic processing sometimes employ a multiple-choice task, in which participants judge how much sense a presented utterance makes by means of, for instance, pressing one of three keys. For example, in studies conducted by Lai et al. (2009) as well as Lai and Curran (2013), participants made their decisions by pressing one of the four keys, which corresponded to either perfect sense, some sense, little sense, and no sense. Consequently, a multiple-choice task allows for examining the degree of sensicality of the presented utterances.

Importantly, even though a semantic decision task is relatively frequently employed in studies on language comprehension, it might be argued that since it requires decisions regarding the meaningfulness of presented utterances, participants are likely to look for a potential meaning of a stimulus more intensively than in natural language comprehension. This might in turn pose a threat to the ecological validity of a semantic decision paradigm.

1.3.2. Strengths and limitations of reaction time research

Reaction time (RT) measures offer a relatively easy method of collecting on-line behavioral data, and are frequently employed to investigate various aspects of language processing. Unlike studies based on off-line measures (e.g., surveys), experiments employing RTs can address research questions that tap into the transient nature of mechanisms engaged in language comprehension. Furthermore, reaction time experiments are fairly easy to design and their results might be easily analyzed using inferential statistics. All of these strengths result in the fact that nowadays, many psycholinguistic experiments are based on measuring response times with a view to providing evidence to leading questions in psycholinguistics. For instance, RT patterns in response to different word types (e.g., abstract vs. concrete words) can provide information regarding the role of particular lexico-semantic variables (e.g., word concreteness) in language processing.

Yet, reaction time methods also possess some limitations, one of them being a speed-accuracy trade-off (Salthouse and Hedden 2002: 859). Namely, in RT experiments, participants are asked to respond as quickly and accurately as possible, which often turns out to be an incompatible requirement. As a result, fast responses are often accompanied by lower accuracy rates. In contrast, longer response times allow for information accumulation, which brings about better performance. The trade-off between accuracy and speed can, however, be modulated by experimental procedures or instructions (Glucksberg 2001: 24). For instance, in the response deadline procedure, participants need to make their decisions within the predefined time interval. Furthermore, instructions can be suggestive and they might encourage participants to focus on either speed or accuracy. Additionally, with a view to minimizing the problem with the speed-accuracy trade-off, many researchers decide to analyze reaction times for correct responses only (Salthouse and Hedden 2002: 859).

Importantly, although reaction time measures provide researchers with information regarding how much time participants need in order to perform a specific task, they do not reveal specific cognitive mechanisms engaged when processing a particular linguistic stimulus. Additionally, equivalent RTs for given stimulus categories do not necessarily reflect equivalent cognitive demands. Reaction time methods can therefore benefit to a great extent when combined with electrophysiological methods, such as elec-

troencephalography (EEG), which offers a continuous measure of brain activity and, unlike reaction times measures, reflects an electrophysiological rather than behavioral response to a stimulus.

1.4. On-line methods in investigating language comprehension: Electrophysiological research

Over the recent decades, research on language processing has benefited greatly from the development of electroencephalography (EEG) as a non-invasive electrophysiological method that offers the temporal resolution up to 1 ms. The EEG method can thus provide the precise time course of language processing, and allows for capturing cognitive mechanisms in the precise time frame, in which a given mental operation takes place (Cohen 2014: 15).

1.4.1. The origins of the EEG signal

The EEG method provides a direct measure of neural activity, which reflects biophysical events from populations of neurons (Cohen 2014: 16). Neurons are defined as functional units of the nervous system, which transmit information by means of electrical and chemical signals (Sanei and Chambers 2007: 5). The move of electrical charges within the central nervous system creates an electrical signal, which can be recorded by the electroencephalogram. The electrical signal pertains to a measure of synaptic ionic currents, which are generated within dendrites as a result of the activation of neurons within the cerebral cortex (Sanei and Chambers 2007: 7; Lopes da Silva 2010: 20).

Fig. 1 presents the structure of a neuron and the current that flows through it in response to a stimulus. The current in the brain is generated by means of pumping the positive ions of sodium, potassium, and calcium, as well as the negative ions of chlorine through the membrane of the neuron in the direction specified by the membrane potential, as a result of which neurons become depolarized (Sanei and Chambers 2007: 7). Upon a complete depolarization of a neuron, which takes place once a stimulus exceeds the

threshold level, an action potential is created. A rapid change that takes place in the membrane potential results in the intracellular potential jumping from negative to positive, and then rapidly returning to the resting intracellular negativity (Tatum et al. 2008: 2; Lopes da Silva 2010: 20). Nerve impulses that are consequently generated travel along the axon, and are released at axon terminals as postsynaptic potentials (Luck 2005: 27).

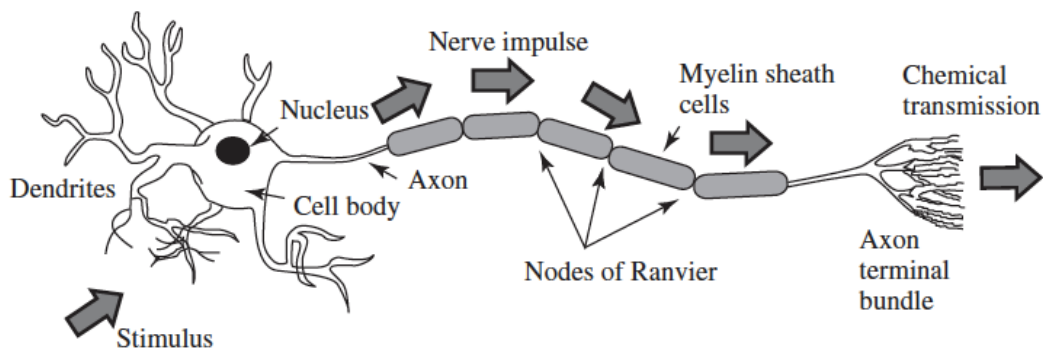


Fig. 1. The structure of a neuron and the direction of a nerve impulse generated in response to a stimulus (after Sanei and Chambers 2007: 8)

An EEG electrode placed on the scalp rarely records action potentials, as these last only about 1 ms and can cancel each other when flowing in different axons at different times (Luck 2005: 29). Consequently, most of the EEG signal reflects postsynaptic potentials, which last tens to hundreds of milliseconds, and are mediated by several neurotransmitter systems (Luck 2005: 29; Lopes da Silva 2010: 20). Two types of postsynaptic potentials can be distinguished. First, in the excitatory potentials (EPSPs), the transmembrane current is carried by positive ions inwards, and are present when the action potential travels along the fiber that ends with an excitatory synapse (Sanei and Chambers 2007: 4). Second, inhibitory potentials (IPSPs) result from the fiber ending with an inhibitory synapse, and involve the transmembrane current carried by either negative ions inwards or by positive ions outwards (Sanei and Chambers 2007: 4; Lopes da Silva 2010: 20). Fig. 2 shows the neuron membrane potential changes, as recorded by intracellular microelectrodes. Action potentials present in the excitatory and inhibitory presynaptic fibers result in the EPSPs and IPSPs in the postsynaptic neuron, respectively.

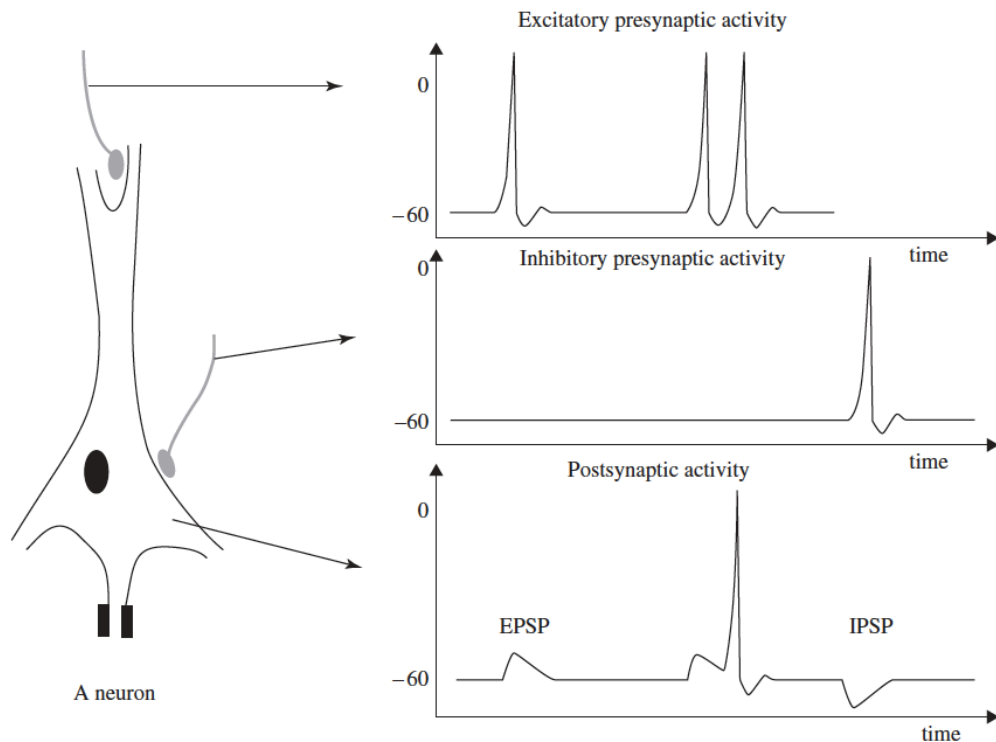


Fig. 2. Excitatory and inhibitory presynaptic and postsynaptic activity (after Sanei and Chambers 2007: 5).

Recording the neuronal activity with electrodes that are distanced from the source of the neuronal activity requires a large population of neurons, which should be temporally and spatially organized. Thus, the recorded EEG signal usually reflects the activation of pyramidal neurons, which are the most numerous, whose dendrites are parallel to each other, and which are perpendicular to the cortical surface (Lopes da Silva 2013: 1112). Importantly, the EEG signal always reflects electrical activity from large populations of neurons, estimated between 1000 to even 10,000 (Kaan 2007: 572).

1.4.2. Principles of EEG recording

The EEG signal is recorded from participants' scalp by means of using electrodes, whose role is to measure the electrical signal generated by the brain. Due to the fact that brain electrical activity is relatively small, with amplitudes up to around 100 μ V (Kamel and Malik 2015: 14), researchers usually apply from 32 up to 256 electrodes with a view to obtaining more precise data. Fig. 3 presents an electrode montage with 64 channels placed

larger numbers representing greater distance from the midline. While odd numbers indicate the left site, even numbers represent the right site of the scalp.

To record the EEG signal, two additional electrodes need to be placed on the scalp: the reference electrode (Ref) and the ground electrode (Gnd). While the former is used in order to differentiate between brain activity in response to an experimental stimulus as opposed to any other source of electrical activity (e.g., muscle movements), the latter is needed to eliminate static differences, and thus to enable the measurement of changes in the neural signal (Luck 2005: 103). EEG amplification systems usually use amplifiers with three electrodes: active (A), reference (R), and ground (G) electrodes so as to amplify the difference between the A-G voltage and R-G voltage.

When recording the EEG signal, researchers can use either active or passive electrodes. Unlike passive electrodes, active electrodes can amplify the electric current and reduce noise (Cencen et al. 2016). Noise observed in the EEG signal is defined as resulting from technical artifacts (power line) or biological artifacts. The biological artifacts usually come from eye movements (electrooculographic artifacts), muscles (electromyogram artifacts), or cardiac activity (electrocardiogram artifacts; Kamel and Malik 2015: 11).

In order to obtain a clean, artifact-free EEG signal, and with a view to alleviating the risk of increased skin potentials as well as a decreased common-mode rejection, it is recommended to keep the impedance of the skin below 5 k Ω , as electricity needs minimal resistance (Luck 2005: 117). To achieve and maintain such conductance between the skin and electrodes, an electro-conductive gel needs to be applied to each electrode before an experiment begins. The EEG signal obtained throughout the experiment can be analyzed by means of running several types of analyses, such as time-frequency or event-related potential (ERP) analyses. The following section of this chapter is further devoted to discussing an ERP analysis, which is performed on the recorded EEG signal, and which involves event-related potentials time-locked to a specific cognitive event (Kutas et al. 2006: 661). The ERP analysis was employed in the experiment reported in this dissertation (Chapter 4).

1.4.3. Event-related potentials (ERPs)

Event-related brain potentials are defined as the averaged electrical brain activity elicited in response to an external event that might be sensory, perceptual, cognitive or motor (Kutas et al. 2006: 661). From a neurobiological perspective, event-related potentials originate from electric signals generated by neurons, and are triggered by the thalamus, which plays a critical role in controlling sensory information that enters the cortex, and modulates information flow in motor, cognitive, and affective cortical regions (Kropotov 2009: 38).

Event-related brain potentials are computed by means of measuring the electrical signal at the scalp, with each electrode recording the electrical brain activity at its location (Kaiser 2013: 157). Differences in temporal intervals of waveforms reflect distinct functional processes as well as different anatomical locations of neurons contributing to the recorded signal (Kutas et al. 2006: 661). Importantly, individual ERPs are considerably smaller as compared with the ongoing EEG activity, and thus ERPs need to be analyzed based on a great number of trials. Therefore, ERPs need to undergo the averaging technique, which involves two steps of the averaging procedure: average within subject, and average between subjects (Jiang 2015: 73).

During within-subject averaging, trials are averaged separately for each participant and for each experimental condition. This type of averaging is required due to the fact that the EEG signal consists of both ERPs and noise. While ERPs, which are elicited in response to an external stimulus, are assumed to be very much similar in phase and shape, noise is usually random, and is not evoked in response to the time-locking event (Jiang 2015: 74). Consequently, due to averaging a number of trials, noise becomes minimized, whereas ERPs become more prominent, resulting in a higher signal-to-noise ratio (SNR).

Once the within-subject average has been computed, between-subjects averaging is used to create an averaged ERP waveform (grand average). The grand average is formed by means of averaging the already averaged waveforms of individual participants. Such a technique allows for an additional control for random noise left after the previously computed within-subject average (Jiang 2015: 76).

When analyzing the EEG signal in terms of event-related potentials, researchers focus on investigating specific ERP components associated with a particular cognitive

event. An ERP component is defined as “scalp-recorded neural activity that is generated in a given neuroanatomical module when a specific computational operation is performed” (Luck 2005: 59). ERP components can be divided into two different types: exogenous (e.g., P100, N200) and endogenous (e.g., N400, LPC) components. While the former components are fully automatic and are sensitive to the physical characteristics of a presented stimulus (e.g., its size), regardless of participants’ alertness and attentiveness, the latter components reflect cognitive operations elicited in response to a specific event or stimulus, and require conscious attention (Kutas et al. 2006: 662). ERP components can be labeled according to a number of factors: their polarity (either P-positive or N-negative), latency relative to stimulus onset (e.g., N100 as a negative component evoked 100 ms after stimulus onset), ordinal position (e.g., N3 as the third negative component), cognitive function (e.g., mismatch negativity), neural generator (e.g., auditory brainstem response), and scalp location (e.g., left anterior negativity as a component observed over left anterior sites). When analyzing particular ERP components, researchers often focus on their amplitude size and latency. From a neurobiological perspective, more pronounced ERP amplitudes mirror the activation of larger postsynaptic potentials, a greater number of neurons, and/or their temporally more synchronous activation (Kutas and Federmeier 2011). The following two subsections are devoted to discussing two language-related ERP components that are crucial for this thesis; namely, the N400 and the late positive complex (LPC).

1.4.3.1. The N400 and lexico-semantic access

The N400 is now one of the best-researched families of language-related ERP components. It pertains to a negative-going brainwave peaking in amplitude at around 400 ms after stimulus onset, which is usually observed over centro-parietal electrode positions. Research on brain-damaged populations that employs intracranially recorded ERPs has pointed to the left anterior medial temporal lobe as the main source of the N400 response (Luck 2005: 45; Kutas et al. 2006: 669; Kutas and Federmeier 2011: 629). For instance, in split-brain patients, whose corpus callosum is severed and therefore the connection between the two brain hemispheres is disrupted, a stimulus presented to the hemisphere

with productive speech capability (i.e., usually the left hemisphere) elicits the N400 response, unlike a stimulus presented to the right hemisphere (Kutas et al. 1988). This points to the left hemisphere as the largest source of the N400. Secondly, the role of the temporal lobe in the N400 response has been supported by observing delayed and attenuated N400 amplitudes in patients suffering from semantic dementia, which usually affects the anterior medial part of the temporal lobe, and results in a progressive loss of semantic knowledge (Mummery et al. 2000; Patterson and Hodges 2000). Importantly, a canonical N400 response is found in patients with damage over the frontal lobe only, indicating that the frontal lobe does not contribute to the N400 response (Swick et al. 1998; Friederici et al. 1999; Hagoort et al. 2003).

Previous research on the N400 in language processing has shown that the N400 response is sensitive to a number of both lexical and semantic variables. Lexical factors influencing the N400 amplitudes include word frequency, word vs. pseudoword vs. non-word status, vocabulary class, word concreteness, as well as orthographic neighborhood size (Kutas et al. 2006: 671). Namely, a more pronounced N400 response has been observed to low-frequency words compared to words that are high in frequency (Allen et al. 2003). Robust N400 amplitudes have also been found for pseudowords, but not for non-words, thus indicating that pseudowords, though not coded in the lexicon, might still contact semantic memory (Kounios and Holcomb 1994). Additionally, the N400 response has been found to be modulated by vocabulary class, with open-class words (i.e., content words) eliciting more pronounced N400 amplitudes compared to closed-class words (i.e., function words; Kutas and Hillyard 1983; Van Petten and Kutas 1991). This difference might, however, result from the abovementioned N400 frequency effect, as closed-class items are less frequent than function words, and thus they evoke an increased N400 amplitude. Furthermore, studies on the N400 component in language processing often report the N400 concreteness effect, showing a larger N400 response to concrete than abstract nouns (Paller et al. 1987). Such an effect is in line with the Dual-Coding Theory and its account of concreteness effect (Paivio 1991), according to which while the mechanisms engaged in concrete word processing are associated with the activation of both a verbal semantic memory store and a non-verbal image-based semantic memory store, abstract words involve the activation of the verbal semantic memory store only (Kutas et al. 2006: 675), which might be reflected in smaller N400 amplitudes for abstract than concrete words. Finally, the N400 response is of larger magnitude when evoked by words with

many orthographic neighbors (i.e., words which can be formed by changing only one letter) than when elicited by lexical items having fewer neighbors (Holcomb et al. 2002). This effect is likely to result from a parallel co-activation of orthographic neighbors, thus leading to a more robust N400 amplitude (Kutas et al. 2006: 675).

In addition to the lexical factors modulating the N400 response, the component has also been found to be sensitive to semantic variables, including plausibility and contextual constraints. Namely, larger N400 amplitudes have been observed for implausible compared to plausible words in a sentence (Kutas and Hillyard 1980; Kutas et al. 2006: 667). Additionally, previous research has shown that contextual constraints can modulate the N400 response, showing more pronounced N400 amplitudes for words presented in non-supportive relative to highly supportive contexts (Kutas and Hillyard 1984; Federmeier and Kutas 1999). Fig. 4 presents the grand average ERPs with the N400 response elicited in response to the critical words of different degree of plausibility.

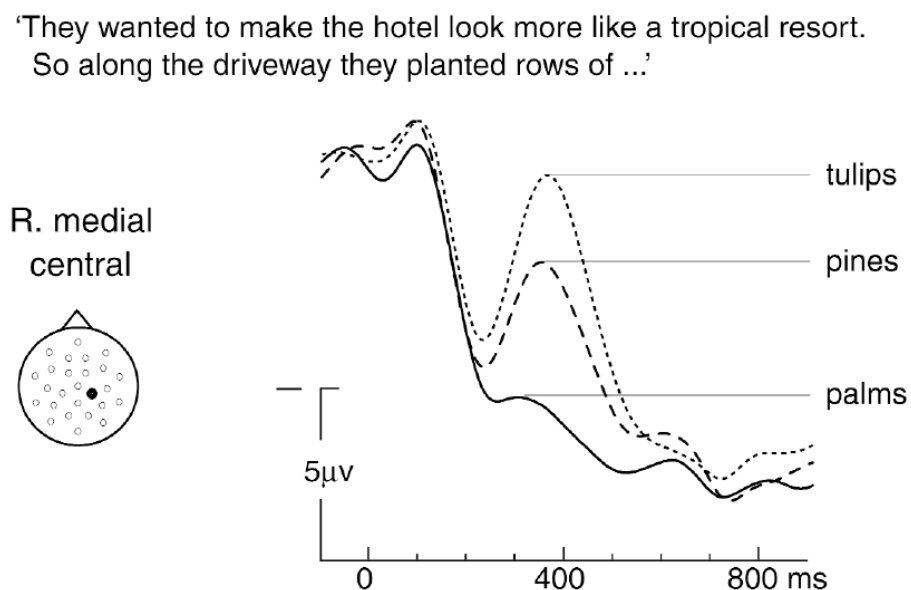


Fig. 4. Grand average ERPs in response to sentence-final critical words, which were expected items (i.e., *palms*), unexpected and implausible representatives of the same semantic category (i.e., *pines*), or unexpected and implausible representatives of a different semantic category (i.e., *tulips*). The presented N400 response varies as a function of plausibility, with the most pronounced N400 amplitudes for *tulips*, followed by *pines*, and *palms*. The effect was observed over the right-medial central electrode position (after Kutas and Federmeier 2000: 466).

Importantly for this dissertation, the N400 has also been examined in the contexts of nonliteral language comprehension as well as bilingual language processing. Previous monolingual research into metaphor comprehension has repeatedly shown that figurative utterances elicit a more pronounced N400 response compared to literal meanings (e.g., Coulson and Van Petten 2002; Lai et al. 2009; Regel et al. 2010). Additionally, studies conducted thus far have shown modulations in the N400 amplitudes for metaphoric utterances of different levels of conventionality, with novel metaphors evoking larger N400 amplitudes than conventional metaphors (e.g., Arzouan et al. 2007; Lai et al. 2009). Differences between novel and conventional metaphors observed within the N400 time frame might indicate that the N400 indexes the difficulty of mapping constructions, which pertains to forming relational correspondences between two concepts (Coulson and Van Petten 2002; Arzouan et al. 2007). In addition to metaphoric meaning comprehension, the N400 has been examined in the context of bilingualism with a view to investigating whether the component is sensitive to language nativeness. Previous research into the N400 response in bilingualism has shown a delayed or/and attenuated N400 response to the non-native language relative to the native tongue (e.g., Phillips et al. 2004; Moreno and Kutas 2005; Newman et al. 2012). Such results suggest that N400 amplitudes might also reflect the automaticity of the lexico-semantic access, as well as the strength of interconnectivity between lexical items within the semantic network.¹

Owing to the fact that the N400 response has been found to be sensitive to both semantic and lexical features of linguistic stimuli, the N400 is frequently interpreted as an index of lexico-semantic memory, with its magnitude reflecting the amount of information that is retrieved from long-term memory (Kotz et al. 2012). Importantly, the component shows how bottom-up and top-down information interact with each other during visual word recognition, which is in line with the Interactive Activation Model (McClelland and Rumelhart 1981), arguing for simultaneous operations engaged in sublexical and discourse-level processing. Bearing in mind the factors that the N400 response is sensitive to, much research has been devoted to establishing the exact functional role of the component (e.g., Brouwer et al. 2012; Schacht et al. 2014; Brouwer and Crocker 2017; Brouwer et al. 2017). As a result, two most prominent theories of the N400 have been put

¹ A more elaborate discussion on research into the N400 response in metaphoric meaning comprehension as well as in bilingual language processing is presented in sections 2.6 and 3.3.1, respectively.

forward: the Integration View (Brown and Hagoort 1993) and the Pre-Recognition View (Deacon et al. 2004).

The role of top-down mechanisms in the N400 response is best explained by the Integration View of the N400 (Brown and Hagoort 1993), which suggests that the N400 response is an indicator of semantic information being integrated with the provided context. This account has been supported in ERP experiments showing more pronounced N400 amplitudes for words presented in low constraining as compared to highly constraining contexts (e.g., Kutas and Hillyard 1984). The theory therefore stresses the role of discourse-level information in the N400 response, and points to top-down control mechanisms engaged in meaning comprehension as the most probable origin of the component.

Even though the Integration View accounts for high-level processes of semantic integration, it does not take into consideration the N400 sensitivity to lexical-level information, including orthographic neighborhood size, lexical frequency, orthographic and phonological similarity. As a result of these limitations, the Pre-Recognition View (Deacon et al. 2004) was proposed, which, unlike the Integration View, assumes that the N400 response reflects lexical retrieval from long-term memory. Namely, the N400 is postulated to mirror lexical processes involved in orthographic and phonological decoding, which take place before semantic analyses. The theory therefore indicates that the N400 response is indifferent to semantic features of linguistic stimuli.

Much as the two aforementioned theories account for certain linguistic factors that the N400 is sensitive to, they are both limited and thus neither of the views provide a clear-cut explanation of the precise functional role of the N400. The Integration View accounts for only top-down semantic features modulating the N400 response, and it therefore undermines the role of bottom-up lexical factors. The Pre-Recognition View, on the other hand, disregards semantic variables modulating the N400 effect. As a result of these limitations, the weaker version of the Pre-Recognition View (Brouwer et al. 2012) was proposed, which highlights the dominance of bottom-up processes in the N400 response, at the same time pointing to the context as excitatory power which adds to the activation pattern, and whose role is reflected in the anticipatory mechanisms present during language processing (Schwanenflugel et al. 1988; Federmaier and Kutas 1999; van Berkum 2009; Brouwer et al. 2012: 134).

Interestingly, when discussing the N400 response, some researchers have argued that the N400 reflects a family of components, and thus each subcomponent elicited within the N400 time frame should to be examined separately (e.g., Paller et al. 2007; Voss and Federmeier 2011; Bridger et al. 2012). For instance, some authors have observed the N400 response pronounced over frontal electrode positions, and have thus postulated that the frontal N400 (henceforth the FN400) might index different cognitive mechanisms than the canonical N400 response with a centro-parietal distribution (e.g., Mecklinger 2000; Paller et al. 2007; Rugg and Curran 2007). Namely, while the N400 is associated with lexico-semantic mechanisms, the FN400 is argued to reflect the old/new effect, therefore indexing familiarity-based recognition processes (Curran 2000; Rugg and Curran 2007; Bridger et al. 2012; Leynes et al. 2017). Thus, the N400 might reflect operations engaged during lexico-semantic access, whereas the FN400 might mirror memory processing (Voss and Federmeier 2011: 3). Importantly, the FN400 and N400 responses have thus far been only little compared within the same experimental settings. Such a limitation was addressed by Voss and Federmeier (2011), who employed a semantic-priming paradigm with a continuous recognition task, in which participants made two decisions regarding the presented word: a semantic judgment and a recognition memory judgment response. The authors found semantic effects within the N400 time frame that were indistinguishable from recognition effects in both its time-course and scalp distribution, thus suggesting that the FN400 correlates of memory might not be distinct from the N400 correlates of lexico-semantic processing.

The study conducted by Voss and Federmeier (2011) has, nonetheless, been criticized, as the semantic priming manipulation used in the experiment was embedded in a recognition test, as a result of which recognition contrasts might have been confounded by semantic operations (Bridger et al. 2012). In another study that examined the functional distinction between the N400 and FN400, Bridger et al. (2012) employed an encoding phase (an incidental learning phase) with a semantic priming paradigm and a semantic judgment task. The encoding phase was followed by a distractor phase, where participants completed a working memory and a visual oddball task. In an oddball task, participants responded to stimuli that were presented infrequently and irregularly (Huettel and McCarthy 2004: 379). In the final phase of the experiment conducted by Bridger et al. (2012), participants performed an old/new recognition test. Unlike Voss and Federmeier (2011), Bridger and colleagues (2012) observed the N400 semantic priming effect

over centro-parietal electrode sites, and the FN400 old/new effect with a fronto-central distribution, thus indicating that the FN400 might be functionally distinct from the N400. More research is, nonetheless, needed in order to further elucidate the functional role of the N400/FN400.

1.4.3.2. The late positive complex (LPC) and meaning integration

Another ERP component frequently discussed in the context of language processing is the late positive complex (LPC). The LPC is defined as a positive-going wave that peaks in amplitude at around 500-900 ms post stimulus onset (Friedman and Johnson 2000). Traditionally, the component used to be referred to as the P600, which has been associated with syntactic processing, as more pronounced P600 amplitudes have frequently been found in response to syntactic violations compared to grammatically correct utterances (e.g., Hagoort et al. 1993; Osterhout and Nicol 1999; Frederici et al. 2002). More recently, the component has also been observed to be modulated by semantic expectancy, thematic role and semantic reversal anomalies, semantic incongruencies on the sentence as well as discourse levels, and conceptual complexity (e.g., Gunter et al. 2000; Coulson and Kutas 2001; Coulson and Van Petten 2002; Regel et al. 2011; Spotorno et al. 2013; Rataj et al. 2018). This indicates that the P600/LPC response also mirrors cognitive operations engaged in semantic processing.

Interestingly, research into the role of the LPC in language comprehension has provided inconclusive results with regard to its scalp distribution. While some authors have reported the LPC response with a broad, posterior distribution (e.g., Münte et al. 1998; Severens and Hartsuiker 2009; Van de Meerendonk et al. 2010; Sanford et al. 2011; Szewczyk and Schriefers 2011), other researchers have found the LPC effect over the frontal sites (e.g., Federmeier et al. 2007; DeLong et al. 2011; Thornhill and Van Petten 2012). Consequently, it has been postulated that differences in the LPC scalp distribution might reflect distinct operations involved in language comprehension. A more posterior LPC effect has been suggested to be sensitive to language violations, while an anterior LPC response has been found to be evoked by plausible yet unpredictable words in constraining contexts (e.g., De Grauwe et al. 2010). Therefore, the anterior LPC might indicate the processes of reanalysis, which are needed in order to arrive at a correct final

interpretation of an input, and thus it may reflect additional working memory processes engaged in meaning integration. It therefore seems that not all LPC effects are generated alike, and differences in the LPC amplitude, peak latency, duration, as well as scalp distribution are likely to stem from distinct sub-processes underlying meaning integration (Brouwer et al. 2012). Fig. 5 presents the grand average ERPs with the posterior P600/LPC effect in response to grammatically correct and incorrect sentence-final critical words.

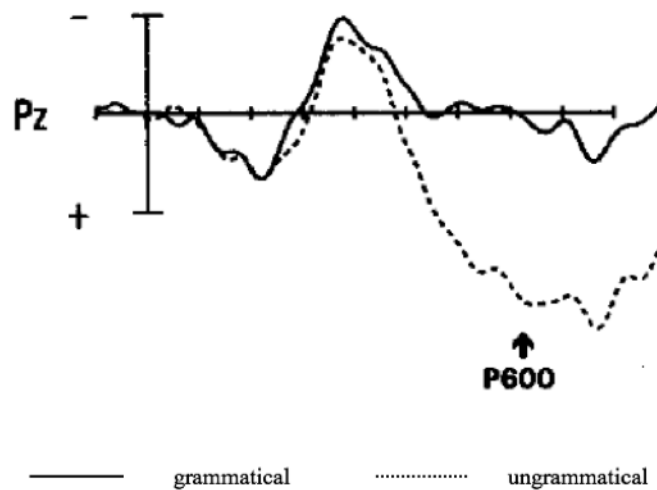


Fig. 5. Grand average ERPs in response to sentence-final critical words, which were either grammatical or ungrammatical. The presented P600/LPC response varies as a function of grammatical correctness, with more pronounced P600/LPC amplitudes for grammatically incorrect compared to grammatically correct words. The effect was observed over the Pz electrode (after Osterhout and Nicol 1999: 299).

Notwithstanding the inconsistencies regarding the scalp distribution of the LPC, the component is generally linked to rechecking and reprocessing operations which take place in response to an unexpected linguistic stimulus, and it therefore reflects an update of the current mental representation by incoming information (van Herten et al. 2005; Brouwer et al. 2012; Rataj 2014). The LPC has therefore been indicated to reflect mental representation composition (the MRC view; Brouwer et al. 2012). The MRC view claims that the construction of the mental representation of what is being communicated is based on logical, causal, as well as pragmatic inferences that are dependent on pragmatic skills and world knowledge that a language user possesses (Brouwer et al. 2012: 136). The

magnitude of the LPC thus indexes cognitive demands engaged in these operations together with additional working memory processes that are required during meaning integration (Brouwer et al. 2012: 137). Therefore, while the N400 response reflects memory retrieval phase that involves the activation of lexico-semantic information from long-term memory, the LPC mirrors operations engaged in revising and integrating this activated information with the provided context (Brouwer et al. 2012).

1.4.4. Strengths and limitations of electrophysiological research

Electrophysiological research into language comprehension offers a lot of advantages, the major one being its temporal resolution. Owing to the fact that cognitive mechanisms engaged in language processing occur within tens of hundreds of milliseconds, only methods having a great temporal resolution are capable of capturing fast, dynamic, and temporally sequenced cognitive events associated with language comprehension. Furthermore, the recorded EEG signal is multidimensional, as the EEG data consist of at least four dimensions: time (voltage changes over time), space (voltage changes over different electrode sites), frequency and power (the strength of frequency-band-specific activity), as well as phase (timing of this activity). An event-related potential analysis seems to be particularly advantageous when studying language processing, as ERPs can be employed to investigate cognitive mechanisms taking place between the presentation of a linguistic stimulus and a response given (Luck 2005: 21). Therefore, unlike behavioral measures, ERPs allow for determining the precise stage of processing influenced by a specific experimental manipulation, even in the absence of a behavioral response.

Notwithstanding these strengths, ERPs also impose some limitations. Compared to behavioral measures (e.g., reaction time methods), conducting an EEG experiment is more expensive, and running ERP analyses is by far more complex, as they require identifying and removing trials contaminated by artifacts, extracting event-related potentials from the overall EEG recording by means of averaging, and finally isolating specific ERP components by performing signal processing techniques (Luck 2005: 12). Furthermore, in order to obtain interpretable data with a low signal-to-noise ratio, ERP experiments require a great number of trials, and are therefore more time-consuming compared to re-

action time studies (Kaiser 2013: 159). Another weakness comes from physiological phenomena including blinks, eye movements, and muscle activity, all of which influence the EEG signal and can therefore mask ERP components (Kaiser 2013: 159). This problem is usually addressed by asking participants to try not to move during the experiment, and to avoid blinking during certain times. Much as such procedures help to obtain high-quality data, they decrease the ecological validity of the study and make the experimental setting more artificial. Finally, ERPs offer a poor spatial resolution due to the fact that electric potentials are distorted by brain fluids as well as tissue and skull irregularities (Kaan 2007: 576). As a result, although ERPs can provide information of exactly when the activity occurs, they are less informative as to where in the brain it is generated.

1.5. Research methods employed in the present dissertation

The main objective of the two experiments described in this thesis was to examine metaphoric language comprehension in the context of bilingualism. To this end, three quantitative research methods were employed: survey research, as well as behavioral and electrophysiological measures. This section provides rationales behind employing each of these research techniques.

Survey research was employed in the normative studies in both experiments reported in this thesis (Experiment 1: Chapter 4; Experiment 2: Chapter 5), and involved self-administered web-based surveys, whose aim was to ensure that experimental stimuli were adequate representatives of categories ascribed to them. Normative studies on experimental materials have frequently been conducted in research on figurative language. Previous studies have often employed rating scales to investigate non-literal utterances in terms of their meaningfulness, familiarity, imageability, naturalness, as well as figurativeness (e.g., Ferretti et al. 2007; Balconi and Amenta 2010; Cardillo et al. 2010; Bambini et al. 2013; Cieślicka et al. 2014; Schneider et al. 2014; Forgács et al. 2015; Rataj et al. 2018). In order to ensure that the experimental materials designed to be included in Experiment 1 and Experiment 2 were perceived as intended, they were pretested on rating scales on concreteness, meaningfulness, familiarity, and metaphoricity. In addition to the rating scales, cloze probability tests were conducted in order to confirm that all of the

utterances used in the experiments were embedded in low constraining contexts, and therefore the critical words were not expected.

Owing to the limitations imposed by rating scales with fewer than 5 or more than 9 categories (see section 1.2.1), all of the conducted surveys included 7-point rating scales. Furthermore, they all consisted of items that fell into both extreme categories, so that the threat of response-set tendency could be minimized. For example, in a survey on word concreteness, both abstract and concrete nouns were included. Importantly, items used in both rating scales and cloze probability tests were always presented in a randomized order, so that each participant could be provided with a different order of utterances, and the potential presentation-order effect could be ruled out. Instructions along with the scales used in the normative tests are provided in Appendices G – J (pp. 269-278).

In addition to survey research, reaction times measures were also employed in the two experiments reported in this thesis. A vast majority of studies investigating figurative language comprehension have employed reaction times in order to test competing theories of nonliteral meaning interpretation, and to provide evidence whether and when literal meanings enjoy the temporal priority over figurative utterances (e.g., McElree and Nordlie 1999; Gentner et al. 2001; Giora 2002; Glucksberg 2003; Camp 2006; see Chapter 2). Additionally, RT methods have been employed to investigate bilingual language processing, and thus to examine the language selective as compared to language non-selective access view (e.g., Ransdell and Fischler 1987; de Groot et al. 2002; Duyck et al. 2008; Gollan et al. 2011; see Chapter 3). The reaction times method was employed in Experiment 1 and Experiment 2, and involved a two-choice semantic decision task, where participants decided whether the utterances presented in their native and non-native language were meaningful or not, at which point response latencies were measured.

Finally, event-related brain potentials were employed in Experiment 1. Due to the fact that ERPs can capture a cognitive operation in the precise time frame in which it takes place (Cohen 2014: 15), they have frequently been investigated in studies on monolingual figurative language comprehension (e.g., Lai et al. 2009; De Grauwe et al. 2010; Lai and Curran 2013; see section 2.6). While behavioral measures (i.e., reaction times) provide information regarding how much time participants need in order to comprehend a specific linguistic stimulus, ERPs show brain responses that directly follow the presentation of a stimulus, thus providing insights into specific stages of nonliteral language comprehension. In studies on figurative meaning interpretation, researchers usually focus

on the N400 and LPC responses, which reflect mechanisms involved in lexico-semantic and integrative operations, respectively. Therefore, unlike behavioral measures, ERPs can be used to examine how the meaning of a nonliteral utterance is accessed and then integrated. Additionally, event-related potentials have previously been employed to investigate bilingual language processing, with a view to showing how the brain computes meaning in the native and non-native tongue, and to examining cognitive mechanisms engaged in target language selection and non-target language inhibition (see section 3.3). Due to the aforementioned strengths, Experiment 1 employed the EEG method along with the ERP analysis in order to examine the precise time course of novel and conventional metaphoric meaning comprehension in bilingualism.

Experiments discussed in this thesis were aimed to make use of the strengths that the aforementioned research methods offer. Rating scales were therefore employed in the normative tests in order to investigate predictability, concreteness, meaningfulness, familiarity, and figurativeness of the experimental stimuli, as all of these features have been found to modulate mechanisms engaged in language processing (e.g., De Grauwe et al. 2010; Rataj et al. 2018). Collecting data from web-based rating scales ensured that the experimental materials were suitable for inclusion in the experiments. Reaction time measurement enabled the investigation of the amount of time bilingual speakers needed in order to decide whether the utterances presented in their native and non-native language were meaningful or meaningless. Finally, event-related brain potentials, which were employed in Experiment 1, aimed to examine the precise stages of metaphoric meaning comprehension in bilingualism.

1.6. Concluding remarks

This chapter discussed three methods that can be employed in quantitative research on language processing; namely, survey research, behavioral measures, as well as event-related brain potentials. Each of these methods offers both strengths and limitations, and thus it is crucial to employ them according to a particular research question that they aim to address. Survey research based on rating scales, though not experimental in its nature, is especially useful when selecting appropriate materials to be included in the experiment proper. Reaction time measures, as an experimental method that is easy and relatively

cheap to employ, are beneficial in providing information regarding the amount of time participants need to respond to a linguistic stimulus. Importantly, though reaction times reflect cognitive demands engaged when processing a particular utterance, they do not allow for examining the specific time course of language processing, and cannot be used to investigate exact mechanisms that govern lexico-semantic access and meaning integration operations taking place before the response is given. These processes can be illuminated by means of employing event-related potentials, which pertain to an average brain response to a linguistic stimulus with an excellent temporal resolution. The following theoretical chapters further discuss reaction time and ERP experiments into metaphoric meaning comprehension (Chapter 2), and bilingual language processing (Chapter 3).

Chapter 2: Metaphoric meaning comprehension

2.1. Introductory remarks

It is assumed that figurative language reflects the poetics of mind (Gibbs 1994a), and thus nonliteral utterances might be used as a tool to examine human creativity by showing how the brain constructs new meanings. This chapter is devoted to discussing mechanisms engaged in metaphor comprehension, as the two experiments reported in this thesis aimed at examining how bilingual speakers compute metaphoric meanings in their native and non-native tongue. Due to the fact that thus far only little attention has been devoted to examining metaphor comprehension in the context of bilingualism, the chapter provides a review of theoretical accounts that elucidate cognitive mechanisms engaged in monolingual metaphoric language comprehension. Furthermore, it discusses factors that have been postulated to modulate processes involved when comprehending metaphoric utterances in the native tongue, such as, for instance, contextual cues and metaphor conventionality. Finally, the chapter presents behavioral as well as electrophysiological research that has thus far been conducted in the field of monolingual metaphoric meaning comprehension.

2.2. Figurative language

Figurative language is defined as involving utterances whose intended meanings do not correspond to their literal interpretations (De Grauwe et al. 2010: 1965; Gibbs and Colston 2012: 1). Owing to the fact that the above definition implies a contrast between

figurative and literal meanings, one of the most crucial points discussed when studying figurative meanings is the actual distinction between literal and nonliteral language. The following subsection is devoted to providing a clear definition of literal and nonliteral meaning that would disentangle these two notions, as this is often perceived as a starting point in further attempts to examine how figurative language is processed in the brain.

2.2.1. From literal to figurative language

The literature discussing the distinction between literal and nonliteral meanings often provides a binary view, where literal language is contrasted with figurative language so that everything that is not literal is figurative (Weiland et al. 2014: 1). Consequently, nonliteral language has traditionally been viewed as a deviation from everyday literal language (Grice 1975). On the other hand, according to less extreme positions, literal meaning represents the meaning of a word stored in the mental lexicon, which is contextually free (Weiland et al. 2014: 1). The integration of a whole figurative sentence, in contrast, requires semantic integration operations, which are not necessarily restricted to the literal interpretation (Gibbs 2002; Récanati 1995). A more elaborate discussion on literal compared to figurative meanings was provided by Gibbs (1994b: 75), who proposed four approaches to differentiating between literal and figurative language. According to the view of conventional literality, figurative language is mostly represented in poetry, and is characterized by indirect and often exaggerated utterances, unlike literal statements, which are hypothesized to convey true meanings. This assumption is further elaborated on in the concept of non-metaphorical literality as well as truth conditional literality, both of which state that figurative utterances never express true meanings, and are therefore always unreliable and subjective. Finally, as argued by context-free literality, literal language is always independent of the context in which it is embedded, unlike figurative language, whose interpretation is argued to be modulated by both verbal and nonverbal contextual cues (Gibbs 1994b, as cited in Evans and Green 2006: 287). Consequently, figurative meanings have been postulated to require the interpretation of not only what is implicated, but also what is pragmatically said.

Other relevant distinctions discussed when differentiating between literal and figurative language are the continuity claim and equivalence claim (Coulson and van Petten

2002: 959). The continuity claim argues for a parallel activation of figurative and literal interpretations, and it therefore postulates a similar time-course of literal and nonliteral meaning comprehension. The equivalence claim, on the other hand, suggests that the comprehension of both nonliteral and literal meanings involves the same cognitive effort. With a view to testing these two assumptions, Coulson and Van Petten (2002) conducted an ERP study², and found evidence in favor of the continuity claim, yet against the equivalence assumption, as larger N400 amplitudes were observed in response to metaphors compared to literal utterances presented in both highly and low constraining contexts. This indicates that although the time-course of mechanisms engaged in literal and figurative meaning interpretation might be very much similar, figurative language comprehension is likely to be more cognitively taxing. Both types of meanings might therefore undergo the process of mapping between different domains, but this might be more resource intensive in the case of figurative compared to literal language comprehension.

An interesting approach to discussing literal and nonliteral language was proposed by Giora (2002) in her Graded Salience Hypothesis (see section 2.3.4). Instead of perceiving figurative and literal meanings as different ends of a continuum, the view claims that it is meaning salience that modulates how language is processed, be it literal or nonliteral. In line with the hypothesis, a salient meaning (i.e., a meaning that is coded in the mental lexicon) is always processed preferentially, regardless of its figurativeness or context in which it is embedded. Consequently, if literal and figurative expressions are equally salient, their comprehension involves similar initial processes.

A clear distinction between literal and figurative language has been perceived as a cornerstone when proposing further theoretical accounts of figurative language comprehension. A discussion on literal and nonliteral meaning processing has resulted in the arousal of the most prominent research in studies on nonliteral meaning comprehension, which refers to whether figurative meanings are activated in a parallel or serial manner (Tartter et al. 2002: 488). While the parallel (direct) processing accounts argue that figurative and literal interpretations are activated simultaneously, serial (indirect) access views point to one interpretation (usually the literal) activated preferentially before the other (usually the figurative). Early research on figurative meaning comprehension fol-

² A more detailed description of the ERP experiment conducted by Coulson and Van Petten (2002) is provided in section 2.6.2.

lowed the serial processing account, which was strongly reflected in one of the most traditional perspectives on nonliteral language processing, the Standard Pragmatic View (Grice 1975), which is discussed in the following section.

2.2.2. A traditional perspective on figurative language comprehension: The Standard Pragmatic View

The serial processing view on figurative language comprehension was supported in the Standard Pragmatic View (Grice 1975), according to which a literal meaning needs to be accessed before a figurative interpretation becomes available, as figurative language pertains to the violation of language use. The theory highlights the role of conversational maxims, which require speakers to be informative, truthful, relevant, and clear when communicating with each other (Grice 1975). These maxims are, however, violated when using figurative utterances, which cannot be interpreted literally, and therefore their comprehension requires both semantic and pragmatic information analyses. Due to the fact that the Standard Pragmatic View claims that semantic processing operates prior to pragmatic analyses, arriving at a figurative interpretation is possible only once several steps have been accomplished. Namely, speakers are supposed to activate the literal meaning of an utterance, check whether it is plausible in the provided context, and only if the literal interpretation turns out to be implausible do speakers arrive at an alternative, figurative interpretation (Grice 1975, as cited in Gibbs 1994a: 420). As a result of these serial steps, nonliteral language interpretation is argued to be more difficult and more time consuming, as it requires the preliminary activation and rejection of the literal interpretation before the figurative meaning is accessed.

However, the Standard Pragmatic View has undergone strong criticism (see Palinkas 2013 for a review). First of all, it does not specify whether figurative language comprehension could be modulated by any additional factors, such as the conventionality of a presented nonliteral utterance or the context in which the utterance is embedded. The level of conventionality reflects how frequent and familiar a given utterance is, based on which figurative meanings are often labeled as either conventional (i.e., familiar and frequent) or novel (i.e., unfamiliar and infrequent; Lai et al. 2009: 145). Novel figurative

language is assumed to be more difficult to interpret, as it requires extra referential processes, and thus novel meaning creation is considered to be more resource intensive and more time-consuming (Gibbs and Colston 2006: 839). Conventional utterances, due to their familiarity and frequency of use, are, however, much easier to comprehend. As a result of such a dichotomy within figurative language, the notion of literal as compared to nonliteral meaning ought to be perceived as a continuum, with some utterances being more literal and other more figurative in their nature (Gibbs and Colston 2012: 19). Another limitation of the Standard Pragmatic View pertains to the fact that it does not account for the role of context in nonliteral meaning processing. Previous research has shown that highly constraining contexts facilitate anticipatory mechanisms and, consequently, words embedded in such contexts are easier to be processed compared to lexical items presented in low constraining contexts (Fischler and Bloom 1985). Therefore, nonliteral meaning comprehension is likely to be modulated by contextual cues, which can lead to a faster activation of a figurative meaning. Finally, due to the fact that figurative language encompasses a number of different types of nonliteral utterances (for example, metaphor, irony, idiom, and metonymy), the Standard Pragmatic View might be too simplistic to provide a thorough explanation of the processes engaged when interpreting different types of figurative expressions.

Additionally, behavioral research has contradicted the serial processing account, according to which a literal meaning is always preferentially accessed before a figurative meaning becomes available. In their seminal study, McElree and Nordlie (1999) instructed their participants to perform a semantic decision task in response to literal, metaphoric, and anomalous utterances (e.g., *Some temples are stone*; *Some hearts are stone*; *Some clouds are stone*, respectively). In contrast to what the Standard Pragmatic View claims, the results revealed similar reaction times for literal and metaphoric sentences. In order to further examine whether the obtained results might have been influenced by task specificity, in Experiment 2, participants decided whether the presented sentence was literally true or not. Similarly, the results showed no differences in the time participants needed to reject a figurative and an anomalous sentence, which indicates that figurative meanings might be accessed in parallel with literal interpretations.

The abovementioned criticism towards the Standard Pragmatic View has resulted in more recent views aimed at explaining processes involved in figurative meaning comprehension. The following section is devoted to reviewing such approaches, and will

mostly focus on how metaphors are processed, as the main objective of the two experiments described in this dissertation was to elucidate cognitive mechanisms engaged in bilingual metaphor comprehension.

2.3. Selected theories and models of metaphoric meaning comprehension

One of the most frequently investigated types of figurative language are metaphors, whose interpretation is postulated to be based on cross-domain mappings, which involve recognizing common features of two presumably distinct concepts (Gibbs and Colston 2012: 129). For instance, in order to understand a metaphor such as *Lawyers are sharks*, a speaker needs to find shared properties of *lawyers* (a metaphor topic/target domain) and *sharks* (a metaphor vehicle/base/source domain). While a metaphor topic pertains to the concept to which a metaphor refers, a metaphor vehicle denotes a basis for understanding the target concept (Su et al. 2017: 300). During the cross-domain mapping construction, speakers therefore need to establish links between target and source domains, which are relatively dissimilar from each other (De Grauwe et al. 2010: 1965). The question how such links are formed has been reflected upon in a number of theoretical accounts as well as models of metaphoric language comprehension.

The following sections discuss the most prominent theories and models that elucidate mechanisms engaged in metaphor comprehension. Theoretical formulations include conceptual descriptions, operational definitions, and propositions, based on which they allow for formulating testable hypotheses that can explain a given phenomenon (Content and Frauenfelder 1996: 3; Karakaş and Başar 2006: 186). A model, on the other hand, pertains to a verbal, visual, or mathematical representation of reality, which can be used to test some theories by means of running simulations. Consequently, while a theory provides a descriptive answer to the questions *what/why*, a model shows *how* by means of presenting underlying mechanisms of a phenomenon under investigation (Content and Frauenfelder 1996: 2).

The first five sections of the following part of the chapter (section 2.3.1 – 2.3.5) are devoted to discussing theoretical accounts, while the two last sections (section 2.3.6 – 2.3.7) focus on models of metaphoric meaning comprehension, thus explaining the underlying mechanisms of metaphor processing. Importantly, sections 2.3.1 – 2.3.4 discuss

views which have originally been proposed to explain how various types of figurative language (including not only metaphor, but also irony, proverb, and metonymy) are comprehended. Sections 2.3.5 – 2.3.7, on the other hand, review accounts which have specifically been proposed to explicate cognitive mechanisms involved in metaphoric meaning comprehension.

2.3.1. The Relevance Theory

The Relevance Theory (Sperber and Wilson 1986) claims that similar mechanisms are engaged in both literal and figurative language comprehension, and these mechanisms are assumed to be modulated by meaning relevance. Namely, listeners are postulated to pay attention only to information that they find relevant, and they continue processing the meaning of an utterance until they begin to consider that any further analyses are not worth cognitive effort (Sperber and Wilson 1986, as cited in Gibbs and Colston 2006: 848). An important point discussed in the Relevance Theory is that human cognition is geared towards relevance maximization, which is reached when an interpretation is worth people's attention, and does not require cognitive load that is unnecessarily high (Sperber and Wilson 1986, as cited in Gibbs and Colston 2012: 100).

In line with the Relevance Theory, when interpreting figurative language, listeners process various possible interpretations until the optimal relevance is achieved, as a result of which nonliteral meaning comprehension might be sometimes more cognitively taxing compared to literal language comprehension. Yet, the theory highlights the fact that there is no specific mechanism engaged in metaphor comprehension that would differ from the cognitive mechanism involved in literal meaning comprehension. Instead, the Relevance Theory introduces the notion of *ad hoc* concepts, which pertain to non-lexicalized, pragmatically-derived meanings, which are constructed during nonliteral language comprehension (Sperber and Wilson 1986, as cited in Carston 2010: 170). For example, in a metaphor *Robert is a bulldozer**, *bulldozer* acquires *ad hoc* properties (referred to with a symbol ‘*’), which are derived from properties associated with the encoded concept of a bulldozer (that is, powerful, crushing, going ahead regardless of obstacles). Thus, the denotation of *bulldozer** is inferred from a wider denotation of *bulldozer* with its encyclopedic properties. Once the lexical form of *bulldozer** is recognized, it activates a lexical

concept *bulldozer* along with associated encyclopedic information regarding bulldozers. Both the explicature (what is said) and implicatures (what is meant) are derived in parallel, with the processes of mutual adjustments operating between them until the interpretation meets an optimal relevance:

(6) Explicature: Robert is a bulldozer*

(7) Implicatures: Robert is powerful.

Robert is crushing.

Robert goes ahead regardless of obstacles.

(Carston 2010: 170)

Importantly, the Relevance Theory differentiates between novel and conventional nonliteral language comprehension by assuming that conventional figurative statements communicate at least one strong implicature, which is essential for an interpretation that satisfies the expectations of relevance. Novel nonliteral meaning comprehension is, on the other hand, more resource intensive due to the fact that novel utterances convey only weak implicatures, which help when constructing relevant interpretations, but are not considered unique, as other implicatures are also suggested by the utterance (Sperber and Wilson 1986, as cited in Gibbs and Colston 2012: 101). Additionally, the Relevance Theory emphasizes the role of context, which can facilitate figurative language comprehension, at the same time reducing the degree of cognitive load engaged when arriving at a nonliteral interpretation. Namely, the theory states that while a highly constraining context can speed up the mechanisms involved in nonliteral language processing, low constraining context can impede these operations to the point that a figurative interpretation may not be even considered. It needs, however, to be noted that the tenets discussed within the Relevance Theory have not been sufficiently tested empirically. In spite of this limitation, the theory is one of the first to rightly emphasize the role of contextual cues in figurative meaning interpretation, which is further discussed in the Constrains Satisfaction View (section 2.3.2) as well as the Underspecification View (section 2.3.3).

2.3.2. The Constraint Satisfaction View

The Constraint Satisfaction View (Katz and Ferretti 2001) states that nonliteral meaning interpretation is highly modulated by both verbal and nonverbal context in which a figurative utterance is embedded. In line with this view, when comprehending figurative language, people follow an interpretation that is better fitted, given the context, than other interpretations. According to the theory, the process of figurative language comprehension is additionally modulated by the level of conventionality, with conventionalized meanings accessed in a parallel manner, and novel meanings requiring serial processing steps, with a preliminary activation of a literal interpretation due to weak lexico-semantic representations of such utterances.

With a view to showing how context and meaning conventionality influence figurative language comprehension, and thus to testing the Constraint Satisfaction View, Katz and Ferretti (2001) examined the comprehension of familiar and unfamiliar proverbs in a self-paced moving window paradigm. Importantly, the proverbs were embedded in contexts that were biasing towards either a literal or a figurative interpretation. The results showed shorter reading latencies for familiar compared to unfamiliar proverbs, and for those stimuli that were embedded in figurative relative to literal contexts. Such results therefore provide support for the Constraint Satisfaction View, showing that contextual cues modulate nonliteral meaning interpretation, and suggesting that semantic operations are not autonomous, but instead, they might be modulated by contextual constraints.

The Constraint Satisfaction View was also supported in a study by Ivanko and Pexman (2003), who presented their participants with literal (e.g., *Tina is a lazy helper*) and ironic (e.g., *Tina is a super helper*) statements embedded in neutral or literal contexts. The results revealed similar reading times for both literal and ironic sentences embedded in the neutral context, and longer reading times for ironic sentences when embedded in the literal context. Such results therefore emphasize the role of contextual cues in nonliteral meaning comprehension, and are against the serial processing account, according to which the comprehension of figurative meanings is always more time consuming relative to literal language comprehension.

2.3.3. The Underspecification View

Similarly to the Constraint Satisfaction View, the Underspecification View (Frisson and Pickering 2001) assumes that the context highly modulates nonliteral meaning comprehension. As explained by Frisson and Pickering (2001), when encountering a figurative utterance, its meaning is initially underspecified with respect to its either literal or figurative sense. Arriving at an intended meaning is, nonetheless, possible owing to the context, which leads to the correct interpretation. When the context is strong enough, it can facilitate the intended meaning comprehension, be it literal or figurative.

The tenets proposed in the model were tested in an eye-tracking experiment carried out by Pickering and Frisson (2001). The researchers recorded participants' eye movements in response to ambiguous words, which conveyed either their dominant sense (for example, *After the capture of the village, we disarmed almost every rebel and sent them to prison for a very long time*) or a subordinate sense (for example, *With his wit and humor, the speaker disarmed almost every critic who was opposed to spending more money on art*). The results showed that sentences with an ambiguous word denoting its subordinate sense were more difficult to process, as reflected in longer total reading times. Importantly, such a difficulty emerged only after the presentation of that critical word. These results therefore indicate that the meaning of ambiguous items might be initially underspecified, but the available context helps to change the underspecified meaning into a specified interpretation. Nonetheless, the Underspecification View has received some criticism, mostly due to the fact that it fails to specify what really constitutes the underspecified meaning (Frisson 2009). Furthermore, the view does not account for how polysemous words change their sense from underspecified to specified, as some of them can capture even a hundred meanings (Gibbs and Colston 2012: 94).

2.3.4. The Graded Salience Hypothesis

Even though the three aforementioned theoretical accounts highlight the role of contextual cues in nonliteral meaning comprehension, its importance is undermined in the Graded Salience Hypothesis (GSH; Giora 2002), which proposes that nonliteral language comprehension is modulated by its salience. Namely, the theory argues for a preferential

activation of the most salient meaning, which pertains to the most frequent, familiar, conventional, and prototypical use of a presented utterance. Unlike the Standard Pragmatic View (Grice 1975), the GSH claims that the preliminary analysis of an expression's literal meaning is not always necessary, as sometimes it is a nonliteral meaning that is higher in salience than the literal interpretation. The hypothesis also accounts for conventionality-related differences in metaphor comprehension, and claims that while conventional metaphors are usually highly salient, novel metaphors are low in salience, and are therefore processed sequentially with more salient meanings being accessed first. Importantly, the theory assumes that salient meanings are always accessed preferentially, regardless of whether they are compatible with the provided context or not.

Interestingly, the GSH also hypothesizes hemispheric differences in salient as compared to non-salient meaning comprehension. Namely, even though both left and right brain hemispheres contribute to language processing, their involvement is argued to be modulated by meaning salience, with a more robust right hemisphere contribution during non-salient meaning comprehension (Giora 2003: 53). Such an assumption is in line with the Fine-Coarse Semantic Coding Theory (Beeman 1998; Jung-Beeman 2005), according to which exposure to a linguistic stimulus is directly followed by the left-hemisphere engagement in fine semantic coding, which results in the activation of semantically related items, as well as the right-hemisphere engagement in coarse semantic coding, which is associated with the activation of a large semantic field with more distant associations and alternative meanings.

Such assumptions have been supported in previous research employing the divided visual field paradigm (Giora et al. 2000; Giora 2003). In the divided visual field (DVF) paradigm, a stimulus is presented to one visual field, and is initially processed by the contralateral hemisphere. Namely, a stimulus presented to the right visual field is initially processed by the left hemisphere, while a stimulus presented to the left visual field is processed by the right hemisphere. Consequently, the DVF provides information regarding hemispheric specializations for specific aspects of language processing (Bourne 2006: 374). Previous DVF studies on semantic processing have pointed to right hemisphere involvement in the processing of less frequent meanings of homonyms, as in the case of the word "bank" referring to its "riverside" meaning (Chiarello 1998; Faust and Chiarello 1998; Faust and Kahane 2002). Similarly, while conventional (salient) metaphors have been observed to preferentially engage the left hemisphere, novel (non-salient)

metaphors triggered a more robust activation in the right hemisphere (Giora 2003: 143). This hypothesis was also strongly supported in a series of clinical studies conducted by Giora et al. (2000), who tested right- and left-brain damaged patients as well as a healthy control group in how they comprehended novel and conventional metaphors. The results revealed an impeded comprehension of conventional metaphors in left-brain damaged individuals, and of novel metaphors in right-brain damaged patients, which highlights the role of the left hemisphere in salient meaning processing, and of the right hemisphere in unfamiliar, non-salient language comprehension.

Notwithstanding the evidence in favor of the Graded Salience Hypothesis, the theory might be criticized as it strongly undermines the role of context in language comprehension. The assumption that salient meaning is accessed regardless of the context is difficult to account for within numerous studies suggesting that even the smallest contextual cues influence mechanisms engaged in nonliteral meaning comprehension (see section 2.6.1). Additionally, though the GSH strongly focuses on the distinction between salient and non-salient meanings in figurative language processing, it fails to specify what cognitive mechanisms govern novel and conventional meaning comprehension. Finally, the theory does not explain whether any factors apart from meaning salience could modulate how novel and conventional figurative meanings are comprehended.

2.3.5. The Conceptual Metaphor Theory

While the aforementioned views on nonliteral meaning comprehension were postulated with a view to elucidating how people comprehend various types of figurative expressions, the Conceptual Metaphor Theory (CMT; Lakoff and Johnson 1980) was proposed to specifically account for metaphor comprehension. The CMT is one of the best-grounded theories of metaphor comprehension within the field of cognitive linguistics, with its fundamental tenet implying that thought is metaphorical in its nature. As put by Lakoff and Johnson (1980: 6), “[m]etaphors as linguistic expressions are possible precisely because there are metaphors in a person’s conceptual system”. Metaphoric language comprehension is therefore based on the premise that the human conceptual system involves basic conceptual metaphor correspondences, and thus any metaphoric utterance is comprehended through conceptual metaphors (Lakoff and Johnson 1980, as cited in

Gibbs and Colston 2012: 144). Consequently, all metaphoric utterances are hypothesized to derive from a set of conventional conceptual metaphors that are often motivated by image schemas referring to an interaction between the human body and the world (McGlone 1996: 547; Górska 2009: 82).

For instance, as shown in Fig. 6, a metaphor *Our marriage was a rollercoaster ride* is understood by means of activating a reference to a conceptual metaphor LOVE IS A JOURNEY. Metaphor comprehension is further postulated to require cross-domain mappings from concrete metaphor source (*a journey*) to abstract metaphor target (*love*), which are based on pre-conceptual embodied experience (Lakoff and Johnson 1980, as cited in Evans and Green 2006: 286).

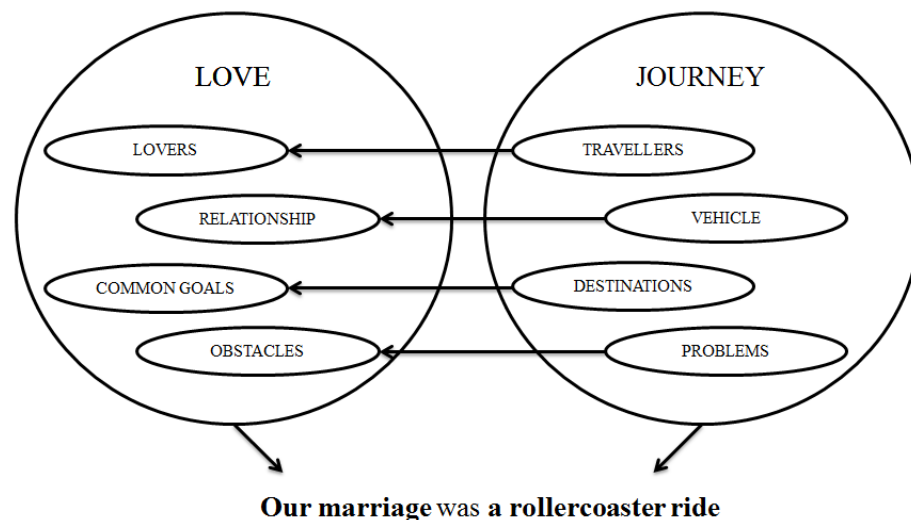


Fig. 6. The Conceptual Metaphor Theory. In line with the assumptions postulated within this theory, a metaphor *Our marriage was a rollercoaster ride* is comprehended through the preactivation of the conceptual metaphor LOVE IS A JOURNEY. Arrows going from JOURNEY to LOVE present the direction of the cross-domain mapping from the conceptual metaphor source JOURNEY to the conceptual metaphor target LOVE (after McGlone 1996: 548, adapted from Lakoff and Johnson 1980).

The Conceptual Metaphor Theory has been supported in studies showing that conceptual metaphors are activated when comprehending metaphoric utterances. For example, Pfaff et al. (1997) found faster reading times for euphemistic sentences embedded in contexts depicting conceptual metaphors (for example, *She's turning my crank* as depict-

ing a conceptual metaphor SEXUAL DESIRE IS AN ACTIVATED MACHINE) compared to contexts without a correspondence to a conceptual metaphor. Furthermore, Gibbs et al. (1997) employed a lexical priming paradigm, and observed a facilitation effect with shorter reaction times for idioms which originated from conceptual metaphors (for instance, *John blew his stack* as originating from a conceptual metaphor ANGER IS HEATED FLUID IN A CONTAINER) compared to idiomatic expressions that did not correspond to conceptual metaphors.

Nonetheless, in a series of experiments, McGlone (1996) showed evidence contradicting the CMT. In Experiment 1, participants paraphrased the presented metaphors, some of which were based on conceptual metaphors. In contrast to what the CMT predicts, when comparing the number of paraphrases that conveyed conceptual metaphors to those that did not, the authors observed no statistically significant differences. Similar results were obtained from Experiment 2, in which participants were explicitly asked to paraphrase the provided metaphors using different metaphors. Even when employing such procedures, participants still produced non-conceptual metaphor paraphrases, thus indicating that they did not refer to conceptual metaphors when interpreting the utterances. Finally, in Experiment 3, participants were exposed to conceptual-metaphor-consistent and inconsistent metaphoric paraphrases adapted from Experiment 2 as well as with original metaphors. Participants rated how similar a particular paraphrase was to the original meaning. The results showed no differences between the similarity ratings for CM-consistent and CM-inconsistent metaphors, which is again contrary to what the Conceptual Metaphor Theory claims. Consequently, these results show that the specific role of conceptual metaphors in metaphoric meaning interpretations remains unclear. Additionally, it needs to be noted that the CMT does not account for specific cognitive mechanisms that might govern the activation of conceptual metaphors, which makes it less feasible to test the tenets of this theory empirically. To address the limitations of the previously discussed theories, other views on metaphor comprehension were proposed, such as the Career of Metaphor Model and computational models of metaphor comprehension, which are presented in the two following sections.

2.3.6. The Career of Metaphor Model

The Career of Metaphor Model (Bowdle and Gentner 2005) is one of the few models of metaphoric language comprehension that thoroughly describes specific cross-domain mapping mechanisms engaged in novel and conventional metaphor comprehension. The model is partially based on the structure-mapping theory of analogy (Gentner 1983; see section 2.4.2), in which analogy is defined as the process of mapping from a base domain to a target domain. As a result, analogy pertains to applying previous knowledge and experience to bear on new situations, and is thus one of the most fundamental cognitive processes. In line with this theory, analogy involves comparison mechanisms between different relational representations that connect entities and ideas. The theory argues for two interrelated mechanisms during mapping operations; namely, the process of alignment and the process of projection. Alignment mechanisms aim at forming structurally consistent matches between different concepts. Once such links have been established, projection mechanisms allow for projecting any further elements from a base domain to a target domain (Bowdle and Gentner 2005: 196).

The Career of Metaphor Model adopts the assumption of cross-domain mappings, and claims that these mechanisms represent core processes engaged in metaphor comprehension. Cross-domain mappings are, however, postulated to differ depending on how lexicalized (conventional) metaphoric utterances are. Importantly, conventionalization is defined by Bowdle and Gentner (2005) as a property of the metaphor vehicle only, and not as a property of the entire metaphoric utterance. Conventional metaphors are characterized by denoting a conventional vehicle, which, due to its repeated and frequent use, has both a literal and a metaphoric reference. For example, a conventional metaphor *A gene is a blueprint* involves a vehicle (*blueprint*) that is polysemous and can denote either a literal (i.e., an architect's design plan) or a figurative sense (i.e., anything that provides a plan; Gentner and Bowdle 2001: 229). As a result of such a dual reference, lexicalized metaphors can be interpreted either as categorizations or comparisons. Categorizations require the target concept to be perceived as a member of a superordinate category specified by the base term. In comparison mechanisms, target concepts are structurally aligned with literal base concepts. Whether a conventionalized metaphor is understood as comparison or categorization depends on factors such as the available context and salience; nonetheless, all else being equal, categorization processes are considered more rapid and

less resource intensive than comparison operations, and thus conventional metaphors are assumed to be usually comprehended as categorizations.

Novel metaphors, on the other hand, involve a vehicle with only its literal reference. Namely, the base terms of novel metaphors refer to a domain-specific concept that denotes a literal meaning, and their correspondences to a domain-general category denoting a figurative meaning are not yet established. Consequently, novel metaphors can be only understood as comparisons, in which a target concept needs to be aligned with a literal base concept (Bowdle and Gentner 1999: 91). A distinction between novel and conventional metaphors is illustrated in Fig. 7.

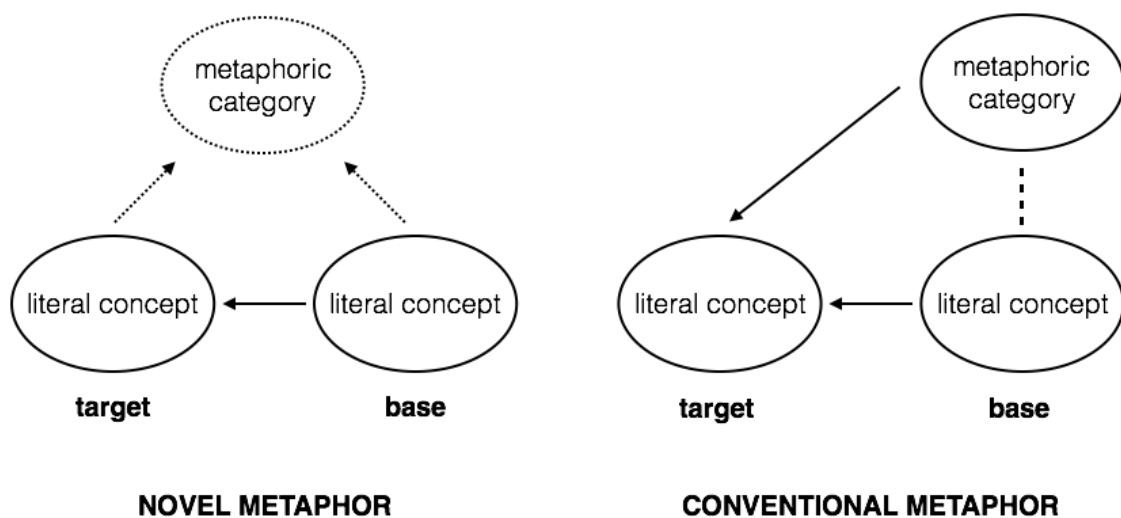


Fig. 7. The distinction between novel and conventional metaphors according to the Career of Metaphor Model. While a conventional metaphoric vehicle (i.e., base) is claimed to have both a literal and metaphoric reference, a novel metaphoric vehicle is assumed to have only a literal reference (after Bowdle and Gentner 1999: 92).

Importantly, novel metaphor comprehension is assumed to be more effortful relative to conventional metaphor comprehension due to the fact that comparison mechanisms are initiated only once the literal meaning has been found as inapplicable (Bowdle and Gentner 2005, as cited in Goldstein et al. 2012: 137). However, since the Career of Metaphor claims that novel metaphors are understood as comparisons, novel similes (e.g.,

Amnesia is like a rubber), whose linguistic form initiates comparison processes, are argued to be easier to comprehend compared to novel nominal metaphors (e.g., *Amnesia is a rubber*).³ All in all, the Career of Metaphor Model assumes that metaphor comprehension is modulated by metaphor conventionality, with more complex mechanisms engaged in meaning construction when processing novel metaphors compared to meaning retrieval in conventional metaphor comprehension.

2.3.7. Computational models of metaphor comprehension

Metaphor comprehension has also been investigated by means of developing computational models that aim to explain how people interpret metaphoric meanings. Computational theories of language processing address linguistic questions by proposing a rule-based approach, in which human language processing is perceived as a form of information processing that consists of a number of processing steps (Dijkstra and De Smedt 1996: 3). These steps are perceived as starting points when developing algorithms, on which computational models are based. The most prominent techniques employed in computational research on semantic processing are distributional models, which are grounded in the assumption that the meaning of a target item can be derived from words co-occurring with that item (Brysbaert et al. 2014: 135; Andrews et al. 2009: 477). Consequently, distributional models are based on tabulating the distributions of word co-occurrence in large corpora (Fig. 8). The co-occurrence of lexical items forms the basis of word associations, which are further translated into networks of nodes connected to one another (Brysbaert et al. 2014: 136).

³ A more in-depth discussion on simile processing is provided in section 2.4.

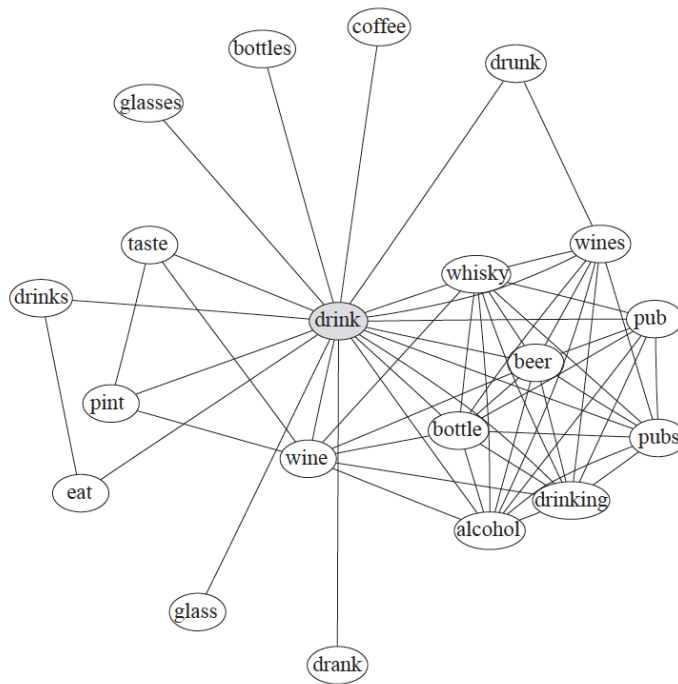


Fig. 8. Neighborhood cliques of the verb *drink* according to the distributional models of word co-occurrence (after Andrews et al. 2009: 477).

A well-established example of this approach is the latent semantic analysis (LSA; Landauer and Dumais 1997). The LSA method extracts how word meanings are used in different contexts by performing statistical computations on a large corpus of text. Based on the analyses of word co-occurrence, the LSA is used to determine semantic maps (spaces) that are of 300-400 dimensions. In the LSA, meanings are represented as vectors in a semantic space that can be calculated based on their co-occurrence in order to specify how close or far apart they are from each other, which can consequently show the semantic relationship between words.

The LSA method was employed by Kintsch (2001) in the Predication Model, which integrates the LSA knowledge base with the Construction-Integration (CI) model of text comprehension (Kintsch 1998). In line with the CI model, upon a presentation of a metaphor such as *My surgeon is a butcher*, the semantic neighborhoods of surgeons and butchers become activated. The semantic features of a metaphor topic specify which semantic features of a metaphor vehicle are relevant for a correct interpretation of an utterance. Consequently, semantic features of *surgeons* which are not related to those of *butchers* become inhibited, while those that are relevant become activated. The Predication

Model further specifies that meanings become contextually appropriate by means of modifying the LSA vector so that it is combined with the features of semantic neighbors related to the critical item. For instance, a metaphor *My surgeon is a butcher* is comprehended by means of modifying the vector representing the meaning of a *surgeon* (a predicate) so that attributes relating to a *butcher* (an argument) could be emphasized. According to the predication, upon the presentation of a word, the process of spreading activation results in the activation of some elements from the word's semantic neighborhood that fit the provided context, which consequently alters the vector of the presented word towards other related items. Importantly, Kintsch (2008) claims that the same mechanisms are involved in literal and figurative meaning interpretation, with, however, more extended processes of spreading activation during nonliteral language analysis (Kintsch 2008: 133).

Fig. 9 illustrates the network representing the metaphor *Happiness is gold*, with an argument (*happiness*), a predicate (*gold*), and three close neighbors of the predicate (*precious*, *nuggets*, *currencies*). *Gold* and *happiness* are only little related, as their $\cos = .11$ (as indicated by the solid line). The neighbors of the predicate compete for activation within the network (the dashed lines), as a result of which the neighbors that are related to both predicate and argument become activated, while those which are related only to the predicate become inhibited. Consequently, *currencies* and *nuggets* are rejected, whereas *precious* is selected to modify the predicate, which results in the selection of a correct interpretation of the presented metaphoric utterance.

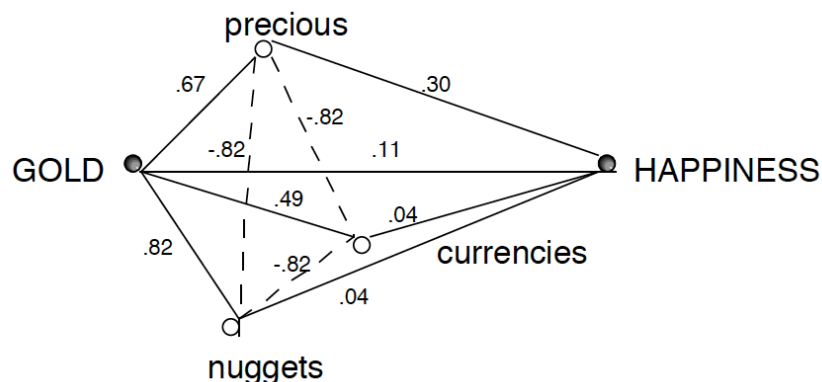


Fig. 9. The network representing the metaphor *Happiness is gold*. The network consists of an argument (*happiness*), a predicate (*gold*), and three close neighbors of the predicate (*precious*, *nuggets*, *currencies*). The solid lines provide cosines between the items in the LSA space, while the dashed lines indicate the activation between the three neighbors of the predicate within the network (after Kintsch and Bowles 2002: 261).

The question whether the LSA can account for how people comprehend metaphors was addressed by Kintsch and Bowles (2002), who asked their participants to read novel or conventional metaphoric sentences (for example, *My lawyer is a shark*) and to finish the subsequent sentence with a literal version of the previously presented metaphor (for instance, *My lawyer is mean*). After completing each sentence, participants decided how easy or difficult it was for them to understand each metaphor on a scale from 1 to 5. The researchers observed consistent results for conventional metaphors, which were relatively easy to understand, yet very erratic responses to novel metaphors, whose perceived level of difficulty varied among participants to a larger extent. In line with these results, the predication model also generated coherent interpretations for conventional metaphors, and very disparate responses for novel metaphors, which indicates that the LSA can account for the comprehension of metaphors of different conventionality levels.

Kintsch's approach to employing the LSA in order to examine metaphor comprehension was further adopted by Utsumi (2011), who used a semantic space model in order to determine whether metaphor comprehension is governed by comparison or categorization processes. As specified in the semantic space model, the meaning of each word is represented as a high-dimensional word vector (a semantic space) that is constructed based on distributional statistics of word occurrence. A semantic similarity between two words is computed by means of comparing their word vectors using the cosine \cos function. The simulation experiment performed by Utsumi (2011) showed that categorization processes were more effective when an entity was semantically rich, and thus when more features of a metaphor source could be attributed to a metaphor vehicle, as in the case of conventionalized metaphors. On the other hand, more semantically diverse entities, such as novel metaphors, needed to be interpreted via a comparison process. The semantic space model is therefore in agreement with the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), which points to the crucial role of comparison mechanisms, yet mostly in novel metaphor processing. The importance of comparison mechanisms in figurative meaning comprehension is further discussed in the following section, which is devoted to presenting theoretical accounts on simile comprehension.

2.4. From metaphor to simile comprehension

Simile is a subtype of metaphoric language, which is characterized by involving the comparison of two concepts (for example, *Lawyers are like sharks*). Mechanisms behind simile comprehension have been subject to discussion, with some perspectives on figurative meaning comprehension assuming similar mechanisms engaged in nominal metaphor (*A is B*) and simile (*A is like B*) interpretation, and other views arguing for distinct processes involved in the comprehension of these two types of utterances (see Glucksberg and Haught 2006 for a review). Such a discordance has resulted in the formulation of two competing views on metaphor and simile processing; namely, the comparison and categorization theories, whose tenets are presented in the following subsection.

2.4.1. Comparison and categorization views on metaphor and simile processing

In line with the comparison views (Ortony 1979; Gentner 1983; Fogelin 1988), any metaphor needs to undergo a property matching process, and thus it is comprehended by means of an implicit simile. Consequently, nominal metaphors (for example, *This lawyer is a shark*) can be paraphrased as similes (*This lawyer is like a shark*), and vice-versa, without any changes in meaning (Glucksberg and Haught 2006: 361). The comparison theories posit that nominal metaphors (*A is B*) are comprehended as similes due to the fact that the comprehension of such metaphors always requires the comparison of the features of metaphor topic (*A*) with the features of metaphor vehicle (*B*; Haught 2013: 255). Common features of metaphor topic and vehicle are identified based on a salience imbalance mechanism (Ortony 1979), in which a perceived similarity between two objects is a function of features shared by these concepts minus the features that are contrastive between them. Namely, in line with this hypothesis, any metaphor is characterized by a salience imbalance, in which the shared characteristic between metaphor topic and vehicle needs to be of low salience to the topic and of high salience to the vehicle of a metaphor (Ortony 1979, as cited in Fawson and Reutzel 1994: 358). For instance, in a metaphor *The man's feet were ice*, the topic's characteristics are more salient (i.e., feet can be small, big, hot, cold, etc.) compared to the characteristics of the vehicle (i.e., ice is cold, hard, slick, made of water).

Importantly, previous research has strongly contradicted the comparison account by showing that nominal metaphors and similes often differ in their interpretability, and can carry different meanings (Glucksberg and Haught 2006: 360). Additionally, nominal metaphors have been found to be comprehended faster than similes (Johnson 1996; Glucksberg 2003), and interpreted as more metaphorical compared to similes (Zharikov and Gentner 2002; Glucksberg 2003). As a result, the categorization view has been proposed, which argues for independent mechanisms governing nominal metaphor and simile comprehension.

The categorization views (Glucksberg 2001, 2003) hold that similes and nominal metaphoric utterances are comprehended in their own right, with similes involving an assertion of similitude, and nominal metaphors bringing about a categorization assertion (Glucksberg and Haught 2006: 361). Therefore, nominal metaphors are likely to be interpreted as categorizations due to the fact that they are identical to literal categorical statements (for example, *A lemon is a fruit*). In contrast, similes are hypothesized to be comprehended by means of comparison, as they are grammatically identical to literal comparisons (for instance, *A lemon is like an orange*; Utsumi 2007: 293).

2.4.2. Structure mapping in analogy and similarity

The view which can account for both categorization and comparison processes in nominal metaphor and simile comprehension is the structure mapping theory (Gentner 1983). Its central assumption is analogy, according to which a relational structure applying to one domain can be also applied to another domain (Gentner 1983: 156). Importantly, the strength of an analogical match is modulated by the degree of relational overlap, but not feature overlap. Consequently, analogy is perceived as a device for conveying that two domains share a relational structure, even though their objects are distinct from each other (Gentner and Markman 1997: 46).

According to the structure mapping theory, metaphor interpretation involves two interrelated mechanisms: alignment and projection. The role of the alignment process is to create a match between two representations. The process of alignment is required to be structurally consistent, that is, it must include a parallel connectivity and one-to-one correspondence. Secondly, it must have a relational focus, with common relations, but not

necessarily common object features. Finally, analogy is required to be systematic, matching connected systems of relations (Gentner and Markman 1997: 47). Once the process of alignment is complete, predicates from the base are projected to the target in the form of candidate inferences.

Metaphors are argued to represent either analogy or similarity. Similarity and analogy are considered to be very similar to each other, since they both involve the process of structural alignment and mapping between mental representations (Gentner and Markman 1997: 45). However, while analogy involves only relational predicates, similarity involves both relational predicates and object attributes. For instance, a simile *Two lovers are like twin compasses* forms a relational comparison, while *A moon is like a silver coin* forms an attribute comparison (Gentner and Markman 1997: 48).

These assumptions were further elaborated on by Aisenman (1999), who proposed the relational precedence hypothesis. The view holds that the distinction between comparisons and categorizations is based on the type of the final interpretation, with nominal metaphors expressing relational meanings, and similes communicating attributional interpretations. Consequently, Aisenman (1999) claims that nominal metaphors are predominantly used to highlight common relations between the target term and the base term, whereas similes communicate common attributes. In her study, Aisenman (1999) confirmed that participants preferred a simile form when the base and target shared object attributes (for instance, *The sun is like an orange – both are round and orange*). When, however, the base and target shared a common relational structure, participants opted for a nominal metaphoric form (for example, *Television is a magnet – both attract*).

The structure mapping theory represents a cornerstone of the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), which further developed the distinction between comparison and categorization mechanisms. Under this view, the preference for one mechanism over the other depends on the conventionality of the base term. While conventional metaphors are processed by means of alignment of a target item with conventional abstraction associated with the base term, novel metaphors undergo the process of comparison between the target domain and the literal meaning of the base domain (Zharikov and Gentner 2002: 976).

2.4.3. The role of topic-vehicle relationship in metaphor and simile comprehension

In favor of the structure mapping theory, previous studies have shown that categorization operations are preferential over comparison processes when metaphor topic and vehicle have a vast number of common features (Chiappe et al. 2003). The role of topic-vehicle relationship was also stressed in the revised version of the Attributive Category Theory (Jones and Estes 2005), which, however, posits that metaphor aptness, and not necessarily metaphor conventionality, might determine mechanisms involved in nominal metaphor and simile comprehension. Metaphor aptness is defined as the extent to which a nonliteral meaning of metaphor vehicle refers to an important feature of metaphor topic (Jones and Estes 2006: 19). Metaphors are apt when their vehicles have a strong association with their attributive property. For example, a nominal metaphor *That fashion model is a rail* is highly apt, as its vehicle (*a rail*) refers to the property which is a relevant attribute of fashion models (i.e., thin). Thus, metaphor aptness is highly related to the interaction between metaphor topic and vehicle. According to Jones and Estes (2005), apt metaphors, whose topic and vehicle are of high similarity, should be understood by means of categorizations, whereas less apt metaphors are argued to be interpreted as comparisons.

Similarly to the revised version of the Attributive Category Theory, the Interpretive Diversity View (Utsumi 2007) emphasizes the role of topic-vehicle relationship in nominal metaphor and simile comprehension. Under this view, the interpretive diversity, which is defined in terms of semantic richness of the figurative meaning of a topic-vehicle pair, leads to distinct mechanisms behind nominal metaphor and simile processing (Utsumi 2007: 292). High interpretive diversity is hypothesized to be a by-product of a large number of features that a nonliteral meaning constitutes as well as of the fact that these features provide a uniform (even) salience distribution, with an equal number of meanings conveyed by topic-vehicle pairs. High semantic richness (high interpretive diversity) is argued to result in a higher preference for a nominal metaphor over a simile form. Namely, semantically rich topic-vehicle pairs are preferentially processed as categorizations, and are therefore easier to be interpreted when presented as nominal metaphors. Less interpretive diverse utterances, on the other hand, require the process of comparisons, and thus a simile form facilitates the comprehension of such meanings.

2.4.4. A Relevance Theory perspective on metaphor and simile comprehension

The Relevance Theory (Sperber and Wilson 1986), which was already discussed in section 2.3.1, also differentiates between nominal metaphor and simile comprehension. In line with the theory, nominal metaphors and similes are bound to communicate different explicatures, with ad hoc concepts required in nominal metaphor comprehension, and lexically encoded concepts involved in simile interpretation (Carston and Wearing 2011: 296), as in the example of *Mr Smith is (like) a mouse*:

(8) Nominal metaphor explicature: *Mr Smith is a mouse**

(9) Simile explicature: *Mr Smith is like a mouse*

(Carston and Wearing 2011: 296)

In (8), *a mouse* pertains to an ad-hoc concept, and its meaning is logically implied through the analysis of typical implications of mice (i.e., timid, quiet, etc.). In (9), on the other hand, *a mouse* cannot correspond to an ad-hoc concept, as *Mr Smith* is a member of the category that includes mice, and he is like other members of this category. Consequently, a simile interpretation requires the activation of encyclopedic assumptions regarding mice, and perceiving them as relevantly applying to *Mr Smith* (Carston and Wearing 2011: 297). Importantly, somewhat similarly to what the Career of Metaphor Model claims, the Relevance Theory also posits that metaphor and simile comprehension is modulated by meaning conventionality. Namely, novel and creative metaphorical uses are rarely understood in terms of ad hoc concepts, unlike familiar utterances. As a result, similes are more likely to be used with a view to expressing novel meanings, while nominal metaphors usually communicate conventionalized ad hoc concepts (Carston and Wearing 2011: 303).

2.5. Factors modulating metaphor comprehension

The aforementioned theoretical accounts of metaphoric language comprehension suggest that metaphor processing is likely to be modulated by numerous factors, such as the available contexts, metaphor salience, conventionality, aptness, interpretive diversity, as well

as by the type of a metaphor itself. Early models of metaphor comprehension have strongly emphasized that it is the context in which a metaphor is embedded that modulates the difficulty of arriving at nonliteral meanings. While low constraining contexts are assumed to lead to a preliminary activation of a literal interpretation, highly constraining contexts are argued to enable a direct activation of a figurative meaning (Sperber and Wilson 1986; Frisson and Pickering 2001; Katz and Ferretti 2001).

Other models have, on the other hand, undermined the role of context, and have instead assumed that metaphor comprehension is modulated by either metaphor salience (Giora 2002), metaphor conventionality (Bowdle and Gentner 2005), metaphor aptness (Chiappe et al. 2003; Jones and Estes 2005), or metaphor interpretive diversity (Utsumi 2007). Meaning salience is defined as dependent on frequency, prototypicality, familiarity, and conventionality of a presented metaphor. Nonetheless, due to the fact that the Graded Salience Hypothesis (Giora 2002) does not specify cognitive mechanisms behind the processing of salient and non-salient meaning comprehension, the role of meaning salience in metaphor comprehension remains unclear. The Career of Metaphor Model (Bowdle and Gentner 2005), on the other hand, thoroughly describes how metaphors undergo the process of conventionalization, and explicates that metaphor comprehension depends on how frequently a metaphor vehicle is used in its nonliteral sense. Finally, metaphor aptness and metaphor interpretive diversity have also been proposed to influence processes involved in metaphor comprehension, with apt and interpretive diverse metaphors usually comprehended as categorizations (Chiappe et al. 2003; Jones and Estes 2005; Jones and Estes 2006; Utsumi 2007). The assumptions emphasizing the role of metaphor aptness as well as metaphor interpretive diversity, both of which emphasize the relationships between metaphor topic and vehicle, contradict the Career of Metaphor Model, which, on the other hand, undermines the role of metaphor topic in metaphoric meaning comprehension.

In addition to the aforementioned factors, metaphor comprehension has also been postulated to depend on the type of a metaphoric utterance. For instance, similes (*A is like B*; for example, *Life is like a journey*) are assumed to be comprehended by comparisons (Bowdle and Gentner 2005). Nominal metaphors (*A is B*; for instance, *Life is a journey*) have, on the other hand, been argued to be processed as categorizations, as their vehicles represent superordinate categories of objects they refer to, and their target items epitomize members of these categories (Glucksberg 2001; Glucksberg 2003; Bowdle and Gentner

2005). Predicative metaphors (for example, *The rumor flew through the office*), in contrast, cannot be processed as categorizations, as the semantic structure of nouns differs from the semantic structure of verbs, which do not refer to objects, but instead describe events and actions (Utsumi and Sakamoto 2011: 301-302). Additionally, while nouns form a hierarchical structure with superordinate and coordinate relations, verbs cannot involve such a structure, as they entail causation, antonymy, as well as entailment, and thus the structure of verbs is non-hierarchical. As a result of these profound differences, predicative metaphors are assumed to be processed by means of an indirect categorization, in which a verb (*to fly*) evokes objects typically associated with that verb (for instance, *planes*, *insects*, and *birds*), which in turn evoke an abstract meaning attributed to the metaphor target (i.e., the rumor that spread rapidly in the office; Utsumi and Sakamoto 2011).

It therefore seems that a part of speech (e.g., a noun or a verb) which is present in a metaphoric utterance might also modulate processes engaged in its comprehension. Such an assumption has also been supported when examining adjectival metaphors (for example, *red voice*), which are postulated to be comprehended by means of the two-stage process of categorization (Utsumi and Sakamoto 2007). In two-stage categorization, a metaphor source (i.e., *red*) triggers the activation of both concrete and abstract red objects (for instance, an apple, blood, passion, danger), which is followed by the selection of objects which are relevant to the metaphor target (i.e., *voice*), as a result of which an abstract meaning can be derived (for example, a voice that is scary and dangerous). The next section of this chapter discusses behavioral as well as electrophysiological research on metaphoric language comprehension, which has been conducted with a view to testing the aforementioned theoretical frameworks as well as models of metaphor processing, and has examined whether and how different factors modulate metaphor comprehension.

2.6. Metaphoric meaning comprehension: Empirical evidence

The following sections report both reaction time and electrophysiological results obtained from studies into metaphoric meaning comprehension. Reaction time patterns can account for the direct as compared to non-direct views on nonliteral meaning comprehension, as longer RTs reflect increased processing difficulty (see section 1.3), and thus

longer response times for metaphoric than literal utterances might indicate a temporal priority of literal meaning activation. In ERP research, two components are frequently discussed when examining metaphor processing. First, the N400 (see section 1.4.2.2) is often reported as an indicator of the amount of information that needs to be retrieved from long-term memory during lexico-semantic access, and thus larger N400 amplitudes for metaphoric than literal utterances could indicate more resource intensive processes engaged in cross-domain mapping (Lai and Curran 2013). The second commonly discussed ERP component is the LPC (see section 1.4.2.2), which is associated with meaning integration. A more pronounced LPC response to metaphoric relative to literal meanings might index additional meaning re-analyses or increased working memory load during nonliteral meaning integration (Brouwer et al. 2012).

Most of the studies discussed in the following subsections have focused on testing direct as compared to non-direct (serial) accounts on metaphoric language processing, and have therefore been aimed at demonstrating if literal meaning is accessed preferentially before a figurative interpretation becomes available, as argued in the Standard Pragmatic View (see section 2.2.2). For instance, the role of literal meaning in metaphor comprehension was examined by Weiland et al. (2014), who employed event-related potentials in a cross-modal masked priming paradigm. As previously explained in section 1.3, a masked priming paradigm is used to investigate automatic language-related processes that are strategy-free. In their study, Weiland and colleagues (2014) used auditorily presented metaphoric (e.g., *These lobbyists are hyenas, if you believe the kindergarten teacher*) and literal sentences (e.g., *These carnivores are hyenas, if you believe the kindergarten teacher*), which were preceded by a visually presented masked prime word that was semantically related to the literal meaning of the previously presented either metaphoric or literal sentence (e.g., *furry*). Participants performed two tasks: a color detection task and a word recognition task. The results of Experiment 1, in which nominal metaphor processing was investigated without priming, showed a biphasic N400-LPC effect, with more pronounced N400 and LPC amplitudes for nominal metaphors relative to control literal sentences. The same pattern of results was observed in Experiment 2, in which the priming paradigm was used. Importantly, in Experiment 2, the authors additionally observed an attenuated N400 response to metaphors preceded by a literal prime word, which indicates that metaphor comprehension might be facilitated by the subliminal processing

of a literal meaning, as it pre-activates the semantic network of the target item, and consequently might ease extra operations engaged in metaphor comprehension. The following subsections are devoted to discussing studies that have addressed the questions whether and to what extent additional factors such as contextual cues, metaphor conventionality, and comparison mechanisms influence metaphoric and literal meaning processing.

2.6.1. How context modulates metaphoric language comprehension

A seminal study on the role of context in metaphor comprehension was conducted by Pynte et al. (1996). In Experiment 1, participants were presented with conventional metaphoric (e.g., *Those fighters are lions*) and literal sentences (*Those animals are lions*), and performed a self-paced reading task. The results revealed a larger N400 response to conventional metaphors than literal meanings, thus pointing to more resource intensive mechanisms engaged in familiar metaphor relative to literal meaning comprehension. In Experiment 2, in addition to conventional metaphors and literal utterances, novel metaphoric sentences (e.g., *Those apprentices are lions*) were presented to participants. Interestingly, results from Experiment 2 showed that both novel and conventional metaphors evoked more pronounced N400 amplitudes compared to literal sentences. Experiment 3 and Experiment 4 were further aimed at examining whether the previously observed effects had been modulated by context-related effects. To this end, in Experiment 3, novel metaphoric sentences were embedded in irrelevant contexts, while conventional metaphoric sentences were presented in relevant contexts. In Experiment 4, on the other hand, novel metaphoric sentences were embedded in relevant contexts, whereas conventional metaphoric sentences were presented in irrelevant contexts. The results revealed a strong effect of context, with the N400 amplitudes larger for both novel and conventional metaphors when these were embedded in non-supportive compared to supportive contexts. Interestingly, when comparing ERP patterns observed in response to novel and conventional metaphors presented in the relevant and irrelevant contexts, the researchers observed more pronounced N400 amplitudes for novel metaphors embedded in the supportive context relative to conventional metaphors presents in the non-supportive context.

Such results therefore point to the crucial role of both metaphor conventionality and contextual cues in metaphoric meaning comprehension, and seem to be against the Graded Salience Hypothesis (Giora 2002; see section 2.3.4), which claims that context does not influence metaphor processing.

The role of context was also examined by Yang et al. (2013), who employed a word-to-sentence matching paradigm, which involved the presentation of two subsequent stimuli. For instance, a metaphoric sentence *The green lung of the city* was followed or preceded by either a relevant (e.g., *park*) or irrelevant stimulus (e.g., *semaphore*). Yang and colleagues (2013) compared ERPs in response to novel metaphors and literal statements that were preceded by different probe words, which were used in order to prime meaning comprehension. In a self-paced reading task, participants were presented with five conditions: congruent target metaphor (e.g., *Sarcastic – The girl is a lemon*), congruent source metaphor (*Sour – The girl is a lemon*), incongruent metaphor (*Idiotic – This girl is a lemon*), congruent literal (*Sour – The fruit is a lemon*), and incongruent literal condition (*Idiotic – The fruit is a lemon*). Interestingly, the results evidenced no modulations within the N400 time frame, but a modulation of the LPC response, with more robust LPC amplitudes for incongruent both metaphoric and literal conditions compared to congruent conditions, which suggests that mapping mechanisms are present in both metaphoric and literal meaning analysis. Importantly, incongruent metaphors elicited a more pronounced LPC response compared to incongruent literal statements, thus indicating that incongruent contextual information poses a greater challenge in metaphor than literal meaning integration. Additionally, a linear effect across the congruent conditions was found, with the largest LPC amplitudes evoked by congruent target metaphors, followed by the congruent literal condition, and finally congruent source metaphors. These results point to potentially more resource intensive mechanisms involved in the integration of contextual information with the metaphor target domain compared to the metaphor source domain.

Further evidence highlighting the role of context in metaphoric meaning interpretation was provided by Bambini et al. (2016). While in Experiment 1, participants were provided with familiar and non-familiar metaphoric as well as literal sentences embedded in a neutral context (for example, *Do you know what that lawyer is? A shark*), in Experiment 2, the utterances were presented in a supportive context (for instance, *That lawyer is really aggressive. He is a shark*). Importantly, the researchers employed an implicit

task so that participants would not be aware of the presence of metaphors in the presented stimuli. To this end, participants were presented with two adjectives that occurred after the presentation of the critical passage, and selected the adjective that better matched the preceding stimulus (e.g., *ferocious* vs. *geographical*). RT results showed no differences between familiar and non-familiar metaphors in the adjective matching task. ERP results revealed more pronounced N400 amplitudes for both familiar and non-familiar metaphoric expressions relative to literal utterances only in the neutral context, while a larger LPC response to metaphors compared to literal sentences was observed in both neutral and supportive contexts. Such results were interpreted by the authors as indicative of the fact the N400 response to metaphoric meanings might be modulated by contextual cues, and not necessarily by stimulus metaphoricity. The LPC effect, in contrast, might reflect pragmatic interpretative mechanisms engaged in metaphoric language comprehension.

2.6.2. How metaphor conventionality modulates metaphoric language comprehension

Apart from investigating the role of context in metaphoric meaning comprehension, ERP research has been frequently devoted to investigating how metaphor conventionality influences cognitive mechanisms engaged in metaphor processing. For instance, De Grauwe et al. (2010) compared ERPs to familiar nominal metaphoric (e.g., *Unemployment is a plague*), literal (*Cholera is a plague*), and anomalous (*Metal is a plague*) sentences in a semantic decision task. The obtained results showed the most pronounced N400 amplitudes for anomalous utterances. Metaphoric sentences elicited a more robust N400 response than the literal condition, yet only within the early N400 time window (325-400 ms), which might indicate an initial conflict between literal and metaphoric meaning, which was, however, resolved as soon as within 400 ms post stimulus onset. Within the LPC time frame, anomalous sentences elicited prolonged negativity in both experiments, which might have indicated a continued difficulty in meaning integration. Furthermore, nominal metaphoric utterances evoked more pronounced LPC amplitudes relative to literal sentences, which was interpreted by the authors as indicative of additional processes engaged in metaphoric compared to literal meaning integration.

Highly conventional metaphors were also examined by Balconi and Amenta (2010), who presented their participants with congruous metaphoric (e.g., *A fighter is a lion*), congruous literal (*A fighter is a soldier*) as well as incongruous metaphoric (*A fighter is a lamb*) and incongruous literal (*A fighter is an employee*) sentences in a self-paced reading task. The results showed a more pronounced N400 response to incongruous conditions, irrespectively of whether these were literal or metaphoric. Such results might indicate that cross-domain mapping in conventional metaphor processing is not necessarily more resource intensive than in literal meaning processing.

In a study by Coulson and Van Petten (2002), in addition to literal (e.g., *He knows that whiskey is a strong intoxicant*) and metaphoric conditions (*He knows that power is a strong intoxicant*), the researchers included the literal mapping condition, where the literal meaning of a word was embedded in highly non-supportive contexts (e.g., *He has used cough syrup as an intoxicant*). The sentences were presented one word at a time, and participants answered yes/no comprehension questions. The results showed a graded N400 response, with the most pronounced N400 amplitudes for metaphoric utterances, followed by the literal mapping condition, and the least negative N400 amplitudes for literal sentences. Such results indicate that meaning literality and metaphoricity represent different ends of a continuum, with some meanings falling in between the two extremes, as in the case of the literal mapping condition. Even though similar mechanisms might govern literal and metaphoric meaning comprehension, they might require a different degree of cognitive effort. In line with this interpretation, increased N400 amplitudes for both metaphoric and literal mapping conditions relative to literal sentences may indicate an initial conflict between source and target domains. Within the LPC time window, on the other hand, metaphoric utterances evoked more pronounced amplitudes compared to both literal and literal mapping utterances, thus suggesting that metaphoric meaning integration might be more resource intensive. It should, however, be noted that Coulson and Van Petten (2002) did not specify how conventional or novel their metaphoric utterances were, and thus it remains unclear whether the obtained results might have been modulated by metaphor conventionality.

Novel metaphor comprehension was examined by Tartter et al. (2002), who presented their participants with sentences whose critical word directed towards literal, novel metaphoric, or anomalous interpretations (e.g., *The flowers were watered by nature's*

rain/tears/laughter, respectively). Participants performed a self-paced reading task followed by a recognition task, in which they decided whether they had previously been exposed to a presented sentence. ERP results revealed that literal sentences started to diverge from both novel metaphoric and anomalous utterances as soon as 160 ms after critical word onset, eliciting an attenuated N200 response. Novel metaphoric sentences branched off the anomalous condition 280 ms post stimulus onset, and evoked a more pronounced N400 response than literal sentences, which was, however, of smaller magnitude compared to anomalous utterances. Such results suggest that novel metaphoric meanings might be processed somewhat similarly to anomalous sentences, at least during very early stages of language processing.

In addition to comparing ERP evoked by novel metaphors only, Arzouan et al. (2007) examined brain responses to both novel and conventional metaphors. To this end, the researchers presented their participants with novel metaphoric (e.g., *conscience storm*), conventional metaphoric (*lucid mind*), semantically related (*burning fire*), and semantically unrelated word dyads (*indirect blanket*). In Experiment 1, participants performed a semantic decision task, while in Experiment 2, they were instructed to read the word dyads and to think about their possible interpretations. Results obtained from both experiments showed that novel metaphors evoked longest RTs, followed by unrelated utterances, conventional metaphors, and semantically related word pairs. Importantly, pairwise comparisons showed no differences in reaction times for conventional metaphoric and semantically related word dyads. ERP results from both experiments showed a linear effect, with the smallest N400 amplitudes for semantically related, followed by conventional metaphoric, novel metaphoric, and finally unrelated word dyads. This points to the least resource intensive mapping operations engaged in semantically related meaning analysis, and the most complex mechanisms involved when processing semantically unrelated utterances. Interestingly, even though novel metaphors did not differ from conventional metaphors, the researchers found a topographic difference, with a right hemisphere bias to novel metaphoric and unrelated word pairs, and a left hemisphere bias to conventional and semantically related word dyads. Such results might be partially in line with the Graded Salience Hypothesis (Giora 2002; see section 2.3.4), which points to the right hemisphere preference when processing meanings that are low in salience. It must, however, be noted that, due to a poor spatial resolution offered by ERPs, the scalp distribution does not provide information with respect to the neural generators of particular

ERP components, and thus the ERP activity observed over the right electrode sites might not necessarily reflect right hemisphere involvement. Finally, within the LPC time window, Arzouan and colleagues (2007) observed sustained negativity to novel metaphors relative to both conventional metaphoric and literal word pairs. Such prolonged negativity in response to novel metaphors might have overlapped with the LPC response, and might indicate the ongoing difficulty of novel meaning integration or access to the non-literal route during novel metaphor comprehension. Importantly, such results might have resulted from the fact that the word dyads used by Arzouan and colleagues (2007) provided minimal contextual cues to facilitate novel metaphoric meaning comprehension.

Novel and conventional metaphors were further investigated by Lai et al. (2009), whose participants decided whether novel metaphoric (e.g., *Every second of our time was attacked*), conventional metaphoric (*Every point in my argument was attacked*), literal (*Every soldier in the frontline was attacked*), and anomalous sentences (*Every drop of rain was attacked*) carried a perfect sense, some sense, little sense, or no sense. Reaction time results showed longest RTs for novel metaphors, followed by anomalous utterances, conventional metaphors, and shortest RTs for literal sentences. The results observed within the early N400 time frame (320-440 ms) showed a more pronounced N400 response to anomalous, novel metaphoric, and conventional metaphoric sentences compared to literal utterances. Importantly, within the late N400 time window (440-560 ms), waveforms for conventional metaphors converged with those evoked in response to literal sentences, both of which elicited smaller N400 amplitudes than novel metaphoric and anomalous utterances. Such results suggest that conventional metaphoric meanings might be more cognitively taxing than literal utterances, yet only during early stages of language processing. Novel metaphor processing, on the other hand, was continuously more resource intensive than literal meaning analysis. The results obtained from the studies by Arzouan et al. (2007) as well as Lai et al. (2009) are therefore in line with the assumption that metaphor comprehension is modulated by metaphor conventionality (Bowdle and Gentner 2005; see section 2.3.6). A more pronounced N400 response to novel metaphors might thus indicate that novel metaphor comprehension, as it involves sense creation mechanisms, is more resource intensive than conventional metaphor interpretation, which engages only sense retrieval processes.

The role of metaphor conventionality in metaphor comprehension, which is emphasized within the Career of Metaphor Model (Bowdle and Gentner 2005), was also

supported by Goldstein et al. (2012). In their study, the researchers tested novel metaphoric, conventional metaphoric, semantically related, and semantically unrelated word dyads, which were all adapted from the above-discussed study by Arzouan et al. (2007). Goldstein and colleagues (2012) used an exposure phase, which took place prior to the ERP experiment, and during which participants either performed a semantic decision task in response to presented word dyads (50% of trials) or provided an explanation for the word pair (50% of trials). In the subsequent ERP test phase, participants performed only a semantic judgment task. Behavioral results from the test phase showed longest RTs for novel metaphors, followed by anomalous utterances, conventional metaphors, and finally literal statements. Additionally, RT results revealed that novel metaphors which had previously been explained elicited significantly shorter response times compared to unexplained novel metaphoric expressions. ERP results further showed a metaphor conventionalization effect, revealing reduced N400 amplitudes for novel metaphoric word dyads which had been explained during the exposure phase compared to the previously unexplained novel metaphors. Such results indicate that explaining the meaning of a novel metaphor might help to establish a relational structure between the source and target domains. Interestingly, such an effect was reverse with regard to conventional metaphors, as a more pronounced N400 response was observed to previously explained conventional metaphoric word pairs compared to those which had not been explained. This effect was interpreted by the researchers as resulting from comparison operations engaged in explaining conventional metaphors, which led to an increased semantic memory activation, and consequently a larger N400 response. Within the LPC time window, the researchers observed reduced LPC amplitudes for novel metaphors. Similarly to the results observed by Arzouan et al. (2007), the reduction within the LPC time window observed in response to novel metaphors might have overlapped with sustained negativity, thus indicating secondary attempts to access meaning and integrate semantic information. Sustained negativity was, however, attenuated in response to previously explained novel metaphors compared to those novel metaphoric word dyads which had not been explained by participants. Such results point to a decreased cost of meaning re-analyses when novel meanings have already been constructed. Overall, the study by Goldstein and colleagues (2012) highlights the dynamics of metaphoric meaning interpretation, which seems to be strongly modulated by metaphor conventionality as well as specific task requirements.

As shown in the studies by Arzouan et al. (2007) as well as Goldstein et al. (2012), novel metaphors might evoke prolonged negativity within the LPC time frame, which may overlap with the LPC response, reflecting ongoing difficulty in meaning integration or access to the nonliteral route during the comprehension of novel meanings. Sustained negativity was also observed in a study by Rutter et al. (2012), who employed a delayed task for novel metaphoric (e.g., *The clouds have danced over the city*), literal (*The clouds have moved over the city*), and anomalous sentences (*The clouds have read over the city*). After the presentation of each sentence, participants first decided whether the sentence was unusual and appropriate. Then, a subsequent question appeared. For unusual utterances, participants decided whether the sentences were familiar or unfamiliar. For appropriate utterances, participants decided if sentences were meaningful or meaningless. Reaction time results showed that novel metaphors evoked longer RTs than anomalous utterances, which elicited longer response times compared to literal sentences. Electrophysiological results further revealed a linear effect within the N400 time frame, with the most robust N400 response to anomalous, followed by novel metaphoric, and finally literal utterances. The graded effect was also found within the LPC time window, where sustained negativity was observed in response to anomalous and novel metaphoric sentences. Such prolonged negativity for anomalous utterances and novel metaphors might have reflected a continuation of the N400 response, possibly resulting from the ongoing difficulty of meaning integration. Additionally, this might have reflected a higher working memory load due to specific task requirements, as Rutter and colleagues (2012) employed a delayed response task, therefore expecting their participants to keep the meaning of an utterance in their working memory for a longer period of time.

More recently, sustained negativity within the LPC time window was observed by Tang et al. (2017a). The researchers used literal sentences (e.g., *The dog is a pet*), nominal conventional metaphors (*Home is a harbor*), as well as nominal novel scientific metaphors, which involved concepts from major disciplines of natural science (*Sound is a wave*). Participants performed a metaphoric judgment task, in which they decided whether the presented sentence had a metaphoric meaning or not. Behavioral results showed longer reaction times for novel scientific metaphors than conventional metaphors and literal conditions. The N400 response was observed to vary as a function of utterance type, with the most pronounced N400 amplitudes for novel scientific metaphors, followed by conventional metaphors, and finally literal statements. Interestingly, within the LPC

time frame, novel scientific metaphors elicited sustained negativity, unlike conventional metaphors and literal utterances, both of which evoked more positive amplitudes. Such results are therefore in line with those obtained by Arzouan et al. (2007) as well as Goldstein et al. (2012), and might point to the ongoing difficulty in novel metaphoric meaning integration.

In another study, Tang et al. (2017b) compared ERP patterns evoked in response to novel scientific metaphors (e.g., *Chromosomes are sisters*), novel poetic metaphors (*The girl is a strawberry*), conventional metaphors (*The language is a bridge*), and literal sentences (*A professor is a scholar*). Behavioral results showed longer RTs for both novel scientific and novel poetic metaphors relative to both conventional metaphors and literal sentences. Within the N400 time frame (350-450 ms), novel scientific metaphors and novel poetic metaphors elicited a similarly robust N400 response, which was of larger magnitude compared to conventional metaphoric and literal sentences. Differences between novel scientific and novel poetic metaphors were, however, observed within the late positivity time frame (550-850 ms), where novel scientific metaphors elicited sustained negativity relative to both novel poetic metaphors and conventional metaphors. Such results might be indicative of potentially more frequent cross-domain mappings in novel poetic metaphors compared to novel scientific metaphors, which denote scientific concepts rarely discussed in everyday language. In contrast, novel poetic metaphors communicate more common meanings, and thus their integration might have been less taxing. It therefore seems that the comprehension of novel metaphors might be highly modulated by the nature of source and target concepts they aim to integrate.

Prolonged negativity in response to novel metaphors was also observed by Rataj et al. (2018), who additionally examined the role task demands in novel meaning comprehension. The researchers presented their participants with novel metaphoric (e.g., *After a while, she saw a procession of bottles*), literal (*Every night you can hear a loud sound of breaking bottles*), and anomalous sentences (*In the market, she bought a bunch of ripe bottles*). In Experiment 1, participants performed a semantic decision task, while in Experiment 2, they read the sentences and answered comprehension questions. In both experiments, novel metaphoric sentences evoked significantly longer RTs than literal utterances. ERP patterns within the N400 time frame further revealed a linear effect across sentence types, with the most pronounced N400 amplitudes for anomalous, followed by

novel metaphoric, and finally literal sentences, which was also observed in both experiments. Differences between the two experiments were, nonetheless, found within the LPC time window. Namely, in both experiments, attenuated LPC amplitudes were evoked in response to novel metaphoric relative to both literal and anomalous sentences over parietal electrode sites; yet, in Experiment 1 (a semantic decision task), this effect had a broad parietal distribution, while in Experiment 2 (a reading task), it was observed over left parietal electrode positions. Such a difference in the LPC scalp distribution was interpreted by Rataj and colleagues (2018) as indicative of the involvement of both recollection and semantic processes when performing a semantic decision task (Experiment 1), and a more profound engagement of recollection mechanisms during a reading task (Experiment 2).

As the above discussion attests, studies on metaphor processing conducted thus far have revealed that novel and conventional metaphor comprehension involves a different degree of resource intensity, with more taxing operations engaged in lexical access as well as early semantic analyses when processing novel metaphoric utterances (Arzouan et al. 2007; Rutter et al. 2012). Thus, the results provide support for the role of metaphor conventionality in metaphoric meaning comprehension, as revealed in longer reaction times for novel than conventional metaphoric utterances, as well as modulations within the N400 time window, in which novel metaphors evoked larger N400 amplitudes compared to conventional metaphors (Coulson and Van Petten 2002; Arzouan et al. 2007; Lai et al. 2009; Tang et al. 2017a; Tang et al. 2017b). At the same time, in some studies, conventional metaphors elicited larger N400 amplitudes than literal statements, which might indicate either an initial activation and rejection of literal meaning in conventional metaphor comprehension, or the processes of selection of a correct meaning from a multiple, parallel activation of other interpretations (Lai et al. 2009; Tang et al. 2017a). Importantly, though the processes engaged in conventional metaphor and literal meaning comprehension are likely to be similar, familiar metaphors might involve more effort when arriving at their final interpretation (Coulson and Van Petten 2002). Within the LPC time frame, familiar metaphors evoked larger LPC amplitudes compared to anomalous utterances (De Grauwe et al. 2010). Novel metaphors, on the other hand, elicited ongoing negativity, which might have overlapped with the LPC response, and might indicate the ongoing difficulty of meaning integration or access to the nonliteral route during novel

metaphor comprehension (Arzouan et al. 2007; Goldstein et al. 2012; Tang et al. 2017a; Rataj et al. 2018).

2.6.3. How comparison mechanisms modulate metaphor and simile comprehension

As postulated within the Career of Metaphor Model, novel metaphor comprehension requires comparison mechanisms, and thus comparison processes initiated when comprehending similes should ease novel metaphor comprehension, as reflected in shorter RTs and smaller N400 amplitudes for novel similes relative to novel nominal metaphors. In one of the few studies on nominal metaphor and simile comprehension, Bowdle and Gentner (2005) asked their participants to perform a preference rating task, in which they were provided with statements presented in the form of categorizations “A is B” (e.g., *A mind is a kitchen*) or comparisons “A is like B” (*A mind is like a kitchen*). The results showed a higher preference for novel metaphors presented as comparisons, and for conventional metaphors presented as categorizations, therefore providing support for the postulates discussed within the Career of Metaphor Model. In the subsequent experiment, reaction times were measured in a self-paced reading task for the same stimuli. In agreement with the Career of Metaphor Model, Bowdle and Gentner (2005) found longer RTs for novel categorizations as compared to novel comparisons, and longer RTs for conventional comparisons than conventional categorizations.

Only little attention has been devoted to investigating metaphor and simile processing by means of employing brain imaging methods. Shibata et al. (2012) employed the functional Magnetic Resonance Imaging (fMRI) method in order to examine the hemispheric activity when processing nominal metaphors and similes. To this end, the researchers investigated brain responses to novel nominal metaphors (e.g., *Memory is a warehouse*), novel similes (*An education is like stairs*), literal sentences (*A dolphin is an animal*), anomalous sentences (*Scissors are dogs*), and anomalous similes (*Time is like a strawberry*) in a semantic judgment task. Behavioral results showed that novel nominal metaphoric sentences elicited longest response time, followed by novel similes, anomalous similes, anomalous utterances, and finally literal conditions. Brain imaging results further revealed differential brain activity in response to nominal metaphors and similes. First, similes evoked an increased activation in fronto-medial regions relative to nominal

metaphors. Since previous research has linked the fronto-medial brain areas to mechanisms engaged in coherence building (e.g., Goel et al. 1997; Ferstl and von Cramon 2002; Zysset et al. 2003), it seems that simile comprehension involves inference processes required in order to establish meaning coherence. Second, nominal metaphors elicited an increased activation in the right inferior frontal gyrus relative to similes. This might be accounted for by the right hemisphere advantage for processing meanings that are low in salience (Giora 2002), and would be in line with the behavioral results showing longer response times for novel nominal metaphors and novel similes, possibly due to the fact that novel nominal metaphors were perceived as less salient than novel similes.

Thus far one ERP study has investigated nominal metaphor and simile processing. In their study, Lai and Curran (2013) employed a priming paradigm, where novel metaphoric (e.g., *Ideas can sometimes be bumpy*), conventional metaphoric (*Life can sometimes be bumpy*), literal (*Roads can sometimes be bumpy*), and anomalous sentences (*Taste can sometimes be bumpy*) were primed by either simile-primers (presented in the form *A is like B*) or sentence-primers (*A is B*). Participants decided whether the presented target stimuli carried a perfect sense, some sense, little sense, or no sense. Behavioral results showed that within each condition, priming effects were observed for conventional metaphors, but not for novel metaphors. ERP results further showed that while simile-primers were effective in priming both novel and conventional metaphoric sentences, sentence-primers primed only conventional metaphors, as indicated by the N400 priming effect, in which attenuated N400 amplitudes were observed for trials preceded by related compared to unrelated items. Owing to the fact that reduced N400 amplitudes reflect a decreased amount of information that needs to be retrieved from the semantic network, the results observed by Lai and Curran (2013) suggest that comparison mechanisms might ease both novel and conventional metaphor comprehension. In contrast, categorization mechanisms present in sentence-primers, seem to facilitate only conventional metaphoric meaning comprehension. Such results therefore lend support to the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), which claims that while conventional metaphor comprehension might involve either categorization or comparison operations, novel metaphor comprehension requires comparison mechanisms.

2.7. Concluding remarks

The above discussion shows that much attention has been devoted to explicating mechanisms engaged in metaphor comprehension by proposing numerous theoretical accounts, models, as well as providing empirical evidence that shows how participants arrive at nonliteral meanings. Although traditional theories of figurative meaning processing have postulated that literal utterances enjoy the temporal priority compared to metaphoric meanings, more recent accounts have argued that metaphor comprehension is modulated by the context in which a metaphor is embedded, as contextual cues can lead to a faster activation of the target meaning. In addition, mechanisms engaged in metaphoric meaning comprehension have been stipulated to depend on the level of conventionality, with more resource intensive operations involved in meaning construction (novel metaphors) relative to meaning retrieval (conventional metaphors). Previous research into metaphoric language processing⁴ has provided strong support for the role of conventionality in metaphor comprehension, and have suggested that comparison mechanisms facilitate novel metaphoric meaning processing, as argued by the Career of Metaphor Model (Bowdle and Gentner 2005). The experiments described in this thesis were aimed to further test the tenets of the Career of Metaphor Model by means of extending research on metaphor comprehension from the monolingual to the bilingual context. To this end, in Experiment 1 (Chapter 4), event-related potentials were analyzed as evoked by novel and conventional metaphoric word dyads presented in the native (L1) and non-native (L2) tongue. Furthermore, the main objective of Experiment 2 (Chapter 5) was to examine the role of comparison in novel meaning comprehension in both L1 and L2 by measuring reaction times elicited in response to novel nominal metaphors and novel similes.

⁴ ERP studies into the N400 and LPC responses in monolingual metaphoric language comprehension are summarized in Appendix A (p. 239).

Chapter 3: Bilingual lexico-semantic processing

3.1. Introductory remarks

Over the recent decades, bilingual language processing has become one of the most widely investigated issues in psycholinguistics. Much attention has been devoted to providing theoretical frameworks and computational models that explicate how bilingual speakers process and comprehend their native (L1) and non-native tongue (L2), and propose factors that modulate the automaticity of bilingual lexico-semantic processing. Since the main objective of the two experiments reported in this dissertation was to examine metaphor comprehension in the context of bilingualism, it is crucial to discuss the aforementioned issues and to provide a background on how semantic information is processed in the bilingual mind.

The present chapter is divided into three major parts. The first part presents the best-grounded structure- and process-oriented views on bilingual visual word processing, pointing to their strengths and limitations. The second part is devoted to discussing electrophysiological research on bilingual language processing, and focuses on the N400 and LPC responses as indexes of lexico-semantic access and meaning integration. The final part of this chapter provides an overview of studies on bilingual figurative language comprehension, discusses potential reasons behind some contradictory results, and identifies gaps in this area of research on bilingualism.

3.2. Selected theories and models of bilingual lexico-semantic processing

Cognitive mechanisms engaged in bilingual language processing have been elucidated in a number of theoretical accounts as well as computational models, which show how bilingual speakers organize, store, and process L1 and L2 linguistic information. While some of them have addressed structure-oriented aspects of bilingualism, others have focused on process-oriented issues (Dijkstra 2007: 251). Structure-oriented perspectives show how L1 and L2 lexical items are stored in the bilingual lexicon, and thus they demonstrate whether memory representations are separate or shared across languages. Process-oriented perspectives, on the other hand, focus on how L1 and L2 information is processed in the bilingual mind, and address the question whether bilingual language processing is selective (language specific) or non-selective (language non-specific). Traditional theories of language processing in bilingualism have mostly focused on the structure-oriented aspects, and have aimed at explicating the organization of bilingual memory by means of describing how native and non-native lexical and conceptual representations are stored in the bilingual lexicon. The following section of this chapter is devoted to discussing three structure-oriented perspectives on the bilingual lexicon: Weinreich's (1953) compound, coordinate, and subordinate bilingualism (section 3.2.1.1), the Bilingual Dual-Coding Theory (Paivio and Desrochers 1980; section 3.2.1.2), and the Distributed Feature Model (de Groot 1992; section 3.2.1.3).

3.2.1. Structure-oriented perspectives on bilingual language processing

3.2.1.1. Compound, coordinate, and subordinate bilingualism

A traditional account on how information is organized in the bilingual lexicon was proposed by Weinreich (1953), who postulated three types of the bilingual mental lexicon: the compound, coordinate, and superordinate bilingual lexicon (Fig. 10). The distinction between these three types is based on how native and non-native lexical and conceptual representations are connected to each other, and originates from differences in language

acquisition and language use. First, in the compound bilingual lexicon, translation equivalents from the two languages share the same conceptual representation, but have separate lexical representations for L1 and L2. Such an organization of the bilingual lexicon is likely to stem from an interchangeable use of both the native and non-native language in the same environment, as a result of which both languages are automatically available through representational mediation processes (Weinreich 1953, as cited in Javier 2007: 39). Second, the coordinate bilingual lexicon is characterized as having two separate conceptual systems, with a signifier (a concept) for each signified (a lexical representation). Coordinate bilingualism is assumed to be preferential when learning languages in different environments, as in the case of the native tongue being acquired at home, and the non-native language being learnt at school. Finally, in subordinate bilingualism, Weinreich (1953) argues for a distinction between a dominant and a subordinate language, which usually stems from one language having been already mastered, and the other still being learnt. Consequently, a lexical representation from a subordinate language (usually the non-native one) can be linked to a semantic representation only by means of the pre-activation of a dominant tongue (usually the native one). The subordinate bilingual lexicon is postulated to stem from the fact that a non-native language is usually acquired through the mediation of the native tongue (Weinreich 1953, as cited in Javier 2007: 40).

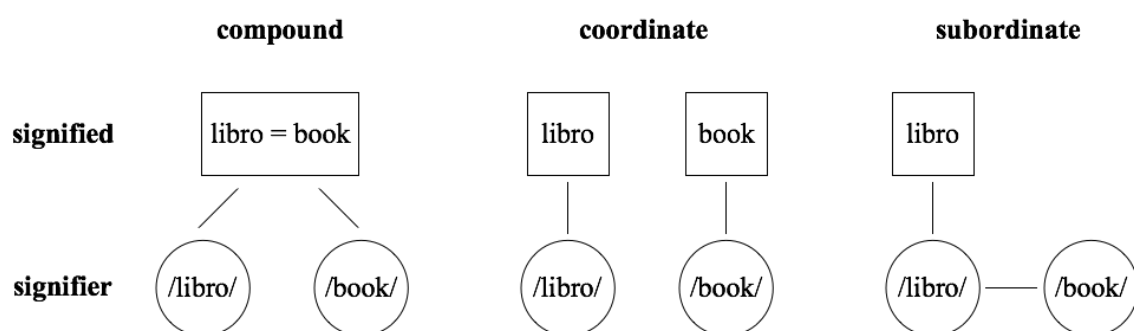


Fig. 10. Three bilingual storage representations (after Heredia and Cieřlicka 2014: 13, adapted from Weinreich 1953)

Importantly, not only are the abovementioned types of the bilingual lexicon not mutually exclusive, but they are even likely to coexist, with some lexical items representing a coordinate organization, while others forming a compound or subordinate lexicon (Weinreich 1953, as cited in De Angelis 2007: 88). Differences in the lexical organization might stem from distinct contexts of language learning, as well as from differences in the age of L2 acquisition (AoA). Namely, in line with the Critical Period Hypothesis (Lenneberg 1967), brain plasticity weakens after puberty, as a result of which languages acquired after the age of 13 are more likely to form the coordinate rather than the compound lexical organization.

Traditionally, neurobiological evidence showing distinct brain activity to L1 and L2 was linked to potentially separate storage for L1 and L2 lexical items, therefore providing support for Weinreich's (1953) proposals. For instance, Ribot (1882, as cited in Green 2008) reported medical cases of patients suffering from aphasia that affected only their non-native language. Furthermore, Dronkers et al. (1995) studied patients with symptoms of Broca's aphasia for one language, and Wernicke's aphasia for the other tongue. Importantly, though such results have previously been interpreted as indicative of a separate lexicon for L1 and L2 words, the observed results might instead reflect language-selective relative to language-non-selective mechanisms engaged in bilingual language processing. Other claims postulated by Weinreich (1953) might seem questionable, too. For instance, Weinreich (1953) perceived the bilingual lexicon as the sum of two separate lexicons for L1 and L2 items, therefore wrongly assuming independent mechanisms engaged in native and non-native lexical processing. Additionally, his assumptions did not account for how different types of lexical items (e.g., concrete vs. abstract words) are processed in the bilingual mind. The theories discussed in the following sections were therefore proposed with a view to providing more thorough insights into how L1 and L2 information is stored.

3.2.1.2. Bilingual Dual-Coding Theory

A theory that adopted and developed Weinreich's (1953) distinction between compound and coordinate bilingualism is the Bilingual Dual-Coding Theory (Bilingual DCT; Paivio and Desrochers 1980). The Bilingual DCT is an extension of the general Dual-Coding

Theory (DCT), which assumes that human cognition involves two independent processing systems, one specialized for verbal stimuli and responses (a logogen system), and the other one responsible for nonverbal objects and events (an imagen system; Paivio and Desrochers 1980, as cited in Paivio 2014: 43). Consequently, the mental lexicon is assumed to include linguistic information that is stored as words (“logogens”), and non-linguistic information stored as images (“imagens”). The structural model of the DCT comprises of representational units, connections between units, and their connections to stimuli and response systems. The model views the system of connections as neural pathways that allow for the activation of representations in response to an external stimulus.

The Bilingual DCT further develops the assumptions argued within the DCT, and it additionally postulates separate, yet interconnected logogen systems for the two languages, which are connected to non-verbal imagen systems. Even though the two language systems are argued to be separate, a word in one language calls up a reference to a word in the other language. The imagen systems can consist of either shared or language-specific imagens, whose structure bears resemblance to the compound-coordinate system proposed by Weinreich (1953). Namely, the presence of shared as opposed to separate imagens is modulated by the learning context. While learning languages simultaneously and in the same context (for instance, in the same environment) results in forming shared imagens, acquiring languages in different contexts and at different ages is likely to bring about separate imagens.

Fig. 11 represents the Bilingual DCT model with two verbal systems for the two languages (V_1 and V_2) and their connections to the imagery system (I). Since the verbal system has both separate and shared referential connections to the image system, translation equivalents from the two languages may not always activate the same imagens, due to potential differences in L1 and L2 contexts of learning. The image system plays a crucial role in the model, as it enables the connections between the two languages. As a result, translation equivalents may become activated indirectly through imagens.

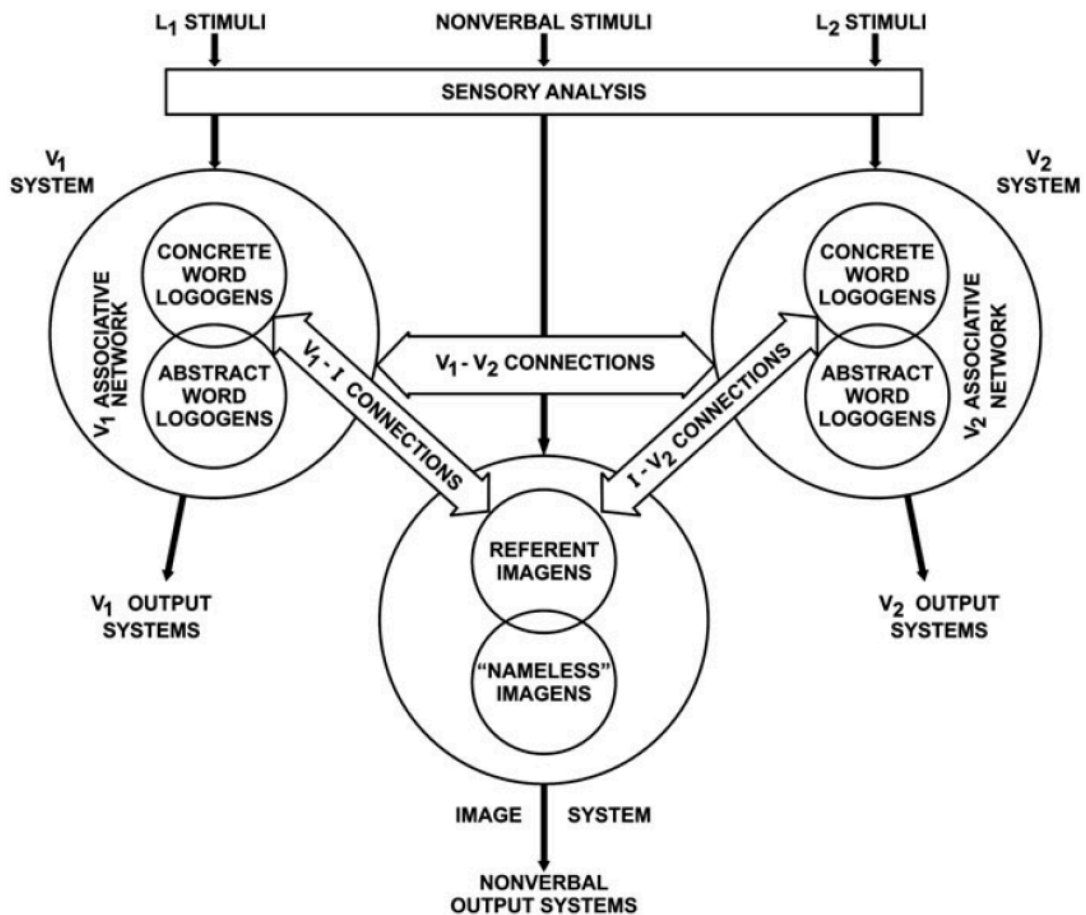


Fig. 11. The Bilingual Dual-Coding Theory (after Paivio and Desrochers 1980: 391)

Importantly, the theory stipulates the priority of concrete words, as only these can activate the image system. Concrete words are also postulated to involve higher imagery compared to abstract words, as they are coded in both the verbal and image systems, and they are thus easier to be recalled. Such an assumption is in line with the bilingual concreteness effect, which is demonstrated in faster and more accurate translations of concrete relative to abstract words (Heredia 2008: 58). The concreteness effect has been postulated to stem from the fact that concrete words bring about more imagery, are acquired earlier than abstract words, and are often of higher frequency. The role of word concreteness in bilingual lexical organization was also discussed in the Distributed Feature Model (de Groot 1992), which points to a potentially distinct organization of concrete and abstract words in the bilingual lexicon, as further presented in the following section.

3.2.1.3. Distributed Feature Model

The Distributed Feature Model (de Groot 1992) was proposed in order to further account for differences in the processing of different word types. The model claims that the structure of the bilingual lexicon is modulated by the lexical category of specific words. For instance, concrete translation equivalents are assumed to have a greater semantic overlap and to share more conceptual nodes across languages compared to abstract translation equivalents. Such an increased semantic overlap of concrete words across languages results in more robust spreading activation processes, and consequently faster translation. As a result, while the conceptual representations of concrete items as well as cognate words (lexical items having a similar form and meaning in the two languages) are argued to be shared across languages, abstract words' representations are assumed to be stored separately (Fig. 12).

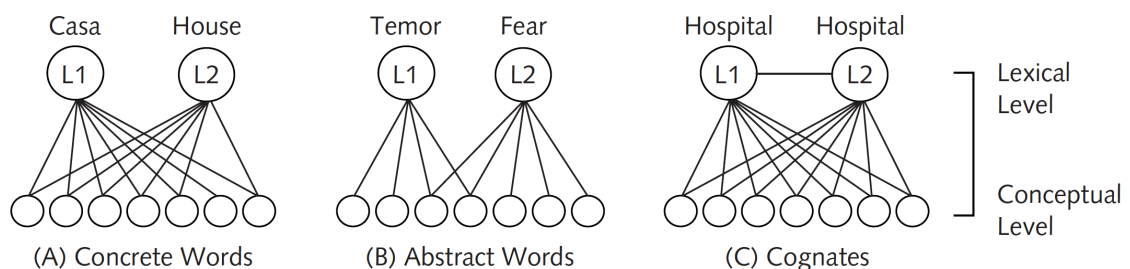


Fig. 12. The Distributed Feature Model (after Heredia and Brown 2006: 239, adapted from de Groot 1992).

Therefore, the model specifies that it is the degree of overlap between translation equivalents from the two languages that determines how fast bilingual speakers are able to access a specific lexical item. Consequently, lexical items whose meanings are similar across languages (e.g., concrete words) tend to share more meaning elements, and are thus postulated to be easier to retrieve from the semantic memory network. Abstract words, in contrast, are more culturally-determined and open to a subjective interpretation, as a result of which they are rarely shared across languages, may not always have one-to-one translation equivalents, and their meaning is consequently more cognitively taxing to

retrieve. It must, however, be noted that the Distributed Feature Model accounts only for single word processing, it does not account for how L1 and L2 information is accessed when processing language at a discourse level (Brysbaert and Duyck 2010: 367).

All of the above-discussed traditional perspectives on language processing in bilingualism suffer from a number of limitations. First of all, they have been proposed mostly to account for the structure-oriented issues, and have shown how information is stored in the bilingual memory system, thus failing to specify mechanisms engaged in bilingual lexico-semantic processing. Secondly, the assumption that L1 and L2 lexical level representations are separate might seem oversimplified, as this would imply that bilingual language processing at the lexical level is always language-selective. Next, even though the views point to different types of conceptual representations depending on the level of abstractness or cognate status of specific lexical items, they do not specify how other lexical variables, including word frequency or neighborhood density, could modulate language processing in bilingualism. The traditional views of bilingual language processing have therefore served mostly as a starting point in the further development of models on bilingual lexico-semantic processing. More current models have thus been proposed with a view to describing specific cognitive mechanisms engaged in bilingual language comprehension. The following two sections are devoted to discussing two best-grounded process-oriented models of bilingual language processing: the Hierarchical Model (section 3.2.2.1) as well as the Bilingual Interactive Activation Models (section 3.2.2.2).

3.2.2. Process-oriented perspectives on bilingual language processing

3.2.2.1. Hierarchical Model

The hierarchical model of bilingual language processing has adopted the traditional perspectives on how information is stored in bilingual memory, and it assumes that lexical items are represented in the human mind at both lexical and conceptual levels, thus having separate and independent representations for a word form and word meaning. Based on such core assumptions, the hierarchical model of bilingual memory further claims that

these representations are separate for the native (L1) and non-native tongue (L2) at the lexical level (word form), yet shared across the languages at the conceptual level (word meaning; Potter et al. 1984). Therefore, translation equivalents are hypothesized to have separate lexical representations, and a shared conceptual (semantic) representation. Furthermore, L1 lexicon is hypothesized to be more developed compared to L2 lexicon, due to a usually larger number of words known in the native compared to the non-native language (Potter et al. 1984, as cited in Heredia 2008: 53).

The hierarchical model of bilingual language processing is based on two core assumptions: word association and the concept mediation. According to the word association assumption, accessing the meaning of an L2 lexical item requires a preliminary activation of an L1 lexical item. Since L2 conceptual representations can be accessed only by means of the pre-activation of L1 representations, retrieving a meaning of an L2 word must be accompanied by L1 mediation (Potter et al. 1984, as cited in Heredia 2008: 54). In contrast, the concept mediation assumption argues for a direct access to a common meaning representation from both L1 and L2 lexical levels of representation, as a result of which both the native and non-native language can operate independently of each other. The two aforementioned assumptions were further developed and incorporated in the Revised Hierarchical Model (RHM; Kroll and Stewart 1994), which argues for a shift from word association into concept mediation once bilinguals become proficient in their L2. Therefore, the model can account for the difference in the strengths of connections between the two languages that are modulated by L2 proficiency level, and can thus demonstrate how language proficiency influences bilingual visual word processing.

The RHM postulates independent lexical-level representations for the two languages, and a common conceptual representation that is shared across the two linguistic systems (Kroll and Stewart 1994, as cited in Kroll and Dussias 2006: 176). The role of L2 proficiency in bilingual language processing is specified by proposing that non-proficient L2 learners access the conceptual representation of an L2 word by means of pre-activating its L1 translation equivalent, as argued in the word association assumption (Kroll and Stewart 1994, as cited in Randall 2007: 118). As a result, the non-native language, which is still being acquired, takes advantage of the already existing pathways between L1 form and meaning (Kroll and Stewart 1994, as cited in Kroll and Dussias 2006: 176). When L2 proficiency level increases, however, L2 speakers become capable

of forming straightforward connections between L2 form and meaning without the mediation of the native tongue, as postulated in the concept mediation assumption (Fig. 13).

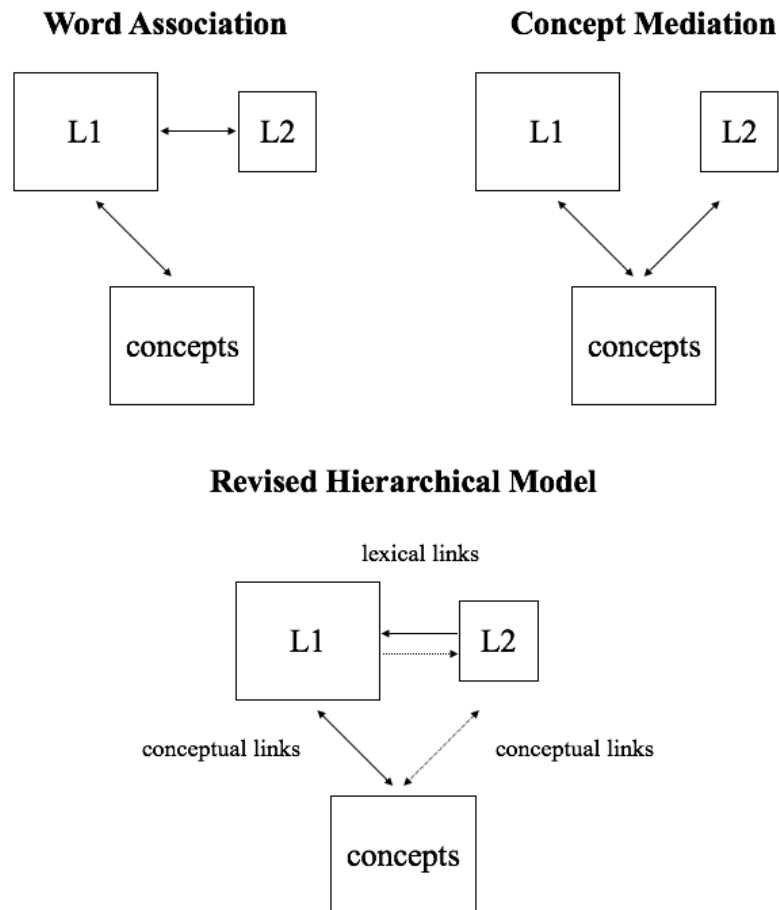


Fig. 13. Word Association, Concept Mediation, and the Revised Hierarchical Model of lexical and conceptual representation in bilingual memory (after Kroll and Stewart 1994: 158)

Importantly, in line with the Revised Hierarchical Model, the considerable reliance on the native language during early stages of L2 acquisition results in diversified connections between various levels of representations proposed by the model. First of all, L1 lexical representations are larger compared to L2 lexical representations, which results in more automatic access to L1 lexicon (Kroll and Stewart 1994, as cited in Kroll and Dussias 2006: 176). Secondly, L1 lexical-level representations have stronger connections with the conceptual representation than L2 lexical representations. Finally, links between

L1 and L2 lexical representations are stronger in the L2-L1 direction than in the L1-L2 direction. Based on the last assumption, the RHM has frequently been tested by means of examining the advantage of backward (L2-L1) over the forward (L1-L2) translation priming effects (Kroll and Stewart 1994, as cited in Heredia 2008: 55). In line with the model, in non-proficient bilingual speakers, L1 has a privileged access to semantic representations, and thus L2 items require the mediation of L1 translation equivalents, as only they can provide access to the conceptual level representations. Consequently, while L2-L1 translation can be accomplished at the lexical level, without meaning mediation, L1-L2 translation needs to be semantically mediated (Kroll et al. 2010: 373).⁵

The tenets proposed in the Revised Hierarchical Model have, nonetheless, undergone some criticism (see Brysbaert and Duyck 2010 for a review). First of all, by postulating a separate lexicon for L1 and L2, the model does not account for a host number of empirical studies showing how bilingual lexico-semantic processing is influenced by both L1 and L2 lexical representations (e.g., van Heuven et al. 1998; Spivey and Marian 1999; Dijkstra et al. 2000; van Hell and Dijkstra 2002; Duyck et al. 2007; van Heuven et al. 2008). Secondly, the assumption suggesting stronger lexical connections, and thus a more robust translation priming effect from L2 to L1, has actually received only little support, especially when testing different-script languages in a masked priming paradigm, which allows for examining mechanisms engaged in language processing that are automatic and strategy-free (e.g., Gollan et al. 1997; Jiang 1999; Jiang and Forster 2001). Next, the RHM does not take into account the fact that not all translation equivalents are one-to-one mappings. For instance, the meaning of abstract words is usually context-dependent, while concrete words are often polysemous (e.g., *pupil*) or might refer to different senses (e.g., *chicken* as an animal and as food), which therefore requires different translations. Furthermore, the assumption that all semantic information is language-independent is contrary to previous research on memory processing, which has shown that language cues activate memory traces, and therefore influence memory retrieval (e.g., Marian and Neisser 2000; Sahlin et al. 2005). Importantly, the model does not explain the role of phonology and orthography in bilingual language processing, and thus it remains unspecified how factors including word frequency or phonological/orthographic neighborhood

⁵ Electrophysiological research testing the tenets proposed in the RHM model is discussed in section 3.3.1.1.

density might influence bilingual lexico-semantic processing. Finally, the RHM formulates no predictions regarding how sentence or language context could affect bilingual word recognition, and does not account for whether and how different task requirements modulate mechanisms engaged in bilingual language processing. Due to the aforementioned challenges that the Revised Hierarchical Model cannot tackle, it seems crucial to discuss the connectionist models of bilingual lexico-semantic processing, the Bilingual Interactive Activation Models, which are based on the well-established and supported model of monolingual visual word recognition, the Interactive Activation Model (McClelland and Rumelhart 1981), and aim to capture neuronlike processing.

3.2.2.2. Bilingual Interactive Activation Models

The role of orthographic, phonological, and semantic features in bilingual visual word recognition, which is not specified in the Revised Hierarchical Model, is thoroughly discussed in Bilingual Interactive Activation Models (Dijkstra and van Heuven 1998; Dijkstra and van Heuven 2002), which argue for language non-selectivity with a shared lexicon across languages, and are based on the best-grounded model of monolingual visual word recognition, the Interactive Activation Model (McClelland and Rumelhart 1981).

The Interactive Activation Model (IA) is the best-established and one of the most influential localist-connectionist (neural-network) cognitive models of word recognition, which provides a link between neural activity and cognition by capturing the properties neurons and neuronal populations (Rogers 2009: 75). The IA model consists of the network with node levels that are connected to each other by means of excitatory and inhibitory links (McClelland and Rumelhart 1981, as cited in Norris 2013: 517). The model postulates three different node levels that contribute to monolingual visual word recognition; namely, word, letter, and feature levels (Fig. 14). When a string of letters is presented, features shared by particular letters become activated, which leads to the subsequent activation of letters sharing those features, and the inhibition of letters that do not share them (McClelland and Rumelhart 1981, as cited in De Angelis 2007: 105). Importantly, the model highlights the uniform role of both top-down and bottom-up operations during language comprehension, as both top-down and bottom-up information is argued to be activated in parallel. While bottom-up mechanisms involve the analyses of

multiple letter features, top-down operations are based on background knowledge as well as contextual information (Sternberg 1996: 215).

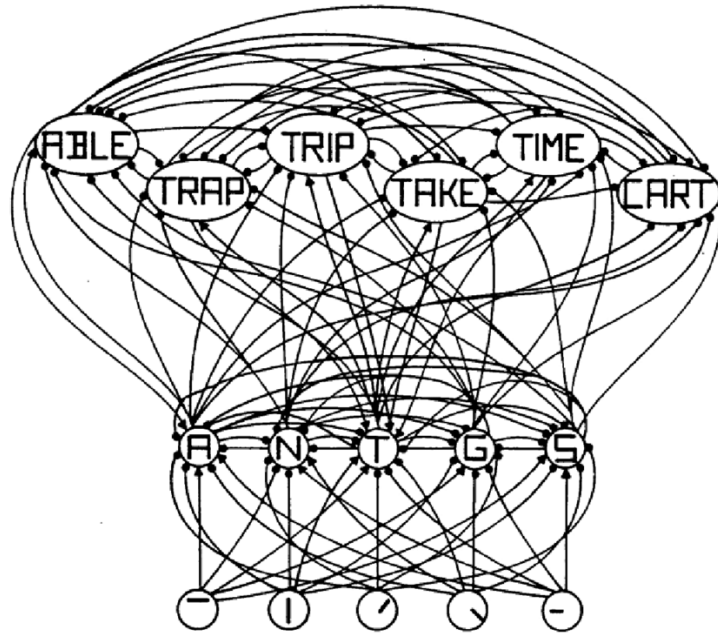


Fig. 14. The Interactive Activation Model of monolingual visual word recognition (after McClelland and Rumelhart 1981: 380).

Similarly to the IA, the Bilingual Interactive Activation Model (BIA; Dijkstra and van Heuven 1998) posits that the process of bilingual visual word recognition takes place at the level of features, letters, and words. Additionally, the model introduces the level of a language node, which is responsible for specifying language membership by means of selecting the target language and inhibiting the non-target tongue (Fig. 15). For instance, when reading cognate words (words sharing a similar form and meaning across languages; e.g., *MUSIC* – *MUZYKA* as English-Polish cognates) or interlingual homographs (words having a similar form, but different meanings across languages; e.g., *EVENTUALLY* – *EWENTUALNIE* as English-Polish interlingual homographs), both the native and non-native tongue become initially activated, but at the end the language node inhibits the non-target language, thus allowing for the activation of the target language only. Therefore, the BIA model argues that the lexicon is integrated up to the word level, and it postulates separate language node levels.

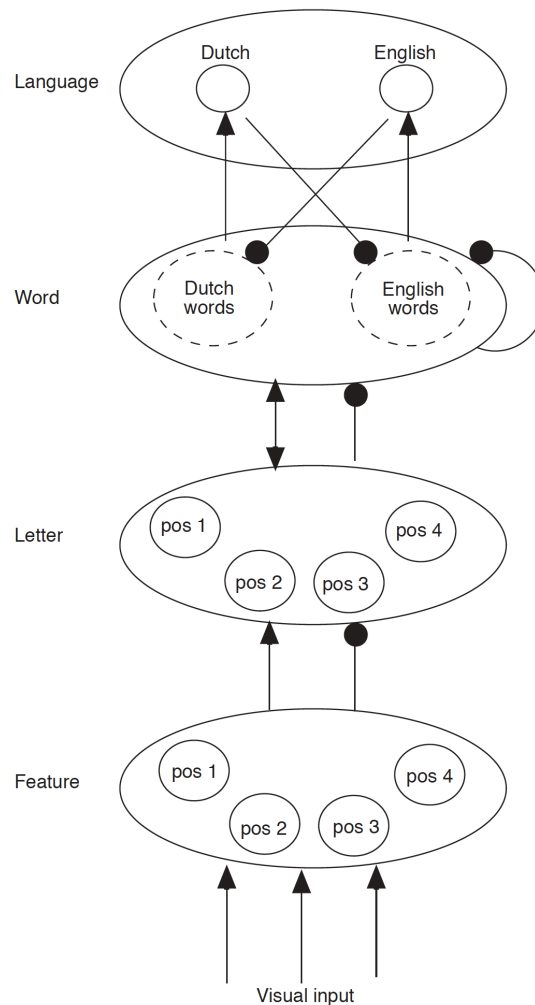


Fig. 15. The Bilingual Interactive Activation Model (BIA) of bilingual word recognition (after Dijkstra and van Heuven 2002: 177, adapted from Dijkstra and van Heuven 1998).

However, the BIA model does not specify the role of phonology and semantics in bilingual word recognition, and it fails to account for how linguistic and non-linguistic contexts modulate bilingual language processing. To address these limitations, the BIA model was further developed into the Bilingual Interactive Activation Plus Model (BIA+; Dijkstra and van Heuven 2002). The BIA+ model claims that bilingual word recognition is influenced by not only orthographic similarity, but also phonological and semantic overlap. The BIA+ model, similarly to the BIA, argues for a nonselective language access within the integrated lexicon of the native and non-native tongue, and highlights the pre-

activation of both languages during bilingual visual word processing. The model further specifies that the process of visual word recognition starts at the level of sublexical orthographic representations, followed by sublexical phonological representations. This is followed by bottom-up and top-down activation of whole-word orthographic and phonological representations (Fig. 16). The activated representations then trigger semantic representations and language membership representations. Importantly, all of the levels of representations involve both bottom-up and top-down flow of activation, except for the language nodes, which involve only the bottom-up activation.

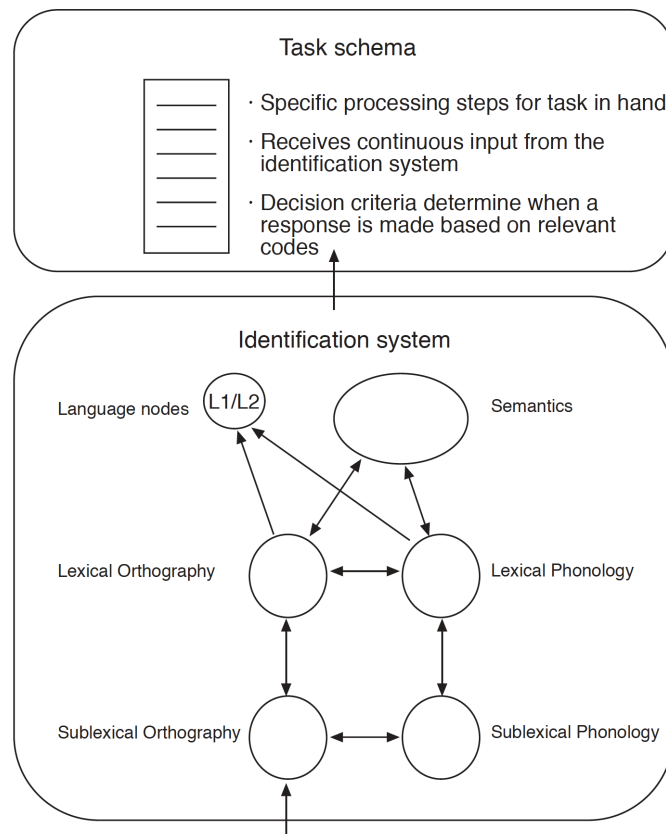


Fig. 16. The Bilingual Interactive Activation Plus Model (BIA+) of bilingual word recognition (after Dijkstra and van Heuven 2002: 182).

Importantly, the model stipulates that the strength and automaticity of connections between different levels of representations is modulated by factors such as language dominance, language context, and everyday language use. As a result, L2 unbalanced (L2

non-dominant) bilingual speakers, who use their L2 less frequently than L1, have a lower resting state activation of the non-native tongue compared to the native language, and thus their non-native language is processed more slowly and less automatically. A lower subjective frequency of L2 codes in less proficient and L2 non-dominant bilinguals results in a temporal delay, as the activation of L2 phonological and semantic representations is argued to occur later compared to the activation of L1 representations (Dijkstra and van Heuven 2002: 183).

The BIA+ model also introduces the node of task schema, which shows how non-linguistic context effects and task-related effects impact the process of word recognition. The distinction between the identification system and a task/decision system is based on the assumption that when performing a task, participants engage both preconscious, automatic processes as well as attention-sensitive mechanisms that are modulated by contextual factors, including task requirements. Consequently, different tasks require different task schemas. The task schema stipulates that performing a specific task requires mental processing steps that need to be carried out. Such operations can be either retrieved from memory or set up during the practice trial, which results in setting specific recognition threshold and temporal deadlines. For example, in a lexical decision task, in which participants need to decide whether a presented string of letters is a real word or not, a selection threshold might be lower in the case of nonwords, which are easy to reject. Importantly, the BIA+ specifies that the identification system is partially independent from the task/decision system, and thus even though the identification system provides a continuous output to the task/decision system, the latter serves also some independent functions. For instance, although a lexical item may be fully recognized by the identification system, the decision regarding that item might be withheld because some information needed to perform a task is not yet available. Furthermore, participants may sometimes make their decision before lexical identification is complete (Dijkstra and van Heuven 2002: 190-191).

Importantly, unlike the original BIA model, the BIA+ can account for both linguistic and non-linguistic context effects. While the linguistic sentence context is assumed to affect the word identification system, the non-linguistic context, including task requirements, instructions, participants' expectations and strategies, should influence the task/decision system. The linguistic context effects, including syntactic and semantic ef-

fects, are further argued to be language non-selective, and therefore semantic and syntactic constraints modulate word recognition regardless of whether the sentence context or the critical word is presented in one language or the other, thus allowing for a cross-linguistic priming effects (Dijkstra and van Heuven 2002: 188).

It needs to be noted that the BIA+ model assumes that language membership information does not affect word recognition, as it is available at the later stages of language processing. Contrary to this assumption, previous research has, however, shown that language membership information is likely to be available prior to semantic as well as lexical information identification (e.g., Vaid and Frenck-Mestre 2002; Casaponsa et al. 2014; Hoversten et al. 2015; Oganian et al. 2016; Hoversten et al. 2017). Consequently, orthographic membership cues might modulate word identification mechanisms by restricting cross-language activation. The role of sublexical features in bilingual visual word recognition was elaborated on within a more recently introduced extended version of the BIA+ model (Van Kesteren et al. 2012). In their proposal, Van Kesteren and colleagues (2012) further emphasize the crucial role of bottom-up mechanisms during bilingual lexico-semantic processing by introducing the sublexical language membership node, which can be accessed directly from the sublexical level representations.

As further specified in the extended BIA+ model (Van Kesteren et al. 2012), the sublexical language membership nodes allow for the activation of the target and the inhibition of the non-target language system based on orthographic properties of a given language. For instance, when presented with orthographically marked lexical items, whose sublexical features (e.g., bigrams) are plausible only in one language, the system can automatically activate the language membership node, without the obligatory mediation of phonological, lexical, and semantic representations (Fig. 17). Importantly, the authors assume that the facilitation effect observed when processing language-specific bigrams is present only when participants' sensitivity to language markers facilitates their performance (e.g., in lexical decision or language decision tasks). Consequently, Van Kesteren et al. (2012: 2147) postulate that such a facilitation effect is dependent on the task/decision system, which uses the language membership information that is retrieved from the sublexical level of representations.

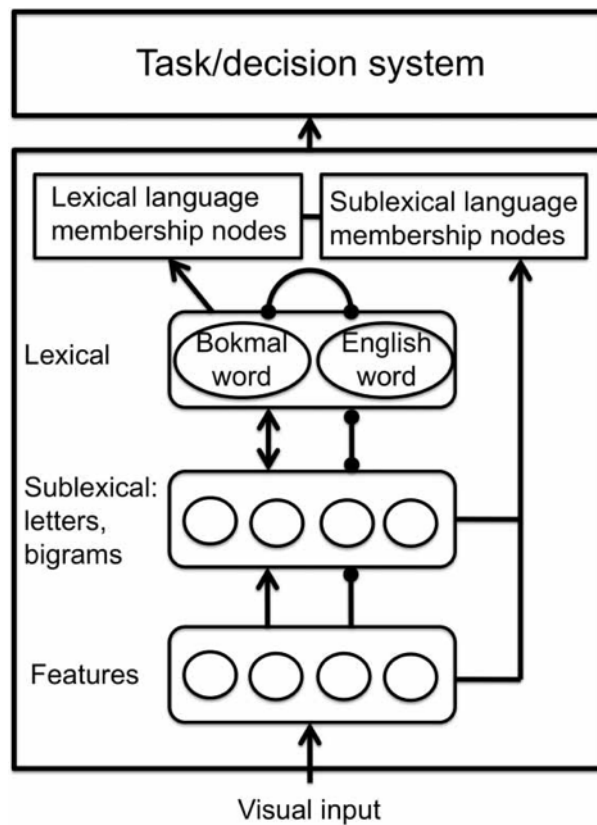


Fig. 17. The extended BIA+ model of bilingual word recognition (after Van Kesteren et al. 2012: 2148).

To date, the above-discussed Bilingual Interactive Activation Models seem to provide the most thorough insights into lexico-semantic processing in bilingualism, and their arguments seem to be much more compelling compared to the previous models and theoretical accounts for several reasons. Firstly, the BIA models are based on the model of monolingual language processing (the Interactive Activation Model; McClelland and Rumelhart 1981), which is already well-established and supported, and thus the BIA models provide a straightforward and systematic extension of the monolingual system by adding the L2 network. Secondly, unlike other theoretical frameworks, the BIA models provide a connectionist perspective on bilingual language processing, thus capturing neuronlike processing during lexico-semantic processing in bilingualism that shows a clear and direct link between neural activity and cognition. Next, the BIA models can account for a great majority of the limitations that the previous views on bilingual language processing suffer from (see sections 3.2.1 and 3.2.2.1). Namely, not only does the BIA+ model specify how language information is processed at the sublexical, lexical, and conceptual levels

of representations, but it also predicts how linguistic context affects bilingual word recognition. The BIA+ model additionally introduces the task/decision system, thus showing how task requirements modulate bilingual lexico-semantic processing. Furthermore, within the temporal delay assumption, the BIA+ model predicts how factors such as subjective frequency of L2 lexical items, which is potentially modulated by language dominance, influence the activation of semantic and phonological information in the non-native tongue. Finally, the arguments discussed within the aforementioned BIA models have been vastly explored empirically by means of conducting behavioral and electrophysiological studies, which have provided strong support for the language non-selective access view.⁶

3.3. Bilingual lexico-semantic processing: Empirical evidence

3.3.1. The N400 and lexico-semantic access in bilingualism

Bilingual studies examining the N400 response, which is an indicator of mechanisms involved in information retrieval from long-term memory, can provide valuable insights into bilingual lexico-semantic processing by showing the amount of information that needs to be retrieved from the semantic network (as reflected in the N400 magnitude) as well as the time course of lexico-semantic access in L1 and L2 (as reflected in the N400 peak latency). Additionally, owing to the competing theories of the functional role of the N400 response (see section 1.4.3.1), research into the N400 response that is extended from the monolingual to the bilingual context might clarify the specific role of the N400 effect in visual word processing. For instance, between-language comparisons of the N400 amplitudes can provide relevant information regarding the degree to which the component is modulated by bottom-up and top-down mechanisms during lexico-semantic access. While the role of top-down mechanisms can be elucidated by examining the influence of context or task requirements on bilingual lexico-semantic processing, bottom-up processes can be explicated when investigating how sublexical features influence ERP patterns in the native and non-native language processing. ERP experiments into bilingual

⁶ Electrophysiological research testing the tenets proposed in the BIA models is discussed in section 3.3.1.2.

language processing conducted thus far have, however, yielded quite inconclusive results, mostly as a result of methodological differences, such as different tasks and stimuli, languages with different orthographic systems, types of bilingual speakers, or types of experimental designs (see Jankowiak and Rataj 2017: 122-123). The following subsections are devoted to discussing studies into the N400 response in bilingual lexico-semantic access which have tested the Revised Hierarchical Model (section 3.3.1.1) as well as the Bilingual Interactive Activation Models (section 3.3.1.2), and have examined the role of language proficiency, language dominance, and age of L2 acquisition in language processing in bilingualism (section 3.3.1.3).

3.3.1.1. The N400 effect in studies on translation priming: Testing the Revised Hierarchical Model

Previous behavioral studies that tested the Revised Hierarchical Model (Kroll and Stewart 1994; see section 3.2.2.1) have revealed shorter reaction times (RTs) when words from the two languages were presented in the backward (L2-L1) compared to the forward (L1-L2) translation direction, therefore providing support for the RHM model (e.g., La Heij et al. 1996; Talamas et al. 1999; Kroll et al. 2002; Sunderman and Kroll 2006). The tenets proposed by the RHM have also been tested by employing event related potentials (ERPs), which provide information regarding cognitive mechanisms involved in lexico-semantic access during forward and backward translation priming, as reflected in the N400 response. In line with the RHM, in non-proficient bilingual speakers, upon a presentation of an L2 lexical item on backward translation trials (L2-L1), an L1 translation equivalent needs to be activated, resulting in a greater amount of information retrieved from the semantic memory store, which should be reflected in larger N400 amplitudes. On forward translation trials (L1-L2), in contrast, an L1 word directly activates the conceptual representation, resulting in an attenuated N400 response compared to the backward translation direction.

Importantly, reduced N400 amplitudes for translation equivalents presented in both translation directions compared to semantically unrelated lexical items might reflect the N400 semantic priming effect, in which attenuated N400 responses are observed to

words preceded by semantically related items. Therefore, the magnitude of the N400 response to L1-L2 as compared to L2-L1 translation direction provides information regarding the strength of the semantic priming effect in the two conditions, and consequently shows how strongly words from the two languages are connected to each other. The following part of this section is devoted to reviewing ERP studies conducted with a view to testing how the automaticity of translation priming in the forward and backward translation direction could be modulated by language proficiency.

Non-proficient bilingual speakers have been tested in three ERP experiments. Alvarez et al. (2003) examined English (L1) - Spanish (L2) bilinguals, who were presented with within-language repetitions (i.e., English-English or Spanish-Spanish) and between-language translation equivalents (i.e., English-Spanish or Spanish-English). A control condition was additionally used, and included semantically unrelated word pairs. Native English speakers, who were beginning and intermediate learners of Spanish (L2) performed a mixed-language semantic categorization task, in which they detected lexical items referring to parts of the body in either English or Spanish. The results revealed a graded N400 effect across the conditions, with the most robust N400 response to the control condition, followed by a between-language condition, and finally a within-language condition. Importantly, though the N400 amplitudes were of similar magnitude for both forward and backward translation directions, the authors observed a delay in the N400 peak latency for the between-language condition in the L1-L2 translation direction as compared to the L2-L1 direction, thus suggesting that translation direction might modulate the time course of bilingual language processing. Nonetheless, such a delay in the N400 peak latency was postulated only based on visual inspection, and thus further research is needed in order to confirm such conclusions.

Non-proficient bilinguals have also been tested by means of employing a masked priming paradigm. Such a paradigm was used by Hoshino et al. (2010), who tested unbalanced intermediate Japanese-English bilinguals. Participants were exposed to cross-script translation equivalents presented in both L1-L2 and L2-L1 translation directions, and performed a semantic categorization task. ERP results revealed the N400 semantic priming effect only in response to translation equivalents presented in the L1-L2 translation direction, which might have resulted from a low L2 proficiency level. The two above-discussed ERP studies on non-proficient bilingual speakers seem to support the tenets of the Revised Hierarchical Model, as they point to the asymmetry of the N400 response,

with more robust N400 amplitudes for the L2-L1 relative to the L1-L2 translation directions. Counterevidence was, on the other hand, provided by Midgley et al. (2009a), who employed a masked priming paradigm, and tested late unbalanced intermediate French-English bilingual speakers in a go/no-go semantic categorization task for animal names. As previously explained in section 1.3.1, in a go/no-go procedure, participants respond to one type of stimuli (the *go* trials), but not to the other type of stimuli (the *no-go* trials). In the study by Midgley and colleagues (2009a), the critical words were preceded by either lexical repetitions, translation equivalents, or semantically unrelated items. Translation equivalents were presented in the L1-L2 direction in Experiment 1, and in the L2-L1 direction in Experiment 2. The authors found the N400 semantic priming effect, with attenuated N400 amplitudes for translation equivalents compared to semantically unrelated items, when the translation equivalents were presented in both L1-L2 and L2-L1 directions. Such results seem difficult to account for within the tenets of the Revised Hierarchical Model, according to which the translation priming effect should be stronger in the forward relative to backward translation direction trials when bilinguals are not proficient in their L2.

The following three ERP experiments have been conducted on highly proficient bilingual speakers. Geyer et al. (2011) tested highly proficient Russian-English bilinguals in a mixed-language lexical decision task, in which participants were presented with translation equivalents in both L1-L2 and L2-L1 directions, and decided whether or not the critical lexical item was a real word in the target language. The authors found a symmetrical N400 semantic priming effect along with similar reaction times for both forward and backward translation directions, thus lending support to the Revised Hierarchical Model, according to which L2 proficiency level modulates translation priming. Highly proficient bilingual speakers were also tested by Moldovan et al. (2016). In their experiment, Catalan-Spanish bilingual speakers performed a translation recognition task in response to Catalan-Spanish pairs of words highly related in meaning, pairs of words moderately related in meaning, as well as pairs of words whose second item had a similar phonological and orthographic form to the correct translation of the first item (i.e., translation neighbors). All of the word pairs were always presented in the Catalan-Spanish direction. ERP results showed a more robust N400 response to translation neighbors compared to words that were highly and moderately related in meaning. Such results are in line with the interference effect, which in a translation recognition task is present when

participants are exposed to semantically related words, whose meanings they need to suppress as these are not task-relevant, and which is reflected in larger N400 amplitudes for trials with semantically related but task-irrelevant items compared to semantically unrelated words.

Similarly to the experiments on non-proficient bilinguals, a masked translation priming paradigm has been employed when studying proficient bilingual participants. Duñabeitia et al. (2010) tested proficient Basque-Spanish bilingual speakers in a lexical decision task in response to between-language translation equivalents (presented in both translation directions), within-language prime-target repetitions (presented in both Basque and Spanish), as well as the control condition with semantically unrelated items. The results showed the N400 semantic priming effect in response to translation equivalents presented in L1-L2 as well as L2-L1 directions, both of which evoked an attenuated N400 response relative to the control condition. Such results further suggest that highly proficient bilingual speakers are able to access translation equivalents similarly automatically in the two translation directions.

The abovementioned evidence in favor of the Revised Hierarchical Model points to a positive correlation between L2 proficiency level and the automaticity of mechanisms engaged in forward and backward translation directions, as a symmetrical N400 response in both translation directions was found in proficient bilingual speakers (Geyer et al. 2011; Duñabeitia et al. 2010), but not in bilinguals at a lower level of L2 proficiency (Hoshino et al. 2010; but see Midgley et al. 2009a). Importantly, notwithstanding the evidence in support of the RHM, the model does not account for different levels of representations, and it therefore does not specify how orthography, phonology, and semantics influence bilingual visual word recognition, as concluded by Jankowiak and Rataj (2017: 125). In addition, the model is selective in its nature, as it argues for separate L1 and L2 lexicons, and therefore cannot account for strong evidence indicating cross-linguistic influence during bilingual language processing (e.g., Brysbaert 1998; van Heuven et al. 1998; Spivey and Marian 1999; Dijkstra et al. 2000; van Hell and Dijkstra 2002; Brysbaert and Dijkstra 2006; Duyck et al. 2007). It thus seems relevant to discuss research aimed at testing the language non-selective access view (i.e., the BIA models), which is discussed in the following section.

3.3.1.2. The N400 effect in studies on language non-selectivity: Testing the Bilingual Interactive Activation Models

Experiments testing the tenets proposed within the BIA and BIA+ models (see section 3.2.2.2) have frequently examined the processing of cognates and interlingual homographs, which are regarded as markers of cross-language co-activation, and thus provide insights into whether bilingual language processing is selective or non-selective. In favor of the language non-selective access view, a number of behavioral experiments have shown faster RTs for cognate words and longer RTs for interlingual homographs relative to control words, thus indicating a simultaneous co-activation of both languages during lexico-semantic processing (e.g., Dijkstra et al. 1998b; Costa et al. 2005; Lemhöfer et al. 2008). Such results suggest that cognate words, due to their cross-language overlap in meaning and form, bring about the cognate facilitation effect, which might result from common morphemic representations across languages (Lalor and Kirsner 2000). Interlingual homographs, on the other hand, are assumed to have two (potentially overlapping) representations in the bilingual lexicon (Dijkstra and van Heuven 2002: 185), and thus they require extended mechanisms in order to activate the target meaning and inhibit the non-target semantic representation. Such assumptions have been tested by means of examining the N400 response to cognates and interlingual homographs. The cognate facilitation effect should result in an attenuated N400 response to cognate relative to non-cognate words. In contrast, the activation of two representations of interlingual homographs should involve an increased activity in long term memory, which should in turn be reflected in more pronounced N400 amplitudes compared to control words. The following part of this section is devoted to presenting ERP studies which have investigated language non-selectivity by examining the role of lexical (e.g., word frequency) as well as sublexical and subphonological features (e.g., language scripting, orthographic neighborhood, orthotactic properties) in cognate and interlingual homograph processing.

In one of the studies on interlingual homograph processing, Hoshino and Thierry (2012) tested unbalanced Spanish-English bilinguals and English monolinguals in an English semantic relatedness task, in which participants decided whether the presented words were semantically related or not. Interlingual homographs used in the study were either related to the Spanish meaning, to the English meaning, or were unrelated. In the monolingual group, the N400 semantic effect was found only in response to homographs

related to English (L1) meanings. In the bilingual group, on the other hand, the effect was elicited in response to both Spanish (L1) and English (L2) homographs, even though Spanish was task-irrelevant. Such results therefore provide support for an automatic co-activation of the target and non-target tongue during bilingual visual word recognition, even when only one language is task-relevant.

Interlingual homographs and cognate words have been further investigated by means of manipulating their frequencies, with a view to examining how lexical features of critical words modulate their processing. For instance, Peeters et al. (2013) tested late advanced French-English bilingual speakers in an English (L2) lexical decision task, in which participants decided whether the presented lexical item was a real English word. Importantly, the researchers manipulated the frequency of the critical words, and included identical cognate words that were either of high frequency in both French (L1) and English (L2), of low frequency in French and high frequency in English, of high frequency in French and low frequency in English, or of low frequency in both French and English. The authors found the cognate facilitation effect in behavioral results, with shorter RTs for cognates than non-cognates. ERP results further revealed that the N400 cognate facilitation effect was attenuated in response to cognates of high frequency in L1 and low frequency in L2. Larger N400 amplitudes for lexical items that were frequent in the task-irrelevant tongue, and at the same time non-frequent in the task-relevant language might have resulted from the fact that participants needed to inhibit the activation of highly frequent words in the task-irrelevant tongue in order to select the target language. In another study, Kerkhofs et al. (2006) manipulated the frequency effect of interlingual homographs, and tested late proficient Dutch-English bilinguals in an English (L2) lexical decision task. The results showed the N400 frequency effect to interlingual homographs of low frequency in English (L2); yet, the effect was reversed in Dutch (L1), in which high-frequency items evoked a more robust N400 response. Such a reverse effect might have stemmed from the fact that Dutch was task-irrelevant, and hence participants might have found it relatively demanding to inhibit the activation of the Dutch meaning of a homograph, which was of high frequency in their native tongue. Overall, the results obtained from the two experiments indicate that the N400 frequency effect in bilingualism depends on task requirements, and may be additionally modulated by the language mode participants are currently in.

In addition to examining lexical features, some researchers have investigated the role of orthography and phonology in interlingual homograph and cognate processing, thus showing how sublexical and subphonological features influence bilingual visual word processing. The interplay between orthography and semantics can be well tapped into when studying different scripting languages. For instance, Jouravlev and Jared (2014) tested balanced proficient Russian-English bilingual speakers and English monolinguals in an English reading task. The materials included sentences with interlingual homographs, English translations of the Russian meaning of a homograph (i.e., English Match), and unrelated control words (i.e., English Mismatch), which were always presented in a mid-sentence position. The results showed a linear effect across the conditions for the bilingual group, with the most robust N400 amplitudes for the English Mismatch, followed by interlingual homographs, and finally the English Match condition. Such results suggest that the Russian meanings of the homographs were activated despite the purely English context of the task, which points to a simultaneous co-activation of both L1 and L2 semantic information, even when the two languages do not share the same script. In another study, Thierry and Wu (2007) employed an English semantic relatedness task, and tested late proficient Chinese-English bilingual speakers as well as English and Chinese monolinguals, who were presented with English words, half of which concealed a character repetition when translated into Chinese. For example, an English word pair *train-ham*, shared a common character when translated into Chinese (*Huo Che-Huo Tui*). Interestingly, in the absence of any differences in reaction times, such items elicited the N400 priming effect, yet only in the bilingual group, which further points to an automatic co-activation of both L1 and L2 semantic information, irrespectively of language scripting.

Other ERP experiments focusing on orthographic features have examined the influence of orthographic neighborhood size (i.e., the number of words that can be formed by changing only one letter, preserving letter positions, such as *HORSE – HOUSE*) and orthotactic plausibility on bilingual language processing. The role of orthographic neighborhood size was investigated by Midgley et al. (2008), who tested late proficient French-English bilingual speakers. Participants performed a go/no-go semantic categorization task for French (L1) and English (L2) lexical items, which were manipulated in terms of the number of cross-language orthographic neighbors in the other language. The results revealed the N400 neighborhood effect, with a more pronounced N400 response to words

having many neighbors in the other tongue. Such results indicate a simultaneous co-activation of lexical representations of the neighbors in both L1 and L2, which therefore suggests that the activation of orthographic level representations is also language non-selective. In another study, Casaponsa et al. (2015) examined orthotactic plausibility by employing a go/no-go paradigm with a masked priming paradigm, and tested early proficient Spanish-Basque bilinguals in their L1 and L2. While target words were always presented in Spanish, prime words were manipulated in terms of their orthotactic plausibility in L1 and L2. Namely, prime words' orthographic regularities were either marked (implausible in L1, yet plausible in L2) or unmarked (plausible in both L1 and L2). The results showed a more pronounced N400 response to marked prime words' bigrams compared to unmarked bigrams, which indicates an automatic access to sublexical orthographic representations of both L1 and L2. Importantly, such results lend support to the extended BIA+ model (Van Kesteren et al. 2012; see section 3.2.2.2), according to which sublexical features can be used to retrieve language membership information directly from the sublexical level, which can thus facilitate the process of bilingual visual word recognition.

In addition to studying how orthographic features influence language non-selectivity, Comesaña et al. (2012) employed a masked priming paradigm, and examined the role of orthographic as compared to phonological overlap in cognate word processing. The authors tested proficient Portuguese-English bilingual speakers, who were presented with both orthographic and phonological cognates words, orthographic cognates, phonological cognates, and non-cognate control words. Participants pressed the spacebar once they read the presented word. The N400 cognate facilitation effect was observed in response to both orthographic and phonological cognates as well as orthographic cognates relative to phonological cognates, which elicited larger N400 amplitudes. Such results point to a more salient role of orthographic compared to phonological overlap in cognate processing. Importantly, the results might be interpreted as in favor of the BIA+ model, which specifies that sublexical orthographic representations are activated prior to sublexical phonological representations during bilingual visual word recognition.

As the discussion above attests, electrophysiological research conducted so far has provided strong support in favor of language non-selectivity, suggesting a simultaneous co-activation of both the native and non-native tongue during bilingual visual word recognition (see Jankowiak and Rataj 2017). The view has been supported even in studies on

different-scripting languages, pointing to a shared lexicon irrespectively of language characteristics (Thierry and Wu 2007; Jouravlev and Jared 2014). Furthermore, research conducted thus far has shown that lexical and sublexical/subphonological features also modulate the N400 response to the native and non-native tongue (Peeters et al. 2013; Kerkhofs et al. 2006; Midgley et al. 2008; Comesaña et al. 2012).

3.3.1.3. The N400 sensitivity to L2 proficiency level, language dominance, and age of L2 acquisition

Studies into the N400 response in bilingual language processing have also investigated the N400 sensitivity to the age of L2 acquisition (AoA), L2 proficiency level, as well as language dominance, and have revealed that these factors might modulate both the N400 magnitude as well as its peak latency. For instance, Weber-Fox and Neville (1996) tested English monolingual participants, as well as five groups of Chinese-English bilinguals. The five groups comprised of bilingual speakers whose L2 exposure started at 1-3, 4-6, 7-10, 11-13, and > 16 years of age. The researchers employed a self-paced reading task, and presented their participants with English sentences which were semantically and grammatically correct or incorrect. First, more pronounced N400 amplitudes were observed for incorrect compared to correct utterances in all groups of bilinguals. Second, the results revealed a delayed N400 response to English stimuli, yet only in late bilinguals, who had acquired their non-native tongue after the age of 10. This points to the sensitivity of the N400 response to the age of L2 acquisition, and indicates that the N400 peak latency might be modulated by years of exposure to L2.

In another study, the role of proficiency level together with age of L2 acquisition was examined by Moreno and Kutas (2005), who tested Spanish-English bilinguals divided into two groups: early proficient English-dominant, and late non-proficient Spanish-dominant bilinguals. Participants were presented with semantically plausible and implausible sentences in a self-paced reading task. ERP results revealed a delay in the N400 peak latency to the non-dominant language in both groups of participants, irrespectively of the condition. The observed results indicate that the N400 latency may not necessarily reflect differences in AoA, but it might instead be modulated by L2 proficiency level. Such an assumption was further supported by Braunstein et al. (2012), who compared

proficient and intermediate German-English bilingual speakers in an L2 plausibility judgment task. The researchers found a negative correlation between the N400 latency and L2 proficiency level, with an earlier N400 peak latency in more proficient bilingual speakers.

The role of proficiency level was further examined by Kotz and Elston-Güttler (2004), who tested highly proficient and non-proficient German-English bilinguals. Participants performed an English (L2) lexical decision task for lexical items presented in an associative and categorical priming paradigm. In an associative priming paradigm, prime and target items are associatively, but not necessarily semantically related to each other due to their high probability of co-occurrence. For instance, *cats* and *dogs* are associatively related, as they co-occur in the fixed phrase *raining cats and dogs* (Matsukawa et al. 2005). In a categorical priming paradigm, prime and target words belong to the same semantic category (e.g., *bed* and *table*). Consequently, while an associative priming paradigm shows how words are linked to each other in the lexicon, a categorical priming paradigm reveals how strong the connections between lexical and semantic representations are. Reaction time results obtained by Kotz and Elston-Güttler (2004) revealed an associative priming effect in both groups of participants; yet, a categorical priming effect was found only in the proficient group. Similarly, while the N400 priming effect in an associative priming condition was observed in both L2 proficient and non-proficient bilingual speakers, a categorical priming condition elicited the N400 priming effect only in the group of proficient bilinguals. This suggests that links between lexical items might be formed prior to connections between lexical and semantic representations, as these may be established with an increase in L2 proficiency level.

Proficient and non-proficient bilingual speakers were also tested by Phillips et al. (2004), who examined English-French bilinguals and employed a semantic categorization task with English and French semantically related as well as unrelated words. The authors found a negative correlation between proficiency level and RTs for L2 items, as longer response times for the non-native tongue were observed in non-proficient bilingual speakers. Regarding ERP results, the N400 priming effect in L2 was observed only in proficient bilinguals; yet, the N400 peak latency was delayed for around 50 ms in both groups of participants, pointing to a delay in bilingual lexico-semantic access, irrespectively of L2 proficiency level. The N400 latency shift is contradictory to what Braunstein et al. (2012) observed, which might have, however, stemmed from differences in task requirements. Namely, while Phillips et al. (2004) employed a semantic categorization task in both L1

and L2, Braunstein et al. (2012) asked their participants to perform a plausibility judgment task in response to L2 items only, which may have resulted in a different resting level state of L1 and L2 in the two experiments.

Importantly, modulations in the N400 peak latency observed in response to the native and non-native tongue might lend support to the temporal delay assumptions postulated within the BIA+ model (Dijkstra and van Heuven 2002; see section 3.2.2.2), according to which the activation of semantic and phonological information is delayed in L2 relative to L1. Such a delay should be mostly observed in less proficient and L2 non-dominant bilingual speakers, who use their native tongue much more frequently than their non-native language, and consequently their subjective frequency of L2 items is hypothesized to be lower relative to L1 words, resulting in slower and less automatic access to L2 lexical items, as reflected in a delayed N400 peak latency.

In addition to studying differences in the N400 magnitude and temporal properties during the native and non-native language comprehension, three ERP experiments have reported clear discrepancies in the N400 scalp distribution to L1 as compared to L2, possibly modulated by the level of proficiency in L2. For example, Proverbio et al. (2002) tested proficient Italian-Slovenian bilingual speakers and Italian monolinguals in a correctness judgment task for semantically or grammatically correct and incorrect sentences. The results revealed that both semantically and grammatically incorrect items elicited an attenuated and left-lateralized N400 response in the non-native relative to the native tongue, in which the effect was more right-lateralized. In another study, Newman et al. (2012) tested proficient Spanish-English bilinguals and English monolinguals in a semantic decision task. The researchers observed a reduced and delayed N400 response to the non-native language in the bilingual group, which was widely distributed, unlike the effect observed in response to the native tongue, which was more left-lateralized. Nonetheless, since the two studies compared L2 users against monolingual speakers, the observed results might have reflected different cortical representations of language in bilingual as compared to monolingual language users (Mechelli et al. 2004; Kovelman et al. 2008), and therefore they might not fully show the extent to which the observed differences in the N400 scalp distribution to L1 and L2 might be accounted for by language proficiency.

To provide further evidence into the role of L2 proficiency in the N400 scalp distribution, Midgley et al. (2009b) tested both intermediate unbalanced and highly profi-

cient French-English bilinguals in a go/no-go semantic categorization task for words presented in L1 and L2. The authors observed a reduction in the posterior N400 response to L2 relative to L1 in the group of intermediate, but not proficient bilingual speakers. These results were interpreted by the authors as indicative of a weaker interconnectivity between L2 lexical items in the semantic memory network when bilinguals are non-proficient. Such an interpretation is in line with the functional role of the N400 response, which claims that the N400 indexes effort related to information retrieval from long term memory (Kutas and Federmeier 2000; Kotz et al. 2012; see section 1.4.3.1). In addition to the reduced posterior N400 amplitudes, Midgley et al. (2009b) observed a delay in the anterior N400 peak latency for around 150 ms in the group of intermediate bilinguals, and for around 50-100 ms in highly proficient bilingual speakers. Such results point to a potentially different functional role of the frontal N400 (FN400) compared to the centro-parietal N400. As previously pointed in section 1.4.3.1, while the N400 response is often associated with lexico-semantic mechanisms, the FN400 response is argued to reflect the old/new effect, therefore indexing familiarity-based recognition processes (Curran 2000; Rugg and Curran 2007; Bridger et al. 2012; Leynes et al. 2017). It might therefore be hypothesized that lexico-semantic mechanisms indexed by the centro-parietal N400 are more sensitive to L2 proficiency level than familiarity-related effects reflected in the frontal N400.

Overall, the above-discussed experiments point to the sensitivity of the N400 magnitude and peak latency to the age of L2 acquisition as well as L2 proficiency level and language dominance. Importantly, a delay in the N400 peak latency in response to L2 that has been observed in previous studies seem to provide support for the temporal delay assumption (BIA+ model; Dijkstra and van Heuven 2002; see section 3.2.2.2), which postulates a less automatic activation of semantic information in the non-dominant tongue. An attenuated N400 response to the non-native relative to the native tongue might further point to a decreased interconnectivity between L2 lexical items within the semantic network (Midgley et al. 2009b). Taking into account the N400 sensitivity to the age of L2 acquisition as well as L2 dominance and proficiency level, all of these variables were controlled for in the two experiments reported in this dissertation (Chapter 4 and Chapter 5), which involved late proficient late unbalanced (L1-dominant) bilingual speakers.

3.3.2. The LPC and semantic integration in bilingualism

Only little attention has thus far been devoted to examining the late positive complex (LPC) in bilingual visual word processing. In one of the few studies, Guo et al. (2012) tested proficient Chinese-English bilinguals, who performed a translation recognition task. The results revealed a larger anterior LPC response to semantically related translation distractors relative to unrelated controls. A more pronounced posterior LPC effect was, however, observed in response to unrelated distractors. Such differences in scalp topography were interpreted by the authors as reflecting either different neural generators involved in the processing of critical relative to control items or two overlapping effects, one with an anterior and the other one with a posterior scalp distribution. In another study, Hoshino and Thierry (2012) tested Spanish-English bilinguals and English monolinguals in an English semantic relatedness task.⁷ The authors found a more pronounced LPC response to related compared to unrelated conditions, which was maximal over the centroparietal region, and did not vary as a function of language group.

Modulations in the LPC scalp distribution by L2 proficiency level were investigated by Van Der Meij et al. (2011), who examined electrophysiological correlates of language switching. To this end, the researchers tested high- and low-proficient Spanish-English bilinguals, who read English (L2) sentences, half of which contained a Spanish (L1) adjective in the middle of a sentence. Within the LPC time window, switching conditions evoked a robust LPC response with a broad fronto-posterior distribution, which was, however, maximal over the frontal sites for the low-proficiency group, and over the posterior sites for the high-proficiency group. This might indicate that the LPC scalp distribution is modulated by L2 proficiency level, as bilingual speakers at different L2 proficiency levels might differ in how they update and integrate semantic information. Yet, the authors did not refer to differences in the anterior and posterior LPC that are described in monolingual research on meaning processing, and did not account for how L2 proficiency level might account for such differences in the LPC scalp distribution. Namely, as previously discussed in section 1.4.3.2, while the anterior LPC is postulated to reflect additional working memory processes, the posterior LPC response is claimed to be sensitive to language violations (van Herten et al. 2005; De Grauwe et al. 2010; Brouwer et

⁷ A more detailed description of the study by Hoshino and Thierry (2012) is provided in section 3.3.1.2.

al. 2012). Further research is therefore needed in order to disentangle specific brain mechanisms behind the anterior as compared to posterior LPC effect in bilingual language processing, as well as to provide more information regarding factors that the LPC response in bilingual lexico-semantic processing might be sensitive to.

Importantly, it needs to be noted that previous ERP studies into bilingual lexico-semantic processing have mostly focused on the processing of either single words or meaningful compared to meaningless sentences, and have not been devoted to examining the role of semantic complexity in bilingual language processing. Additionally, even though previous research has focused on a number of different aspects of bilingual language processing, including the comprehension of cognates, interlingual homographs, and translation equivalents, only little attention has been devoted to investigating how bilingual speakers process nonliteral language in their non-native tongue. Importantly, studies into bilingual figurative language comprehension could provide valuable insights into mechanisms engaged in bilingual lexico-semantic processing at different levels of semantic complexity, as semantically simple meanings (i.e., literal) can be contrasted with semantically complex meanings (i.e., metaphoric). The two experiments reported in the present dissertation (see Chapter 4 and Chapter 5) were therefore aimed to examine bilingual language comprehension at different levels of semantic complexity by means of investigating bilingual metaphoric meaning processing. The following section discusses previous research conducted in the field of bilingual nonliteral language comprehension.

3.4. Bilingual figurative language comprehension

Initial research into bilingual figurative language comprehension was mostly conducted for didactic purposes, with a view to investigating how L2 learners perceive, learn, and use humor, idioms, and metaphors (Danesi 1992; Cooper 1999; Schmitz 2002). Conclusions from such research point to a decreased sensitivity of L2 learners to nonliteral meanings presented in their non-native language (Danesi 1992). This effect is postulated to stem from the fact that figurative language is often at least partially culturally-dependent, and since L2 learners might lack L2 culture awareness, they are likely to interpret each word separately when exposed to a figurative expression (Littlemore and Low 2006: 6).

Bilinguals are therefore assumed to be less skilled in integrating meaning with the provided context, and they are claimed to pay attention to contextual cues determined by their native tongue and their native culture, which might consequently lead them to different, often incorrect, interpretations of nonliteral utterances. Such a decreased sensitivity to L2 figurative meanings is explained within the Literal Salience Model (Cieślicka 2006). The model is partially based on the monolingual Graded Salience Hypothesis (Giora 2002; see section 2.3.4), and claims that in the non-native tongue, a literal meaning is always more salient compared to nonliteral meanings, and thus in L2, a literal interpretation is preferentially accessed before a figurative interpretation becomes available.

Additionally, it has been claimed that the ability to arrive at a correct figurative interpretation might be positively correlated with L2 proficiency level (Matlock and Heredia 2002; Cieślicka 2006; Heredia and Cieślicka 2014; Zhao et al. 2014). Namely, low-proficient L2 learners are suggested to analyze a figurative utterance presented in L2 by means of translating it literally into their L1, due to the fact that they use their L1 semantic representations in order to comprehend L2 meaning (Kecskes 2000). However, with increase in L2 proficiency level, L2 learners are hypothesized to become capable of interpreting figurative language as effectively as native language users (Matlock and Heredia 2002; Cieślicka 2006; Heredia and Cieślicka 2014; Zhao et al. 2014).

Thus far, most studies into bilingual figurative language comprehension have focused on how bilingual speakers comprehend idiomatic, ironic, and metaphoric utterances. Though the comprehension of idioms, irony, and metaphors requires the process of meaning retrieval, integration, and construction, the intensity of these processes varies according to the type of a figurative utterance. Namely, as idioms are usually lexicalized, research into idiom processing mostly taps into lexical access from the semantic memory network. On the other hand, the comprehension of ironic utterances is postulated to require the activation of the speaker-intended meaning, which is different from or opposite to the literal interpretation of the presented utterance (Gibbs and Colston 2012: 129). Consequently, studies on irony comprehension provide insights into meaning integration mechanisms. Finally, the comprehension of metaphoric utterances might differ depending on their types. Namely, while conventional metaphoric meanings are lexicalized and their comprehension requires meaning retrieval mechanisms, novel metaphoric meanings need to be created, and hence their comprehension provides insights into meaning construction processes. Importantly, both novel and conventional metaphors might involve a different

degree of meaning integration processes, depending on whether they are embedded in highly or low constraining contexts. In order to discuss how bilingual speakers retrieve, integrate, and construct nonliteral meanings, the following subsection presents bilingual research that has so far been conducted on idiomatic, ironic, and metaphoric meaning comprehension.

3.4.1. Behavioral and electrophysiological research into bilingual nonliteral meaning comprehension

Behavioral research into bilingual nonliteral meaning comprehension has repeatedly pointed to increased cognitive demands when processing non-native compared to native figurative utterances. This assumption has been vastly supported in studies on bilingual idiom comprehension. As the correct interpretation of idioms requires them to be lexicalized, experiments into idiom comprehension mostly tap into meaning retrieval mechanisms. In one of the studies, Cieślicka (2010) tested Polish-English bilingual speakers, who read the beginning of English (L2) idioms presented on the screen, and then finished the expression by saying the final words out loud. On critical trials, participants were already provided with the target word, which they were instructed to read. The results revealed shorter naming times for critical words that were semantically and phonologically related to the literal interpretation of a presented idiomatic expression. Such results suggest an initial activation of literal interpretation when comprehending L2 idioms.

Similar results were found by Cieślicka and Heredia (2011), who also tested Polish-English bilinguals in bilingual idiom comprehension, and employed a divided visual field paradigm with a go/no-go lexical decision task. In their study, idiomatic expressions were provided in either non-constraining ambiguous (biased towards either a literal or an idiomatic interpretation) or constraining unambiguous (biased towards an idiomatic interpretation) sentence contexts, and were followed by a lateral presentation of critical words, which were either related to a literal or a figurative meaning of an idiom. Additionally, the researchers manipulated the duration of an inter-stimulus interval (ISI) between sentence offset and the onset of a critical word, which lasted 0 ms (Experiment 1), 300 ms (Experiment 2), and 800 ms (Experiment 3). The results showed a more robust

priming effect in the right visual field (left hemisphere) for literal than figurative meanings in the non-native language at both 0 ms and 800 ms ISI, thus pointing to a special status of a literal meaning in non-native idiom comprehension, which seems to be in line with the Literal Salience Model (Cieślicka 2006). Importantly, the fact that the results were observed in the right visual field (left hemisphere) is in accordance with neuroimaging evidence suggesting an increased activity in the left prefrontal cortex when processing a non-native language (Indefrey 2006; Abutalebi 2008). The lack of significant priming effect at 300 ms ISI was interpreted by Cieślicka and Heredia (2011) as consistent with previous hemifield studies showing no priming in either visual field at a moderate ISI (e.g., Coney and Evans 2000; Mashal et al. 2008). As for the role of context in bilingual idiom comprehension, the researchers found the left visual field (right hemisphere) advantage for contextually compatible figurative meanings at 0 ms ISI in the case of L1 idiomatic utterances, and at 300 ms ISI in the case of L2 idioms, thus pointing to the right hemisphere sensitivity to the contextual bias in both languages.

In another study, Cieślicka et al. (2014) employed an eye-tracking method in order to examine whether the degree of literal and figurative meaning activation in bilingual idiom comprehension is modulated by language dominance. To this end, the researchers tested Spanish-English bilingual speakers, who were dominant either in Spanish or English. Participants were presented with ambiguous (biased towards either literal or idiomatic interpretation) English idioms in either neutral contexts or rich supportive contexts biasing towards a literal or idiomatic interpretation, and were instructed to read the sentences and answer yes/no comprehension questions. The results revealed that both context and language dominance affected idiom comprehension. Namely, Cieślicka and colleagues (2014) found longer total reading times, more fixations, and more regressions in response to idioms presented in the neutral context relative to the supportive context in both groups of participants. Importantly, total reading times for idioms were shorter in English-dominant bilingual speakers, which therefore indicates that bilingual idiom comprehension might be modulated by language dominance.

More recently, Cieślicka et al. (2017) examined task effects in bilingual idiom comprehension. To this end, the researchers tested Spanish-English bilingual speakers, who differed in their dominance in English or Spanish, and were either Spanish-dominant or English-dominant bilinguals. In Experiment 1, participants performed an implicit lexical decision task, in which they made their decisions as fast as possible by selecting an

answer that first came to their mind. In Experiment 2, in contrast, participants performed an explicit semantic judgment task, which involved a recollection of the previously encountered items. Experimental materials included English idioms that could be interpreted as either literal or figurative utterances (e.g., *to kick the bucket*), and which were embedded in either literal, figurative, or control contexts. The results from Experiment 1 showed that targets were responded to faster when they were preceded by both literal and figurative contexts. This effect was, however, absent in control contexts, thus indicating that an idiomatic meaning of a critical word was activated even if the utterance was used literally. Similarly, a literal meaning of a target item might have been activated when the expression was used figuratively. Additionally, Experiment 1 showed that English idioms were processed faster by English-dominant relative to Spanish-dominant bilinguals, thus indicating that language dominance might modulate the speed of idiom processing. The results of Experiment 2 further revealed that figurative targets elicited shorter response times when preceded by the figurative context, and literal targets were responded to faster when preceded by the literal context. Taken together, the results obtained from the two experiments suggest that task specificity modulates bilingual idiom comprehension, as the difference between English-dominant as compared to Spanish-dominant bilinguals was mostly observed in a lexical decision task (Experiment 1), whereas the results obtained from an explicit semantic judgment task (Experiment 2) emphasized the role of context in idiom processing. To conclude, the four above-discussed experiments into bilingual idiom comprehension have shown that L2 idiomatic expressions might not necessarily be fully lexicalized due to the fact that idioms are usually culture-dependent and are not shared across languages, as a result of which their comprehension may involve serial processing, with a literal meaning accessed prior to an idiomatic interpretation.

In addition to studying bilingual idiom comprehension, previous research has also looked into irony processing in the context of bilingualism. Since irony comprehension is highly modulated by contextual cues, experiments into how bilingual speakers comprehend ironic utterances provide insights into meaning integration mechanisms in bilingual nonliteral language comprehension. In one of the studies on bilingual irony comprehension, Bromberek-Dyzman et al. (2010) tested proficient Polish-English bilingual speakers, who read mini-stories ending with either an ironic or a literal commentary, and performed a self-paced reading task. Behavioral results showed longer reading times for ironic than literal conditions in both languages, with no between-language differences,

therefore demonstrating that irony comprehension was more cognitively taxing than literal meaning comprehension in both the native and non-native tongue. Such results might indicate that Theory of Mind mechanisms, which play a crucial role in irony comprehension, might not be modulated by language nativeness, and might be purely cognitive-based.

More recently, irony comprehension in the bilingual context was examined by Bromberek-Dyzman and Rataj (2016). The researchers instructed highly proficient Polish (L1) – English (L2) bilingual speakers to read mini-stories, and to decide whether the final comment had a critical (ironic) or praising (literal) meaning. Importantly, the study employed the limited response window procedure, in which participants had to make their decisions within 1000 ms after critical word onset. As a result of using such a procedure, participants were encouraged to focus on response speed rather than accuracy, which minimized the role of potential strategies when responding. The results revealed lower accuracy rates along with longer response times for ironic than literal comments in both the native and non-native language, thus indicating increased processing demands when comprehending ironic meanings in both languages. Since no between-language differences were found in response to ironic meanings, it seems that irony comprehension might be similarly taxing in both languages when participants are highly proficient in their L2. The two experiments into bilingual irony comprehension suggest that meaning integration mechanisms might be less sensitive to language nativeness, as bilingual speakers might be similarly competent in both native and non-native ironic meanings comprehension. Such skills might, however, be modulated by L2 proficiency levels, since highly proficient bilingual speakers were tested in the two experiments. Future research is therefore needed in order to show whether the same patterns of results could be obtained when testing less proficient bilinguals.

Finally, previous studies into bilingual figurative language comprehension have also focused on how bilingual speakers interpret metaphoric meanings. Studies into metaphor comprehension provide information into how bilingual speakers retrieve meanings (as in the case of conventionalized metaphors), construct meanings (as in the case of novel metaphors) as well as integrate a metaphoric utterance with the provided context. In one of the studies on bilingual metaphor comprehension, Mashal et al. (2015) tested native speakers of Hebrew and balanced English-Hebrew bilingual speakers using the divided visual field paradigm and a semantic decision task for Hebrew conventional metaphoric

(e.g., *heated debate*), literal (*loyal friend*), and unrelated word dyads (*angle laundry*). Reaction time results revealed a left hemisphere advantage for conventional metaphor comprehension in the native tongue, but a right hemisphere advantage for conventional metaphor interpretation in the non-native language. Importantly, no between-language differences in hemispheric involvement were observed for literal word pairs, which were responded to faster than conventional metaphors in both languages. The obtained results might be in line with the Graded Salience Hypothesis (Giora 2002; see section 2.3.4), which postulates a right hemisphere preference for less salient meanings. Therefore, it seems that conventional metaphors might be salient only in the native tongue, and might thus be preferentially processed by the left hemisphere. In the non-native language, on the other hand, conventional metaphors seem to be less conventionalized, and might therefore be processed somewhat similarly to novel metaphors.

In another study on bilingual metaphor comprehension, Vaid et al. (2015) employed the metaphor interference paradigm, in which participants decided whether a presented sentence (metaphoric, literal or anomalous) is literally true or false. A metaphor interference effect (MIE) is regarded as a measure of the automatic activation of a metaphoric meaning, as participants are assumed to need more time to reject a metaphoric sentence (e.g., *memory is a warehouse*) than an anomalous utterance (e.g., *an insult is a warehouse*) as literally false, thus indicating that metaphor processing is automatic (Pierce et al. 2010). In Experiment 1, Vaid et al. (2015) tested native speakers of English and two groups of English-Spanish bilingual speakers varying in language dominance: English-dominant bilinguals and balanced bilinguals. All of the participants were presented with sentences in English. The results showed a metaphor interference effect, with longer RTs for metaphoric than anomalous sentences in all groups of participants. Such results indicate the automatic activation of a figurative meaning in both monolingual and bilingual speakers, irrespectively of their language dominance. In Experiment 2, Vaid et al. (2015) tested unbalanced as well as balanced English-Spanish bilinguals, who were exposed to both English and Spanish stimuli. This time, the results revealed a metaphor interference effect only in the group of balanced bilinguals, therefore additionally pointing to the role of language dominance in figurative meaning activation. It thus seems that in the non-native language, figurative meanings might be automatically activated in balanced, but not necessarily in unbalanced bilingual speakers.

The role of language dominance in metaphor processing was also examined by Heredia and Cieřlicka (2016), who employed an eye-tracking method, and tested English-dominant, Spanish-dominant, and balanced English-Spanish bilingual speakers. Participants were provided with English passages biasing towards either a literal (e.g., a bakery shop) or metaphoric (e.g., a weak and ineffectual person) meaning of a critical metaphoric referential description (e.g., *creampuff*), and responded to yes/no comprehension questions. The results revealed that metaphoric meaning interpretation was modulated by language dominance, as longer reading times were observed for figurative than literal interpretations, yet only in Spanish-dominant bilinguals. This points to a preference for literal meaning activation when processing metaphoric language in the non-dominant tongue. English-dominant as well as balanced bilingual speakers, in contrast, might have accessed both literal and figurative interpretations with a similar efficiency. Overall, the experiments reported above point to the significant role of language dominance in metaphor comprehension, and seem to suggest a preferential activation of a literal meaning in unbalanced bilingual speakers. Furthermore, as shown by Mashal et al. (2015), bilingual speakers might be less sensitive to the levels of meaning conventionality, and may consequently perceive both novel and conventional metaphors as similarly taxing when presented in their non-native tongue.

Importantly, previous studies into bilingual metaphor comprehension have not been devoted to examining specific mechanisms engaged when comprehending novel and conventional metaphors in the native and non-native tongue. For instance, they have not investigated simile comprehension, and thus it remains under-investigated whether comparison mechanisms modulate novel metaphoric meaning comprehension in the bilingual context. Furthermore, only little research has focused on novel figurative meaning comprehension. In one of the few studies, Onysko (2016) investigated bilingual creativity, and tested English monolinguals and Maori-English bilinguals. Participants performed a meaning interpretation task for novel English compounds (e.g., *bucket philosopher*). The results showed similar patterns of diversity of figurative associations in the monolingual and bilingual group. Yet, the two groups differed in the associative strategy of meaning interpretations, as bilingual speakers tended to associate the existing idiomatic expressions with homophones related to some constituents of the presented compound. Such results point to a generally increased associative flexibility of bilinguals as compared to

monolingual speakers, which might result from an overall more robust activation in lexico-semantic networks of bilingual speakers (Onysko 2016).

Electrophysiological research has been scarcely devoted to examining figurative meaning comprehension in the bilingual context. Thus far, only two ERP experiments have investigated bilingual nonliteral meaning comprehension. Paulmann et al. (2015) tested English native speakers and late Arabic-English bilinguals in a rapid serial visual presentation with randomly presented comprehension questions. Participants were exposed to English verb-preposition strings (e.g., *to run over*), which were presented in two different contexts, biasing towards either a literal or a figurative interpretation. Surprisingly, in both groups of participants, a more robust N400 response was observed to literal compared to figurative interpretations, thus pointing to more cognitively taxing mechanisms engaged in literal than figurative meaning comprehension. Nonetheless, the observed results might have been affected by the experimental materials, which had not been controlled for in terms of their familiarity and frequency of co-occurrence. Since these factors have previously been found to influence language processing, with familiar and frequent lexical items accessed faster than unfamiliar and less frequent words (Connine et al. 1990; Kutas et al. 2006), they are also highly likely to influence nonliteral meaning comprehension in bilingualism.

In another ERP experiment, Chen et al. (2013) tested proficient Chinese-English bilinguals, who were presented with metaphoric (e.g., *Zeal is fire*) and literal sentences (*Jim is a teacher*) in both Chinese and English, and decided whether the sentences had a metaphorical meaning. The results revealed no differences within the N400 time window in response to metaphorical and literal sentences presented in Chinese (L1); yet, the authors observed a more pronounced N400 response to English (L2) metaphors as compared to L2 literal meanings. The obtained results might have, however, resulted from the fact that the metaphors had not been controlled for in terms of their level of conventionality, as the researchers had not examined whether L1 and L2 metaphoric sentences were novel or conventional.

As the above discussion shows, research into bilingual metaphoric language comprehension is vastly limited. Behavioral studies have suggested that idiomatic and metaphoric language comprehension lags behind in L2, especially when bilingual speakers are non-dominant in their non-native language. Additionally, a decreased sensitivity to met-

aphor conventionality has been observed, which might suggest that conventional metaphors are less salient in the non-native relative to the native tongue (Mashal et al. 2015). Importantly, experiments conducted thus far have not been devoted to investigating the comprehension of novel and conventional metaphors in L1 and L2, and have not examined the role of comparison mechanisms in bilingual metaphor processing. Future research is therefore needed in order to provide more information regarding these issues, which can further contribute to elucidating the role of semantic complexity in bilingual language processing.

3.5. Concluding remarks

Research into bilingual lexico-semantic processing conducted thus far has provided strong support in favor of the language non-selective account, pointing to a simultaneous co-activation of both the native and non-native tongue during bilingual language processing, as reflected in behavioral and electrophysiological measures.⁸ Nonetheless, as pointed in the sections above, only little attention has thus far been devoted to investigating bilingual metaphoric language comprehension, while studies on metaphor comprehension in bilingualism could show how semantic complexity modulates bilingual lexico-semantic processing, as well as how bilingual speakers retrieve conventionalized and construct novel meanings in their native and non-native language. With a view to filling the gap in existing research, experiments described in this thesis were devoted to examining how late proficient unbalanced (L1-dominant) bilingual speakers compute novel and conventional metaphors. The main objective of Experiment 1 (Chapter 4) was to investigate whether bilinguals are sensitive to the levels of metaphor conventionality, as reflected in ERP and behavioral patterns. Additionally, Experiment 2 (Chapter 5) aimed to examine whether comparison processes initiated when comprehending similes might ease novel meaning comprehension in the bilingual context, as reflected in behavioral data. The following chapters report these two experiments (Chapter 4 and Chapter 5, respectively), and discuss their results with reference to the previously reviewed theoretical frameworks (Chapter 6).

⁸ ERP studies into the N400 and LPC responses in bilingual language processing are summarized in Appendix B (p. 247).

Chapter 4: Experiment 1: Novel and conventional metaphor comprehension in bilingualism

4.1. Introductory remarks

Research into bilingual figurative language comprehension conducted so far has been rarely devoted to investigating whether metaphor conventionality modulates metaphoric meaning processing in both the native (L1) and non-native tongue (L2). The main objective of the present experiment was therefore to examine how the bilingual brain computes novel and conventional metaphors in L1 and L2, as reflected in behavioral (reaction times and accuracy rates) as well as electrophysiological (ERP) measures. To this end, late proficient unbalanced Polish (L1) – English (L2) bilingual speakers performed a two-choice semantic decision task in response to novel metaphoric, conventional metaphoric, literal, and anomalous verb-noun word dyads presented in their native and non-native language.⁹

4.2. Aims, research questions and hypotheses

Thus far, the role of metaphor conventionality in metaphoric meaning comprehension has been examined mostly in the monolingual context. In favor of the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), previous monolingual research has shown that novel metaphors are more difficult to comprehend compared to conventional metaphors, as reflected in longer response times as well as the graded N400 response, with larger N400 amplitudes for novel than conventional metaphoric utterances (e.g.,

⁹ The experiment reported in this chapter has appeared in Jankowiak et al. (2017).

Bowdle and Gentner 2005; Arzouan et al. 2007; Lai et al. 2009; Goldstein et al. 2012; see section 2.6.2). Yet, it remains unclear whether bilingual speakers are similarly sensitive to the levels of metaphor conventionality when metaphoric meanings are presented in their non-native tongue. Consequently, the present experiment was aimed at examining how the bilingual brain computes novel and conventional metaphoric word pairs presented in the native and non-native language.

In addition, by examining the comprehension of metaphoric as well as literal word dyads in L1 and L2, the study provides insights into how bilingual speakers process meanings of different levels of semantic complexity, with literal utterances representing semantically simple meanings and metaphoric word pairs exemplifying meanings that are semantically complex. In line with the temporal delay assumption postulated within the Bilingual Interactive Activation Plus Model (BIA+; Dijkstra and van Heuven 2002; see section 3.2.2.2), the activation of semantic information is delayed in the non-dominant tongue. This assumption has been supported in previous electrophysiological research showing longer response times as well as a delay in the N400 peak latency in response to the non-native relative to the native tongue (Weber-Fox and Neville 1996; Dijkstra et al. 1998a; Dijkstra et al. 1999; de Groot et al. 2002; Phillips et al. 2004; Moreno and Kutas 2005; Braunstein et al. 2012). Previous studies have, however, not been devoted to examining whether this temporal delay could be affected by the semantic complexity of experimental materials. The next objective of the present experiment was therefore to show whether and how mechanisms engaged in bilingual lexico-semantic access and meaning integration might be modulated by semantic complexity in L1 and L2.

Two independent variables were tested in the present experiment: language (Polish: native and dominant language; English: non-native and non-dominant language) and utterance type (novel metaphoric, conventional metaphoric, literal, and anomalous verb-noun word dyads). Dependent variables included event-related potentials, reaction times, and accuracy rates. Two main research questions were addressed in the study:

- 1) Are bilingual speakers sensitive to the levels of metaphor conventionality in both their native and non-native tongue?
- 2) Are cognitive mechanisms engaged in bilingual lexico-semantic processing less automatic and delayed in the non-native and non-dominant relative to the native and dominant tongue?

Six research hypotheses have been formulated in order to address the two aforementioned research questions:

Hypothesis 1: *In the native language, novel metaphors will elicit longer reaction times and lower accuracy rates compared to both conventional metaphors and literal utterances.*

Previous monolingual research into the role of metaphor conventionality has repeatedly shown that metaphor comprehension is modulated by the degree of meaning conventionality, as novel (unfamiliar) metaphors have been observed to evoke longer reaction times compared to conventional (familiar) metaphoric utterances (e.g., Blank 1988; Gentner and Wolff 1997; Bowdle and Gentner 2005). Such results are in line with the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), which claims that while novel metaphor processing is serial, with a literal interpretation accessed prior to a target metaphoric meaning, conventional metaphors, due to their high frequency of use, are accessed directly, without a preliminary activation of a literal meaning. Consequently, a difference in response speed to novel compared to conventional metaphors is interpreted as resulting from the fact that novel metaphors require the processes of meaning construction, whereas the comprehension of conventional metaphoric utterances involves only meaning retrieval processes, and thus conventionalized metaphors are easier and faster to comprehend (Bowdle and Gentner 2005). Therefore, faster RTs along with higher accuracy rates elicited in response to conventional metaphoric compared to novel metaphoric word pairs will corroborate the previously observed results.

Furthermore, although the Standard Pragmatic View (Grice 1975; see section 2.2.2) claims that literal meanings enjoy the temporal priority over metaphoric utterances, the view does not account for whether such a priority could be modulated by metaphor conventionality. Namely, since conventional metaphors involve sense retrieval mechanisms, the comprehension of conventional and literal meanings is likely to be comparably automatic. Similar reaction times and accuracy rates for conventional metaphoric and literal utterances have been observed in previous research (e.g., Arzouan et al. 2007; Tang et al. 2017a), and thus it is expected that the results of the present experiment will show a similar pattern of results.

Hypothesis 2: *In the non-native language, novel and conventional metaphors will elicit similar reaction times and accuracy rates, both of which will evoke longer response times and lower accuracy ratings than literal word pairs.*

Studies on bilingual metaphoric meaning comprehension have so far been limited, and they have rarely been aimed at investigating whether metaphor conventionality modulates metaphor comprehension in the non-native tongue. Yet, due to the fact that metaphor conventionality is postulated to reflect how frequently a metaphor vehicle is used in its nonliteral sense (Bowdle and Gentner 2005), it might be assumed that the vehicles of conventional metaphors have been less frequently referred to their metaphoric meaning in L2 than L1. Consequently, L2 conventional metaphors might actually be comprehended similarly to novel metaphors, as their metaphoric meaning frequency is lower for L2 users compared to native speakers. In line with such an assumption, a behavioral study conducted by Mashal et al. (2015) showed that conventional metaphors are less salient in the non-native compared to the native tongue, as they might not be fully conventionalized in L2. It is therefore hypothesized that L2 novel and conventional metaphors will elicit similar reaction times and accuracy rates.

In addition, since L2 conventional metaphors are likely to be of lower frequency compared to L2 literal utterances, it is expected that both novel and conventional metaphors presented in the non-native tongue will elicit longer response times along with lower accuracy rates compared to literal word pairs. Such results would provide support for the Literal Salience Model (Cieślicka 2006), which claims that in bilingual figurative language processing, literal meaning is more salient relative to the nonliteral interpretation, and is thus preferentially accessed before the figurative interpretation becomes available.

Hypothesis 3: *Overall, all word pair types presented in the non-native language will elicit longer reaction times compared to stimuli presented in the native tongue.*

According to the temporal delay assumption discussed within the Bilingual Interactive Activation Plus Model (Dijkstra and van Heuven 2002; see section 3.2.2.2), the activation of L2 semantic representations is delayed relative to L1 due to a lower subjective frequency of non-native items. This assumption has been supported in previous behavioral research on bilingual language processing, showing longer response times for L2 compared to L1 stimuli (e.g., Dijkstra et al. 1998a; Dijkstra et al. 1999; de Groot et al. 2002;

Moreno et al. 2008). Overall longer reaction times for all types of the experimental utterances presented in L2 than L1 will therefore corroborate previous results.

Hypothesis 4: *Within the N400 time window, a linear effect across the utterance types will be observed in both the native and non-native language.*

Previous ERP research into monolingual metaphor processing has shown a graded N400 effect of utterance type, with the most pronounced N400 response to anomalous conditions, followed by novel metaphors, conventional metaphors, and finally literal utterances (Arzouan et al. 2007; Lai et al. 2009; Tang et al. 2017a). Larger N400 amplitudes for novel than conventional metaphors seem to suggest more resource intensive lexico-semantic processes engaged in meaning construction (i.e., novel metaphor comprehension) compared to meaning retrieval (i.e., conventional metaphor comprehension). An increased N400 response to conventional metaphoric than literal utterances might additionally indicate the increased difficulty of cross-domain mappings in conventional metaphoric compared to literal meaning processing.

In the bilingual context, on the other hand, novel and conventional metaphor comprehension has thus far been scarcely examined, and thus it remains under-investigated whether the graded N400 effect observed in the native tongue might also be present in the non-native language. If the same pattern of ERP results is observed in both languages, this will indicate that lexico-semantic processes reflected in the N400 response are independent of language. Alternatively, if conventional metaphors are less salient in the non-native relative to the native tongue, as suggested in previous research (Mashal et al. 2015), this might be reflected in the converged N400 waveforms for novel and conventional metaphoric word pairs in L2.

Hypothesis 5: *In both the native and non-native language, novel metaphoric and anomalous word pairs will elicit attenuated LPC amplitudes compared to conventional metaphoric and literal word dyads.*

In addition to utterance-specific differences observed within the N400 time frame, previous monolingual ERP experiments into metaphor comprehension have revealed an attenuated LPC response to novel metaphors and anomalous utterances relative to conventional metaphoric and literal meanings (Arzouan et al. 2007; De Grauwe et al. 2010; Tang

et al. 2017a; Rataj et al. 2018). Such reduced LPC amplitudes for novel metaphoric meanings might reflect sustained negativity, possibly overlapping with the LPC response, and might indicate the ongoing difficulty of novel metaphor integration or/and access to the nonliteral route (Arzouan et al. 2007). For anomalous utterances, this prolonged negativity is postulated to reflect continued effort in meaning integration or additional meaning re-analyses (De Grauwe et al. 2010). It is thus hypothesized that in the native language, novel metaphoric and anomalous word dyads will evoke reduced LPC amplitudes compared to conventional metaphoric and literal word pairs.

A similar pattern of results observed in the non-native language will suggest that meaning integration processes, as indexed by the LPC response, are language-independent. If, on the other hand, bilingual speakers are less sensitive to the levels of metaphor conventionality, as observed in previous research (Mashal et al. 2015) and already pointed in Hypothesis 2 and Hypothesis 4, this might be reflected in a similar magnitude of the LPC response to both novel and conventional metaphors in L2. Thus, in the non-native language, LPC waveforms in response to novel and conventional metaphors might converge, which will further indicate that conventional metaphors are less salient in L2, and thus their integration is more taxing.

Hypothesis 6: *Overall, all word pair types presented in the non-native language will elicit attenuated and/or delayed N400 and LPC amplitudes compared to the word dyads presented in the native tongue.*

Previous electrophysiological research into bilingual semantic processing has shown an attenuated (Proverbio et al. 2002; Moreno et al. 2008; Midgley et al. 2009b; Newman et al. 2012) or delayed (Weber-Fox and Neville 1996; Phillips et al. 2004; Moreno and Kutas 2005; Braunstein et al. 2012) N400 response to the non-native relative to the native language. A delay in the N400 peak latency to the non-native tongue is in line with the temporal delay assumption proposed in the BIA+ model (Dijkstra and van Heuven 2002; see section 3.2.2.2), which points to a delayed activation of L2 semantic representations in L2 non-dominant bilingual speakers. An attenuated N400 response to the non-native language is further interpreted in accordance with the functional role of the N400 effect, according to which the N400 response indexes information retrieval from long-term memory (Kutas and Federmeier 2000; Kutas and Federmeier 2011; Kotz et al. 2012). Following such an assumption, a reduction in the N400 magnitude indicates a decreased

activity in long-term memory, possibly due to weaker interconnectivity for L2 lexical items within the semantic memory network in non-proficient and/or L2 non-dominant bilingual speakers (Midgley et al. 2009b). Results obtained from the present experiment are expected to corroborate the previously observed results, and thus it is hypothesized that overall, all word pair types will elicit attenuated and/or delayed N400 amplitudes in L2 compared to L1.

Previous ERP studies on bilingual semantic processing have, however, rarely focused on the LPC response to the native compared to non-native stimuli. In the study by Hoshino and Thierry (2012), a more pronounced LPC response to semantically related relative to unrelated items was observed, and its magnitude did not vary as a function of language group (i.e., L2 learners of English vs. English native speakers). Therefore, the LPC amplitudes might be of a similar magnitude for both Polish (L1) and English (L2) utterances. Nonetheless, taking into account the hypothesized delay in the activation of semantic representations in L2, it might be predicted that the process of semantic integration, as indexed by the LPC response, may also be delayed in L2, as reflected in a delayed LPC latency to the non-native relative to the native tongue.

4.3. Method

4.3.1. Participants

The original sample included 28 participants. Four of them had to be excluded from analyses owing to low accuracy rates on literal or anomalous trials (below 70%). Additionally, one participant was removed from analyses as a result of an increased impedance during the EEG recording. This resulted in a final sample of 23 participants (16 women, $M_{age} = 22.83$, $SD = 1.03$), who were MA students of the Faculty of English (Adam Mickiewicz University in Poznań), and who participated in the experiment for course credits. All of the participants were right handed ($M_{right\ hand\ preference} = 89.80\%$, $SD = 12.38$), as indicated by scores obtained from an online Handedness Questionnaire (Cohen 2008) based on the Edinburgh Inventory (Oldfield 1971). Participants were all native speakers of Polish (L1), and were late proficient unbalanced learners of English as their second

language (L2), as they had acquired their L2 after their L1 ($M_{age\ of\ L2\ acquisition} = 9.61$, $SD = 3.19$). They used their native tongue more frequently than their non-native language, and they were thus L1-dominant. Although their L2 proficiency level was not directly tested during the experiment, all participants were students of English Studies, and had passed their Practical English Language Exam, which was equivalent to the Cambridge Proficiency Examination that is administered by Cambridge University. All of the participants had normal or corrected to normal vision, and they did not suffer from any language or neurological disorder (Jankowiak et al. 2017: 5).

4.3.2. Materials

Experimental stimuli used in the present study included 304 Polish (L1) and 304 English (L2) verb-noun word pairs, which were divided into four categories: 76 novel metaphoric (e.g., *to harvest courage*), 76 conventional metaphoric (*to gather courage*), 76 literal (*to experience courage*), and 76 anomalous word dyads (*to move courage*) in each language. Each set shared the same critical noun, 8 of which were Polish-English cognate words and 1 was an interlingual homograph, which, however, constituted only 6% of all critical words (Jankowiak et al. 2017: 5). The criteria under which the critical words were selected are presented in Table 1. The experimental materials used in the study are provided in Appendix C (Polish stimuli; p. 257) and Appendix D (English stimuli; p. 260).

Table 1. Polish and English critical words' characteristics, including their frequency per million, number of syllables, and number of letters (adapted from Jankowiak et al. 2017: 6).

| | Frequency per million | Number of syllables | Number of letters |
|-----------------------|--|---|---|
| Polish critical words | 15-36/million $M = 23.53$, $SD = 5.91$ (as indicated by the National Corpus of Polish Language) | 2-3 syllables $M = 2.49$, $SD = 0.50$ | 4-13 letters $M = 7.29$, $SD = 1.88$ |
| | English critical words | 14-37/million $M = 23.75$, $SD = 6.41$ (as indicated by the Corpus of Contemporary American English) | 2-3 syllables $M = 2.51$, $SD = 0.50$ |

All of the experimental materials used in the present study were pretested by means of administering web-based surveys. While critical words were pretested with a view to evaluating their level of abstractness, word pairs were pretested in four normative studies: cloze probability tests, as well as meaningfulness, familiarity, and metaphoricity rating scales. Instructions used in the surveys are provided in Appendix G (Polish surveys; p. 269) and Appendix H (English surveys; p. 272). Polish materials were evaluated by Polish native speakers, whereas English materials were assessed by English native speakers. Raters whose scores were more than 3 SDs from the mean were excluded from final analyses. Table 2 provides the number of raters included in the analyses together with their demographic data. Raters who took part in these pretests did not participate in the ERP experiment.

Table 2. Demographic information regarding the participants of the five normative studies in each language, including the number of raters, their gender, and mean age (adapted from Jankowiak et al. 2017: 7).

| Survey type | Language | Number of raters included in the analyses | Mean age |
|------------------------------|----------|---|------------------------------|
| Critical words' concreteness | Polish | N = 33 (23 women) | $M_{age} = 22.12, SD = 3.52$ |
| | English | N = 34 (19 women) | $M_{age} = 24.53, SD = 5.43$ |
| Cloze probability | Polish | N = 140 (112 women) | $M_{age} = 21.07, SD = 2.62$ |
| | English | N = 140 (65 women) | $M_{age} = 22.76, SD = 4.86$ |
| Meaningfulness | Polish | N = 137 (99 women) | $M_{age} = 21.52, SD = 3.03$ |
| | English | N = 133 (61 women) | $M_{age} = 22.15, SD = 5.21$ |
| Familiarity | Polish | N = 103 (82 women) | $M_{age} = 21.74, SD = 3.33$ |
| | English | N = 101 (55 women) | $M_{age} = 22.94, SD = 5.46$ |
| Metaphoricity | Polish | N = 101 (83 women) | $M_{age} = 21.27, SD = 2.89$ |
| | English | N = 102 (59 women) | $M_{age} = 22.15, SD = 5.21$ |

4.3.2.1. Normative studies: Critical words

With a view to ensuring that all of the critical nouns were abstract, raters evaluated 76 abstract critical nouns together with 76 concrete filler words on a 7-point rating scale, where 1 represented abstract words and 7 represented concrete words. Paired samples *t*-tests showed that Polish critical words ($M = 3.12$, $SD = 0.58$) were assessed as more abstract than concrete filler words ($M = 5.75$, $SD = 0.75$), $p < .001$. Similarly, English critical words ($M = 2.96$, $SD = 0.91$) were rated as more abstract as compared to concrete filler words ($M = 5.72$, $SD = 0.65$), $p < .001$ (Jankowiak et al. 2017: 6).

4.3.2.2. Normative studies: Word dyads

Cloze probability tests were conducted in order to ensure that all of the critical nouns were not expected due to the preceding context. Raters were presented with a prime word (a verb), and were instructed to provide a critical word (a noun) which first came to their mind, so that the two-word utterance would be meaningful. The final list of stimuli included only those word pairs whose critical words had been elicited less than three times in the cloze probability test (Jankowiak et al. 2017: 6).

For the normative studies with rating scales on stimuli meaningfulness, familiarity, and metaphoricity, analyses of variance (*ANOVAs*) were conducted, whose results are reported below. Significance values for pairwise comparisons were corrected for multiple comparisons using the Bonferroni correction. If Mauchly's tests indicated that the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied. In such cases, the original degrees of freedom are reported with the corrected *p* value. While novel metaphoric, conventional metaphoric, and literal word dyads were pretested with regard to their meaningfulness, familiarity, and metaphoricity, anomalous word pairs were tested only on their level of meaningfulness.

To evaluate the meaningfulness of the word dyads, raters assessed them on a scale from 1 (totally meaningless) to 7 (totally meaningful). The analysis conducted on Polish surveys showed a main effect of utterance type, $F(3, 399) = 1204.32$, $p < .001$, $\epsilon = .666$, $\eta_p^2 = .901$. Pairwise comparisons further revealed that literal word pairs ($M = 5.63$, $SE = 0.06$) were rated as more meaningful than conventional metaphors ($M = 5.44$,

$SE = 0.07$), $p < .001$, conventional metaphors were rated as more meaningful than novel metaphors ($M = 3.70$, $SE = 0.08$), $p < .001$, and novel metaphors were assessed as more meaningful compared to anomalous word pairs ($M = 1.89$, $SE = 0.06$), $p < .001$.

Similarly, the results obtained from English rating scales revealed a main effect of utterance type, $F(3, 387) = 1611.54$, $p < .001$, $\epsilon = .799$, $\eta_p^2 = .926$. Pairwise comparisons further showed that literal utterances ($M = 5.99$, $SE = 0.05$) were rated as more meaningful than conventional metaphors ($M = 5.17$, $SE = 0.06$), $p < .001$, conventional metaphors were rated as more meaningful than novel metaphors ($M = 4.09$, $SE = 0.08$), $p < .001$, and novel metaphors were assessed as more meaningful compared to anomalous utterances ($M = 2.33$, $SE = 0.06$), $p < .001$. Fig. 18 presents meaningfulness ratings for Polish and English experimental materials.

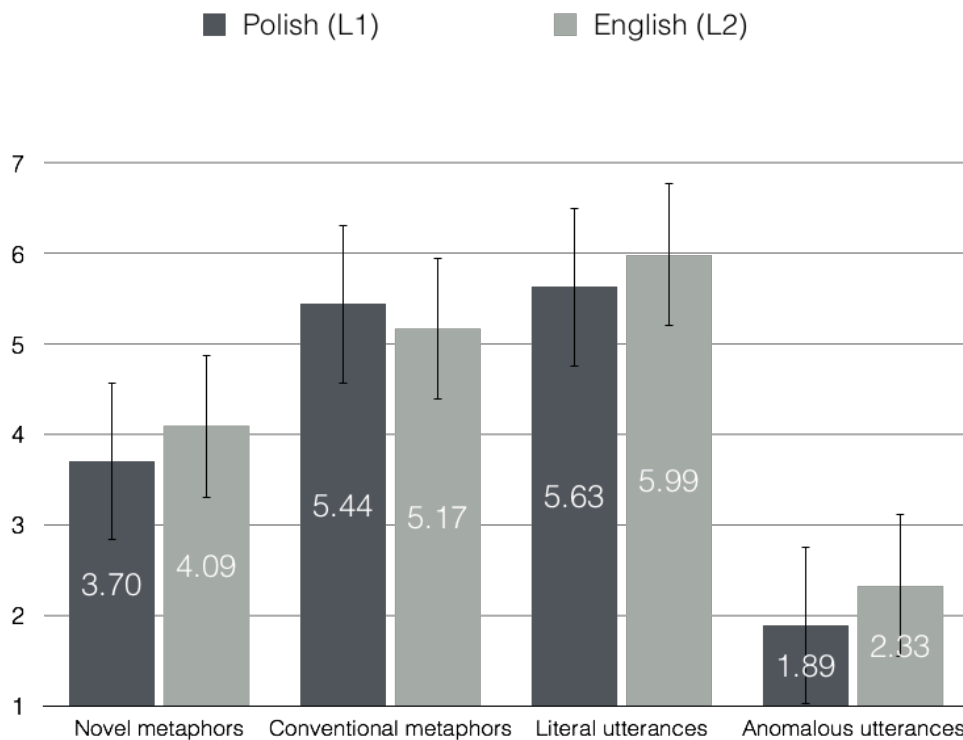


Fig. 18. Meaningfulness ratings for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, literal, and anomalous word pairs (1 = totally meaningless, 7 = totally meaningful).

To assess the familiarity of the stimuli, raters decided how often they encountered the presented word pairs on a scale from 1 (very rarely) to 7 (very frequently). The results obtained from rating scales on Polish stimuli showed a main effect of utterance type, $F(2,$

200) = 684.63, $p < .001$, $\varepsilon = .920$, $\eta_p^2 = .873$. Pairwise comparisons confirmed that novel metaphors ($M = 2.39$, $SE = 0.07$) were less familiar than both literal utterances ($M = 4.10$, $SE = 0.09$), $p < .001$, and conventional metaphors ($M = 4.28$, $SE = 0.08$), $p < .001$. Furthermore, literal word pairs were less familiar than conventional metaphors, $p = .002$.

With respect to the rating scales on English word dyads, a main effect of utterance type was found, $F(2, 296) = 470.97$, $p < .001$, $\varepsilon = .801$, $\eta_p^2 = .827$. Pairwise comparisons showed that English novel metaphors ($M = 2.15$, $SE = 0.07$) were rated as less familiar than conventional metaphors ($M = 2.97$, $SE = 0.08$), $p < .001$, as well as than literal utterances ($M = 3.85$, $SE = 0.09$), $p < .001$. Furthermore, conventional metaphors were less familiar than literal word dyads, $p < .001$. Fig. 19 presents familiarity ratings for Polish and English experimental materials.

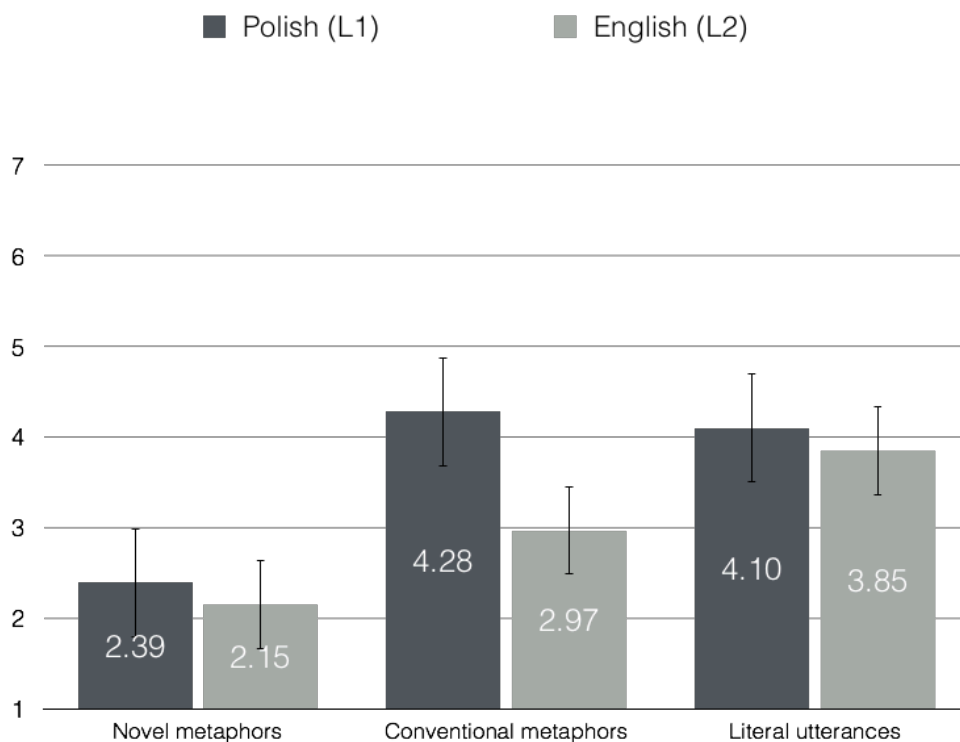


Fig. 19. Familiarity ratings for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, and literal word pairs (1 = very rarely, 7 = very frequently).

In order to evaluate the metaphoricity of the stimuli, raters decided how metaphorical given word dyads were on a scale from 1 (very literal) to 7 (very metaphorical). The

results from rating scales on Polish stimuli showed a main effect of utterance type, $F(2, 196) = 605.55, p < .001, \epsilon = .732, \eta_p^2 = .861$. Pairwise comparisons further showed that novel metaphors ($M = 5.47, SE = 0.06$) were judged as more metaphorical than conventional metaphors ($M = 4.05, SE = 0.08$), $p < .001$, and conventional metaphors were rated as more metaphorical than literal word dyads ($M = 2.65, SE = 0.09$), $p < .001$.

The results obtained from English surveys also revealed a main effect of utterance type, $F(2, 198) = 588.82, p < .001, \epsilon = .738, \eta_p^2 = .856$. Pairwise comparisons confirmed that novel metaphors ($M = 5.00, SE = 0.06$) were rated as more metaphorical than conventional metaphors ($M = 3.98, SE = 0.06$), $p < .001$, and conventional metaphors were rated as more metaphorical than literal utterances ($M = 2.74, SE = 0.07$), $p < .001$. Fig. 20 presents metaphoricity ratings for Polish and English experimental materials.

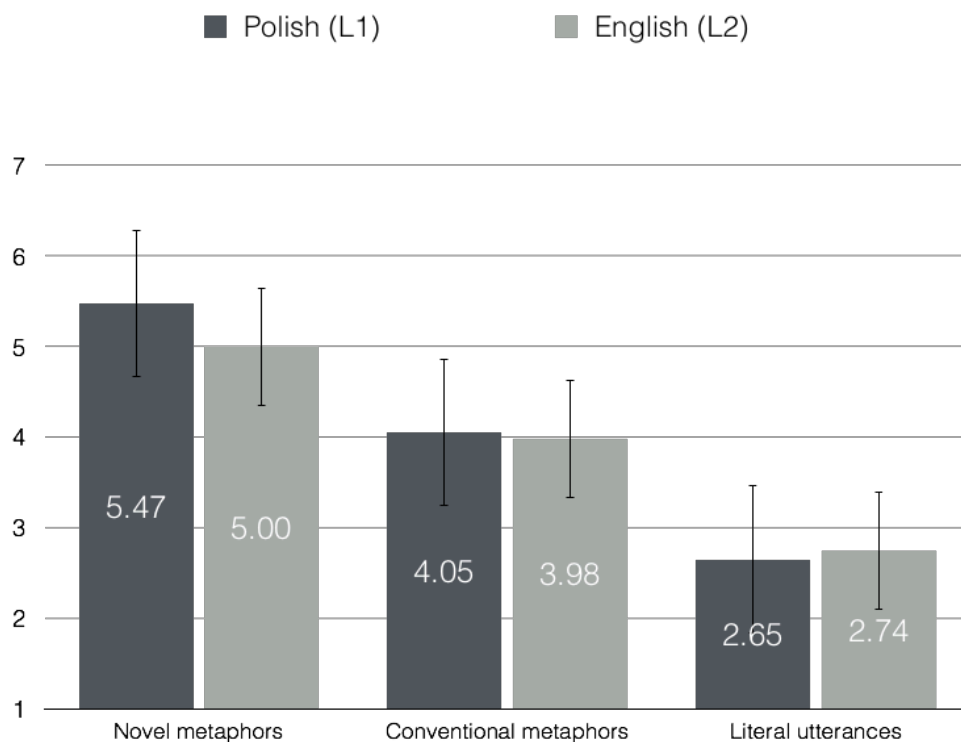


Fig. 20. Metaphoricity ratings for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, and literal word pairs (1 = very literal, 7 = very metaphorical).

The normative studies resulted in the selection of the final 304 Polish and 304 English word pairs that were used in the ERP experiment. The results of the normative

studies for those word dyads that were included in the final list are reported in Table 3. For the EEG experiment, the stimuli were divided into 8 blocks (4 blocks in Polish and 4 blocks in English) in order to avoid the repetition of critical words within one block. Each block consisted of 19 novel metaphoric, 19 conventional metaphoric, 19 literal, and 19 anomalous word pairs. Additionally, 30 filler word dyads were added to each block, which differed from the experimental stimuli in their syntactic structure, and instead of containing verb-noun word pairs, they included adjective-noun or noun-noun word dyads. Furthermore, in order to balance out the number of positive and negative responses to the utterances, all of the filler word dyads were meaningless. Altogether, 106 randomized expressions were used in each block. Each participant completed all 8 blocks, whose presentation was randomized (Jankowiak et al. 2017: 6-8).

Table 3. Results of the normative tests on the stimuli included in the Experiment 1 (adapted from Jankowiak et al. 2017: 8).

| | Meaningfulness (1 – totally meaningless, 7 – totally meaningful) | | Familiarity (1 – very rarely, 7 – very frequently) | | Metaphoricity (1 – very literal, 7 – very metaphorical) | |
|------------------------|--|-------------------------------------|--|-------------------------------------|---|-------------------------------------|
| | Polish | English | Polish | English | Polish | English |
| Novel metaphors | <i>M</i> = 3.70 <i>SE</i> = 0.08 | <i>M</i> = 4.09 <i>SE</i> = 0.08 | <i>M</i> = 2.39 <i>SE</i> = 0.07 | <i>M</i> = 2.15 <i>SE</i> = 0.07 | <i>M</i> = 5.47 <i>SE</i> = 0.06 | <i>M</i> = 5.00 <i>SE</i> = 0.06 |
| Conventional metaphors | <i>M</i> = 5.44 <i>SE</i> = 0.07 | <i>M</i> = 5.17 <i>SE</i> = 0.06 | <i>M</i> = 4.28 <i>SE</i> = 0.08 | <i>M</i> = 2.97 <i>SE</i> = 0.08 | <i>M</i> = 4.05 <i>SE</i> = 0.08 | <i>M</i> = 3.98 <i>SE</i> = 0.06 |
| Literal utterances | <i>M</i> = 5.63 <i>SE</i> = 0.06 | <i>M</i> = 5.99 <i>SE</i> = 0.05 | <i>M</i> = 4.10 <i>SE</i> = 0.09 | <i>M</i> = 3.85 <i>SE</i> = 0.09 | <i>M</i> = 2.65 <i>SE</i> = 0.09 | <i>M</i> = 2.74 <i>SE</i> = 0.07 |
| Anomalous utterances | <i>M</i> = 1.89 <i>SE</i> = 0.06 | <i>M</i> = 2.33 <i>SE</i> = 0.06 | - | - | - | - |

4.3.3. Procedure

The procedures applied in the EEG experiment were approved by the Adam Mickiewicz University Human Research Ethics Committee, and followed ethical guidelines for studies with human participants. An informed consent form along with the information about

the procedures of the experiment was provided to participants before the experiment began. The informed consent form is provided in Appendix K (p. 281).

The experiment was conducted at the Vision and Neuroscience Laboratory, Nano-BioMedical Centre, at Adam Mickiewicz University in Poznań. Participants were seated in a dim and quiet testing cabin, 50 cm from the computer screen. During the experiment, the word pairs were randomly presented one word at a time on a computer screen using black letters, and were centered on a gray background. The stimuli were presented in the following time sequence: the first fixation cross (500 ms), followed by a prime word (700 ms), a blank screen (300 ms), the second fixation cross (600 ms), and finally a critical word (1500 ms). After the critical word appeared, a blank screen was displayed until participants made their decisions. Trials were separated by an intertrial interval (2000 ms), which was presented as a blank screen. The time sequence of stimulus presentation is provided in Fig. 21.

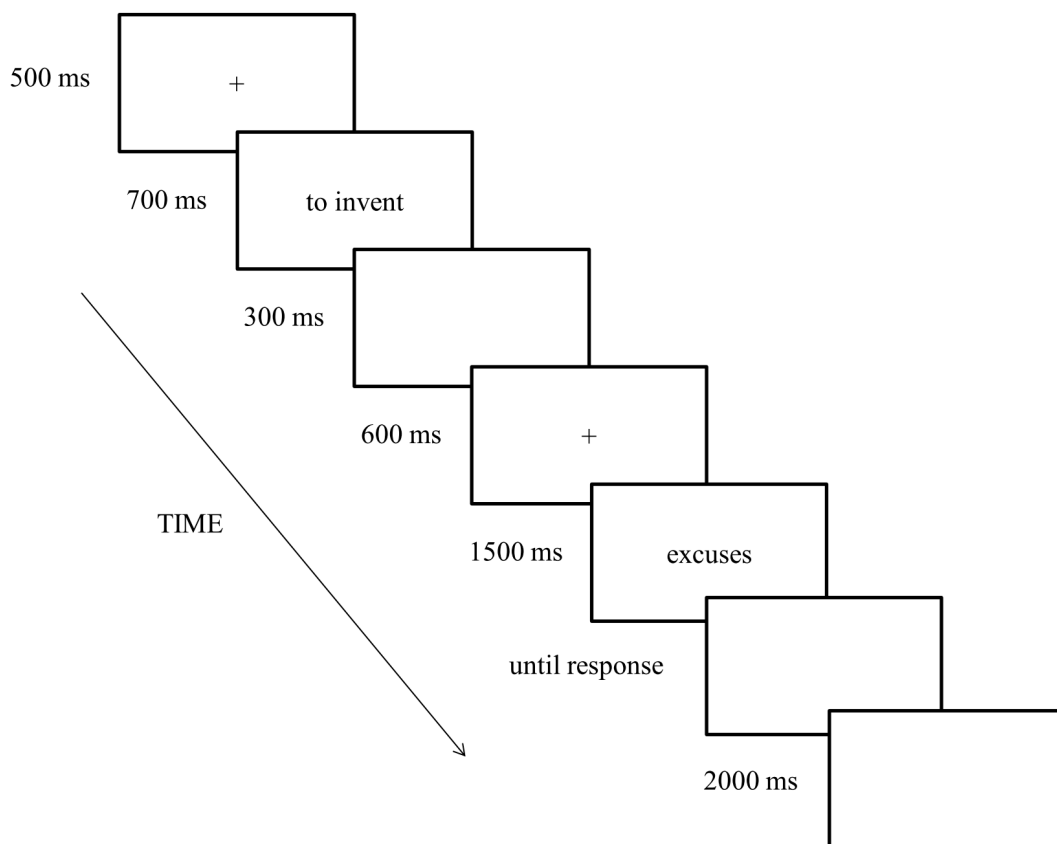


Fig. 21. Time sequence of stimulus presentation (after Jankowiak et al. 2017: 9).

Participants decided if the presented word pair was meaningful or meaningless by pressing a corresponding key, whose designation was counterbalanced. Prior to the eight experimental blocks, participants completed a practice block with 20 word dyads not included in the experimental trials (Jankowiak et al. 2017: 8). Participants who began the experiment with English blocks were instructed in English, and those who started the experiment with Polish blocks were provided with Polish instructions. Such a procedure was employed due to the fact that the language of instructions has been suggested to influence the level of activation of a particular linguistic system, thus putting participants in a specific language mode (Grosjean 1998: 140; Marian and Spivey 2003; Canseco-Gonzales et al. 2010).

4.3.3.1. Electrophysiological recording

EEG signals were recorded from 64 active Ag/AgCl electrodes (Brain Products): FP1, FP2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, C2, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2, PO10, AF7, AF3, AF4, AF8, F5, F1, F2, F6, FPz, FT7, FC3, FC4, FT8, FCz, C5, C1, Cz, C6, TP7, CP3, CPz, CP4, TP8, P5, P1, P2, P6, PO7, PO3, POz, PO4, PO8 at the standard extended 10/20 positions, with the ground placed at AFz. Additionally, bipolar electrodes were used to monitor horizontal (hEOG) and vertical (vEOG) eye movements. For horizontal eye movements, hEOG electrodes were situated horizontally from positions next to the outer rims of the eyes, while for vertical eye movements, vEOG electrodes were placed above and below the right eye. The EEG signal was amplified by the QuickAmp amplifier (Brain Products, Germany) with average reference, filtered with low-pass (cutoff frequency = 0.016 Hz), and stored at 1000 Hz per channel. All data were stored on a computer for offline analyses. Impedances were kept below 5 k Ω for each electrode. ERPs were time-locked to the onset of the second word of the word dyad.

Offline data analyses were performed using BrainVision Analyzer 2.0 software (Brain Products, Germany). Data were segmented from 100 ms before stimulus onset to 900 ms afterward, and filtered offline (Butterworth zero phase filter) with a high-pass filter set at 0.5 Hz (slope 24 dB/octave) and a low-pass filter set at 30 Hz (slope 24 dB/octave). Next, data were referred to baseline -100 to 0 ms before stimulus onset, and edited

for artifacts (rejecting trials with zero lines, rejecting trials with voltage differences higher than 150 μ V or voltage steps higher than 50 μ V). Ocular artifacts were corrected by the Gratton & Coles method. The following electrodes were selected for the statistical analyses: FC1, FC3, FCz, FC2, FC4, C1, C3, Cz, C2, C4, CP1, CP3, CPz, CP2, CP4, P1, P3, Pz, P2, P4.

The N400 peak detection analysis was computed separately for Polish and English, and involved averaging amplitudes for all electrodes over which between-language differences within the N400 time frame were most pronounced (i.e., C3, C1, CP3, CP1, P3, and P1). The peak detection analysis was performed using a semi-automatic mode, with a detection method based on global maxima in the predefined interval of 300-500 ms post stimulus onset. After the N400 peaks for both L1 and L2 were detected, a paired samples *t*-test was performed on the N400 peak latencies to L1 and L2 word dyads (Jan-kowiak et al. 2017: 8-10).

4.3.3.2. Overview of statistical analyses

Both accuracy ratings and reaction times were analyzed by performing 2 language (Polish/native dominant vs. English/non-native non-dominant) \times 4 utterance type (novel vs. conventional vs. literal vs. anomalous) repeated measures *ANOVAs*. For event-related potential analyses, mean amplitudes from 20 electrodes for each condition in each block were chosen. Along the anterior-posterior axis, the following electrodes were selected: FC3, FC1, FCz, FC2, FC4 (fronto-central), C3, C1, Cz, C2, C4 (central), CP3, CP1, CPz, CP2, CP4 (centro-parietal), P3, P1, Pz, P2, P4 (parietal). Along the left-right axis, the following electrodes were chosen: FC3, C3, CP3, P3 (left), FC1, C1, CP1, P1 (left medial), FCz, Cz, CPz, Pz (midline), FC4, C4, CP4, P4 (right), FC2, C2, CP2, P2 (right medial). Statistical analyses were performed between 100-900 ms after the presentation of a critical word. In order to avoid potential confounds caused by brain activity resulting from correct responses rather than a specific utterance type (VanRullen 2011), the ERP analyses were carried out on all responses. Visual inspection indicated three clear peaks, and the following three time frames were chosen for ERP analyses: the 150-250 ms time window (P200), the 300-500 ms time window (N400), and the 500-800 ms time window

(LPC). Furthermore, based on visual inspection, the N400 amplitudes for individual utterance types converged between 300-400 ms, and started to diverge at around 400 ms post stimulus onset. Consequently, the N400 time frame was further divided into the early (300-400 ms) and late (400-500 ms) N400 time window.

Mean amplitudes were analyzed using 2 language (Polish/native dominant vs. English/non-native non-dominant) \times 4 utterance type (novel metaphoric vs. conventional metaphoric vs. literal vs. anomalous) \times 4 anterior-posterior electrode position (fronto-central vs. central vs. centro-parietal vs. parietal) \times 5 laterality (left vs. left medial vs. midline vs. right medial vs. right) \times 4 block order (first vs. second vs. third vs. fourth block) repeated measures *ANOVAs*. In all analyses, significance values for pairwise comparisons were corrected for multiple comparisons using the Bonferroni correction. If Mauchly's tests indicated that the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied. In such cases, the original degrees of freedom are reported with the corrected *p* value (Jankowiak et al. 2017: 10). Block order did not interact with either utterance type or language within the P200, N400, and LPC time windows.

4.4. Results

4.4.1. Behavioral results

4.4.1.1. Accuracy rates

Accuracy ratings are reported as percentage of correct responses observed in the semantic decision task. For statistical analyses, accuracy rates were arcsin-transformed. The results revealed an interaction between language and utterance type, $F(3, 66) = 19.06, p < .001, \epsilon = .530, \eta_p^2 = .464$. Follow up analyses were carried out for each language separately. A repeated measures *ANOVA* with utterance types as factor performed on accuracy rates for Polish utterances showed a main effect of utterance type, $F(3, 66) = 55.65, p < .001, \epsilon = .442, \eta_p^2 = .717$. Pairwise comparisons confirmed that novel metaphors ($M = 48.85, SE = 3.89$) were rated less accurately than conventional metaphors ($M = 85.58, SE = 2.16$),

$p < .001$, as well as than literal ($M = 90.10$, $SE = 1.29$), $p < .001$, and anomalous utterances ($M = 92.85$, $SE = 1.89$), $p < .001$. There was no statistically significant difference between conventional metaphoric and literal, between conventional metaphoric and anomalous, as well as between literal and anomalous word pairs, $ps > .05$.

Similarly, a repeated measures *ANOVA* with utterance types as factor performed on accuracy rates for English utterances showed a main effect of utterance type, $F(3, 66) = 49.13$, $p < .001$, $\varepsilon = .377$, $\eta_p^2 = .678$. Pairwise comparisons revealed that novel metaphors ($M = 44.79$, $SE = 3.87$) differed from conventional metaphors ($M = 62.87$, $SE = 2.52$), $p < .001$, from literal utterances ($M = 84.90$, $SE = 1.39$), $p < .001$, as well as from anomalous utterances ($M = 86.90$, $SE = 2.93$), $p < .001$. Additionally, conventional metaphors differed from literal, $p < .001$, and from anomalous word pairs, $p < .001$. There was no statistically significant difference between accuracy ratings for literal and anomalous utterances, $p > .05$.

With regard to between-language differences, post-hoc tests revealed that conventional metaphors were rated as more meaningful in Polish than in English, $p < .001$, and the same pattern of results was observed for Polish and English literal expressions, $p < .001$, and Polish and English anomalous utterances, $p < .001$. There was no statistically significant difference between Polish and English novel metaphors, $p > .05$.¹⁰

In addition to the interaction, a main effect of language was found, $F(1, 22) = 103.95$, $p < .001$, $\eta_p^2 = .825$. The accuracy rate was higher for Polish ($M = 79.35$, $SE = 1.38$) than for English utterances ($M = 69.87$, $SE = 1.38$). Furthermore, a main effect of utterance type was found, $F(3, 66) = 55.93$, $p < .001$, $\varepsilon = .391$, $\eta_p^2 = .718$. Pairwise comparisons confirmed that novel metaphors ($M = 46.82$, $SE = 3.62$) differed from conventional metaphors ($M = 74.23$, $SE = 2.05$), $p < .001$, from literal utterances ($M = 87.50$, $SE = 1.18$), $p < .001$, as well as from anomalous utterances ($M = 89.87$, $SE = 2.34$), $p < .001$. Furthermore, conventional metaphors differed from literal, $p < .001$, and from anomalous utterances, $p = .002$. There was no statistically significant difference between accuracy rates for literal and anomalous word dyads, $p > .05$ (Jankowiak et al. 2017: 10-11). Mean accuracy rates per each utterance type in each language are provided in Fig. 22.

¹⁰ Correction for multiple comparisons was applied here, and the critical p level for significance was set to .012.

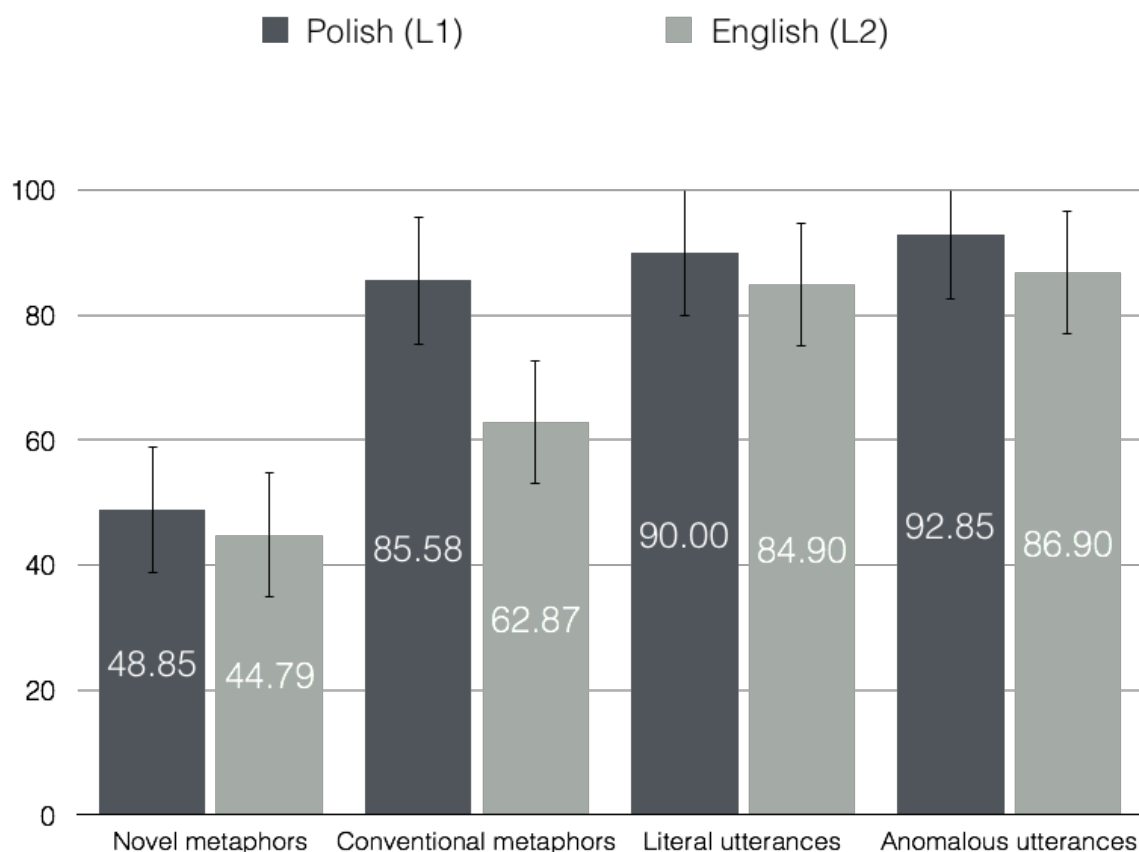


Fig. 22. Accuracy rates (%) for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, literal, and anomalous word dyads (adapted from Jankowiak et al. 2017: 11).

4.4.1.2. Reaction times

Reaction times (RTs) were measured time-locked to critical word onset. Only correct responses were included in statistical tests. For statistical analyses, reaction times were log-transformed. The results showed an interaction between language and utterance type, $F(3, 66) = 7.46, p < .001, \eta_p^2 = .253$. Follow up analyses were carried out for each language separately. A repeated measures *ANOVA* with utterance types as factor performed on reaction times for Polish utterances further revealed a main effect of utterance type, $F(3, 66) = 16.25, p < .001, \epsilon = .733, \eta_p^2 = .425$. Pairwise comparisons confirmed that novel metaphors ($M = 1121.25$ ms, $SE = 77.17$) elicited significantly longer RTs than conventional metaphoric ($M = 986.84$ ms, $SE = 79.19$), $p < .001$, literal ($M = 988.44$ ms, $SE = 80.87$), $p < .001$, and anomalous utterances ($M = 987.06$ ms, $SE = 73.05$), $p = .002$. There was no statistically significant difference between conventional metaphoric and

literal, between conventional metaphoric and anomalous, as well as between literal and anomalous word pairs, $ps > .05$.

Similarly, a repeated measures *ANOVA* with utterance types as factor performed on reaction times for English utterances showed a main effect of utterance type, $F(3, 66) = 9.54$, $p = .001$, $\varepsilon = .591$, $\eta_p^2 = .302$. However, pairwise comparisons revealed a somewhat different pattern of results in English as compared to Polish. Namely, English novel metaphors elicited longest RTs ($M = 1189.73$ ms, $SE = 87.09$), which were, however, similar to those evoked by conventional metaphors ($M = 1155.19$ ms, $SE = 92.36$) and anomalous word pairs ($M = 1113.68$ ms, $SE = 86.48$), $ps > .05$. Shortest reaction times were elicited by literal word pairs ($M = 1090.07$ ms, $SE = 85.87$), which differed significantly from both novel, $p < .001$, and conventional metaphors, $p = .001$. There was no statistically significant difference between anomalous and literal, and between anomalous and conventional metaphoric word pairs, $ps > .05$.

With regard to between-language differences, paired samples *t*-tests showed that English conventional metaphors differed from Polish conventional metaphors, $p < .001$. There was no statistically significant difference between Polish and English novel metaphoric, literal, as well as anomalous word pairs, $ps > .05$.¹¹

In addition to the interaction, a main effect of language was observed, $F(1, 22) = 8.24$, $p = .009$, $\eta_p^2 = .272$, with longer reaction times elicited by English ($M = 1137.17$ ms, $SE = 86.94$) than Polish word dyads ($M = 1020.90$ ms, $SE = 76.14$). Moreover, a main effect of utterance type was found, $F(3, 66) = 16.01$, $p < .001$, $\varepsilon = .659$, $\eta_p^2 = .421$. Pairwise comparisons confirmed that novel metaphors ($M = 1155.49$ ms, $SE = 79.00$) elicited longer RTs compared to literal utterances ($M = 1039.26$ ms, $SE = 80.54$), $p < .001$, conventional metaphors ($M = 1071.02$ ms, $SE = 83.20$), $p < .001$, and anomalous utterances ($M = 1050.37$ ms, $SE = 75.58$), $p = .004$. Additionally, conventional metaphors evoked longer response times than literal word pairs ($p = .039$). There was no statistically significant difference between anomalous and literal, and between anomalous and conventional metaphoric word pairs, $ps > .05$ (Jankowiak et al. 2017: 12-13). Mean reaction times per each utterance type in each language are provided in Fig. 23.

¹¹ Correction for multiple comparisons was applied here, and the critical *p* level for significance was set to .012.

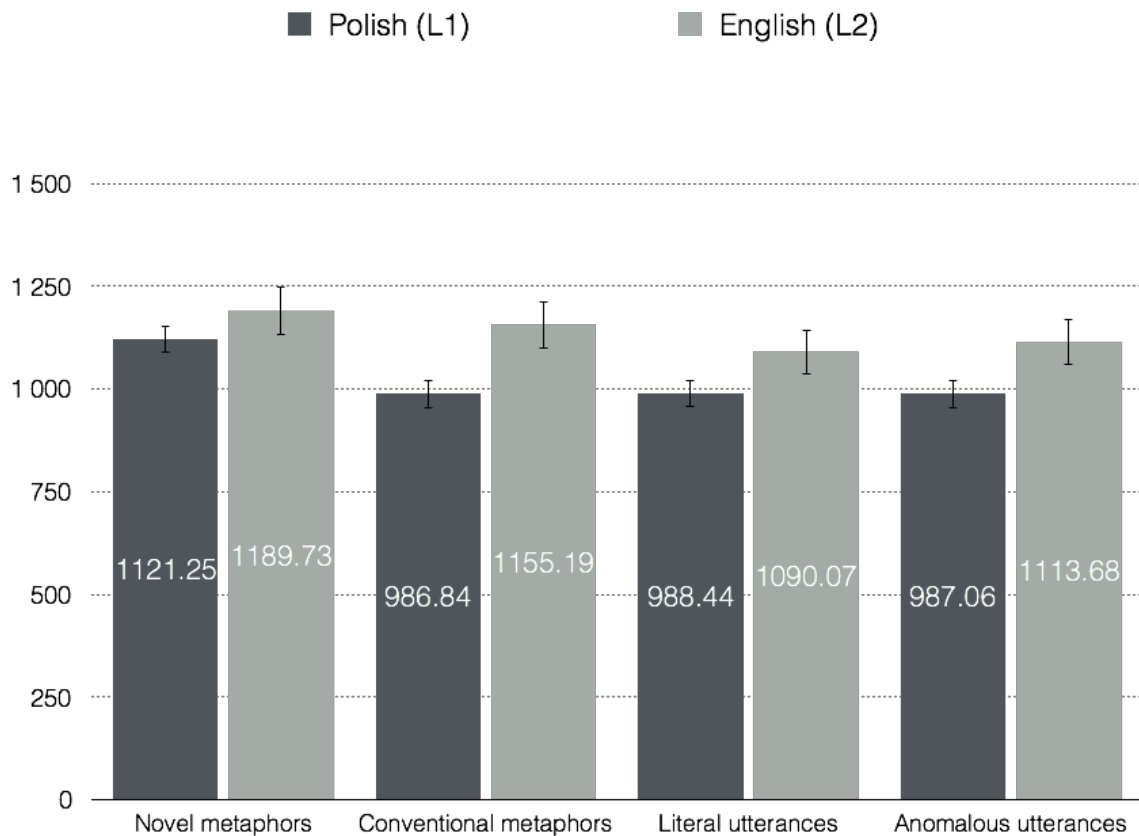


Fig. 23. Reaction times (ms) for Polish (dark grey) and English (light grey) novel metaphoric, conventional metaphoric, literal, and anomalous word dyads (adapted from Jankowiak et al. 2017: 13).

4.4.1.3. Correlation analyses

Since accuracy rates for Polish stimuli were higher than those for English utterances, correlation analyses were conducted in order to check whether the word dyads were evaluated similarly by the normative study participants (native speakers of Polish and English) and the EEG experiment participants (native speakers of Polish and L2 learners of English). The correlation analyses were carried out separately for each language, and were conducted on averaged values for all novel metaphoric, conventional metaphoric, as well as literal word dyads (Jankowiak et al. 2017: 12). The results showed a strong positive correlation between the accuracy rates for novel metaphoric, conventional metaphoric, and literal word pairs in the EEG experiment and the results from the meaningfulness ratings for both Polish, $r(226) = .86$, $p < .001$, and English word dyads, $r(226) = .73$, $p < .001$. Fig. 24 shows scatterplots representing the correlations observed for Polish and English stimuli.

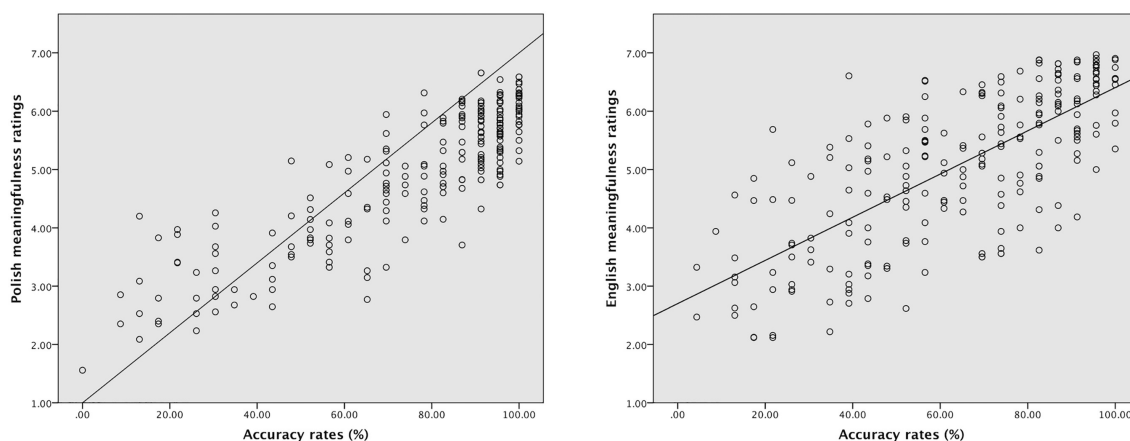


Fig. 24. Scatterplots showing the correlation between meaningfulness ratings (y-axis) and accuracy rates (x-axis) for Polish (left-hand side) and English (right-hand side) novel metaphoric, conventional metaphoric, and literal word dyads (after Jankowiak et al. 2017: 12).

Additionally, a follow-up survey was conducted on the familiarity of English stimuli, in which English L2 learners ($N = 35$; 23 females; $M_{age} = 20.46$, $SD = 0.82$) assessed how frequently they encountered English novel metaphoric, conventional metaphoric, and literal word dyads on a 7-point rating scale (1 = very rarely; 7 = very frequently).¹² The raters taking part in the familiarity ratings were all native speakers of Polish, who were L2 learners of English ($M_{age\ of\ acquisition} = 11.74$, $SD = 2.44$). Similarly to the participants of the EEG experiment, the raters were all students of the Faculty of English at Adam Mickiewicz University in Poznań, and they were thus proficient learners of English as their second language.

The results obtained from the rating test showed a main effect of utterance type, $F(2, 68) = 117.99$, $p < .001$, $\varepsilon = .566$, $\eta_p^2 = .776$. Pairwise comparisons revealed that English novel metaphors ($M = 2.62$, $SE = 0.15$) were rated as less familiar than conventional metaphors ($M = 3.21$, $SE = 0.18$), $p < .001$, as well as than literal expressions ($M = 4.23$, $SE = 0.14$), $p < .001$. Furthermore, conventional metaphors were less familiar than literal word dyads, $p < .001$. A correlation analysis was conducted in order to examine whether the familiarity of English word dyads was evaluated similarly by the native

¹² The procedures applied in the survey as well as the statistical test used for data analyses were the same as those used in the original normative study on stimuli familiarity, in which English native speakers assessed the familiarity of English experimental materials (see section 4.3.2.2).

speakers of English taking part in the original pretests (see section 4.3.2.2) and L2 learners of English. The results showed a strong positive correlation, $r(226) = .79, p < .001$ (Fig. 25).

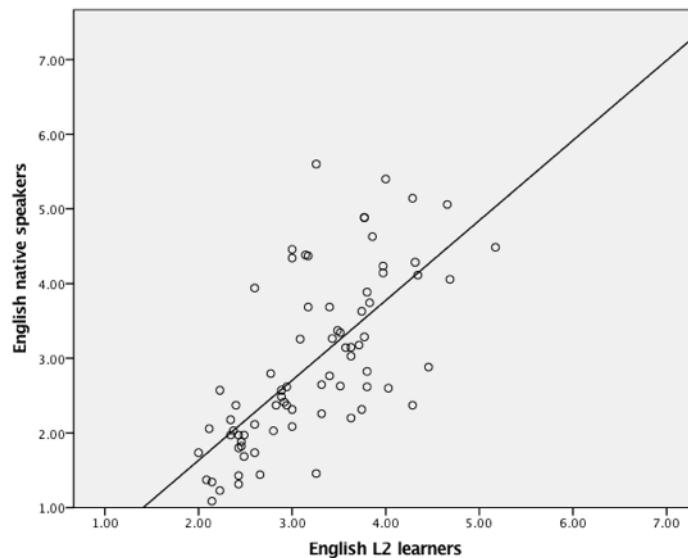


Fig. 25. Scatterplot showing the correlation between familiarity ratings obtained from English native speakers (y-axis) and English L2 learners (x-axis) for English novel metaphoric, conventional metaphoric, and literal word pairs.

4.4.2. Electrophysiological results

4.4.2.1. P200 (150-250 ms)

Within the 150-250 ms time window, an interaction was found between anterior-posterior electrode position, laterality, and language, $F(12, 264) = 2.33, p = .048, \varepsilon = .408, \eta_p^2 = .096$. Furthermore, within the same time window, an interaction was observed between laterality and language, $F(4, 88) = 12.38, p < .001, \varepsilon = .558, \eta_p^2 = .360$. To deconstruct the observed interactions, averaged amplitudes were computed for all left (FC3, FC1, C3, C1, CP3, CP1, P3, P1) and all right electrodes (FC2, FC4, C2, C4, CP2, CP4, P2, P4) separately for fronto-central, central, centro-parietal, and parietal electrode posi-

tions. For each of these levels of the anterior-posterior axis, a 2 language (Polish vs. English) \times 2 laterality (left vs. right) repeated measures *ANOVA* was performed. An interaction between language and laterality was observed over central, centro-parietal, and parietal electrodes (Table 4), but not over fronto-central electrodes ($p > .05$) To test whether between-language differences were more pronounced over left or right sites, paired samples *t*-tests were performed on amplitudes for English and Polish utterances averaged separately for left central, left centro-parietal, left parietal, right central, right centro-parietal, and right parietal electrodes (Table 5). The results showed maximal between-language differences over left central and left centro-parietal electrodes, and less pronounced over left parietal electrodes, with reduced P200 amplitudes for Polish compared to English word pairs (Fig. 26). No between-language differences were observed over right sites, $ps > .05$ (Jankowiak et al. 2017: 13-14).

Table 4. An interaction between language and laterality over central, centro-parietal, and parietal electrode positions within the 150-250 ms time window (adapted from Jankowiak et al. 2017: 14).

| Electrode position | An interaction between language and laterality |
|--------------------|--|
| Central | $F(4, 88) = 8.00, p < .001, \eta_p^2 = .267$ |
| Centro-parietal | $F(4, 88) = 10.46, p < .001, \eta_p^2 = .322$ |
| Parietal | $F(4, 88) = 9.08, p < .001, \eta_p^2 = .292$ |

Table 5. Between-language differences within the 150-250 ms time window, with mean amplitudes for Polish and English utterances over left and right central, centro-parietal, and parietal electrode positions (adapted from Jankowiak et al. 2017: 14).

| Electrode position | Mean amplitude for Polish utterances | Mean amplitude for English utterances | Between-language difference |
|--------------------|---|---------------------------------------|---|
| Central | Left (C3, C1) $M = 1.28,$ $SE = 0.21$ | $M = 1.66,$ $SE = 0.23$ | $t(22) = 4.33, p < .001,$ $r = .68$ |
| | Right (C4, C2) $M = 1.34,$ $SE = 0.17$ | $M = 1.47,$ $SE = 0.19$ | $t(22) = 1.39, p = .178,$ $r = .28$ |
| Centro-parietal | Left (CP3, CP1) $M = 0.23,$ $SE = 0.18$ | $M = 0.58,$ $SE = 0.23$ | $t(22) = 4.31, p < .001,$ $r = .68$ |
| | Right (CP4, CP2) $M = 0.45,$ $SE = 0.19$ | $M = 0.48,$ $SE = 0.19$ | $t(22) = 0.31, p = .760,$ $r = .06$ |
| Parietal | Left (P3, P1) $M = -0.70,$ $SE = 0.18$ | $M = -0.42,$ $SE = 0.22$ | $t(22) = 2.47, p = .022,$ $r = .47$ |
| | Right (P4, P2) $M = 0.06,$ $SE = 0.29$ | $M = -0.13,$ $SE = 0.34$ | $t(22) = -1.19, p = .245,$ $r = .24$ |

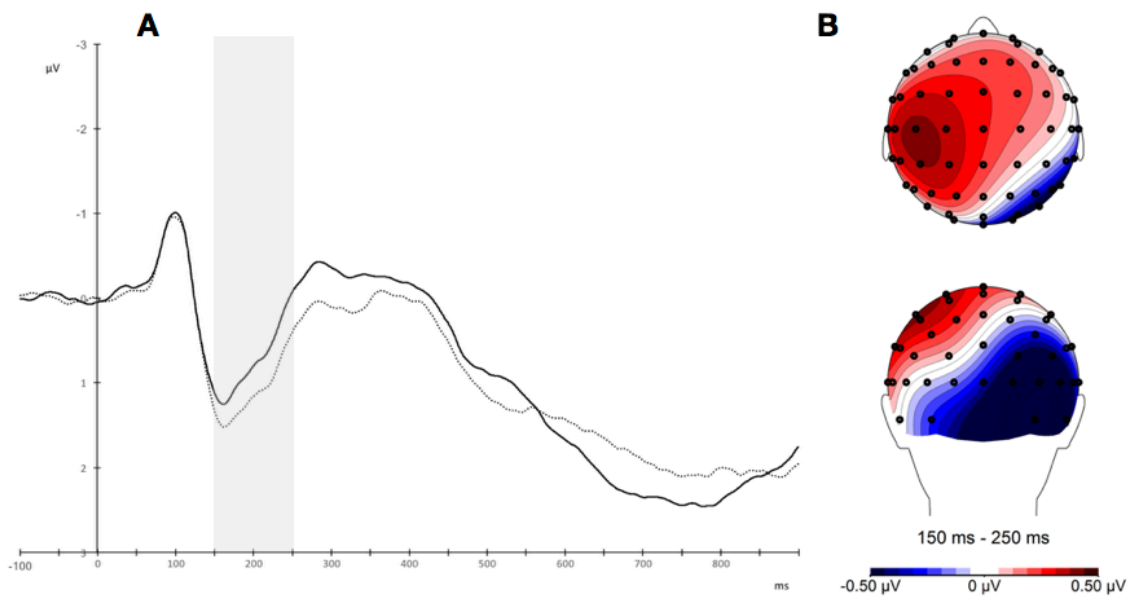


Fig. 26. A: Grand averages for Polish (solid line) and English (dotted line) utterances over left central and centro-parietal electrodes, where the P200 effect was maximal (adapted from Jankowiak et al. 2017: 15). B: Topographic distribution of Polish and English word dyads in the 150-250 ms time window. Voltage maps were obtained for the averaged value of difference waves (English word pairs minus Polish word pairs; after Jankowiak et al. 2017: 16).

4.4.2.2. Early N400 (300-400 ms)

The analysis of the early N400 time window (300-400 ms) revealed an interaction between anterior-posterior electrode position and language, $F(3, 66) = 5.12$, $p = .027$, $\varepsilon = .394$, $\eta_p^2 = .189$, and an interaction between laterality and language, $F(4, 88) = 14.60$, $p < .001$, $\varepsilon = .522$, $\eta_p^2 = .399$. To deconstruct the interactions, averaged amplitudes were computed for all left and all right electrodes separately for fronto-central, central, centro-parietal, and parietal electrode positions, similarly to the analyses performed within the P200 time frame (see section 4.4.2.1). For each of these levels of the anterior-posterior axis, a 2 language (Polish vs. English) \times 2 laterality (left vs. right) repeated measures *ANOVA* was performed. The analyses revealed a main effect of language over central, centro-parietal, and parietal electrodes, where English utterances evoked attenuated N400 responses compared to Polish word dyads (Table 6). The effect was not observed over fronto-central electrode sites ($p > .05$).

Table 6. Main effect of language within the 300-400 ms time window, with mean amplitudes for Polish and English utterances over central, centro-parietal, and parietal electrode positions (adapted from Jankowiak et al. 2017: 17).

| Electrode position | Mean amplitude for Polish utterances | Mean amplitude for English utterances | Main effect of language |
|--------------------|--------------------------------------|---------------------------------------|---|
| Central | $M = -0.77, SE = 0.23$ | $M = -0.39, SE = 0.20$ | $F(1, 22) = 6.49, p = .018, \eta_p^2 = .228$ |
| Centro-parietal | $M = -0.20, SE = 0.24$ | $M = 0.49, SE = 0.22$ | $F(1, 22) = 16.46, p = .001, \eta_p^2 = .428$ |
| Parietal | $M = 0.90, SE = 0.34$ | $M = 1.63, SE = 0.33$ | $F(1, 22) = 20.25, p < .001, \eta_p^2 = .479$ |

Next, amplitudes across all left and all right electrodes for all central, centro-parietal, and parietal electrode positions were averaged. A 2 language (Polish vs. English) \times 2 laterality (left vs. right) repeated measures *ANOVA* was performed, and showed an interaction between language and laterality, $F(1, 22) = 24.11, p < .001, \eta_p^2 = .523$. To decompose this interaction, paired samples *t*-tests were performed on averaged amplitudes separately for all left central, centro-parietal, and parietal electrodes, as well as all right central, centro-parietal, and parietal electrodes. The results revealed attenuated amplitudes for English relative to Polish word pairs (Table 7, Fig. 27). The effect was most pronounced over right electrodes, where the mean difference between amplitudes for English vs. Polish utterances was $.828 \mu\text{V}$ ($r = .79$). Over left sites, the mean difference was much smaller ($.305 \mu\text{V}$, $r = .44$; Jankowiak et al. 2017: 14-15).

Table 7. Between-language differences within the 300-400 ms time window, with mean amplitudes for Polish and English utterances over left and right electrode positions (adapted from Jankowiak et al. 2017: 17).

| Electrode position | Mean amplitude for Polish utterances | Mean amplitude for English utterances | Between-language difference |
|-------------------------------------|--------------------------------------|---------------------------------------|------------------------------------|
| Left (C3, C1, CP3, CP1, P3, P1) | $M = 0.22, SE = 0.20$ | $M = 0.52, SE = 0.20$ | $t(22) = -2.31, p = .031, r = .44$ |
| Right (C4, C2, CP4, CP2, P4, P2) | $M = 0.06, SE = 0.27$ | $M = 0.90, SE = 0.22$ | $t(22) = -6.13, p < .001, r = .79$ |

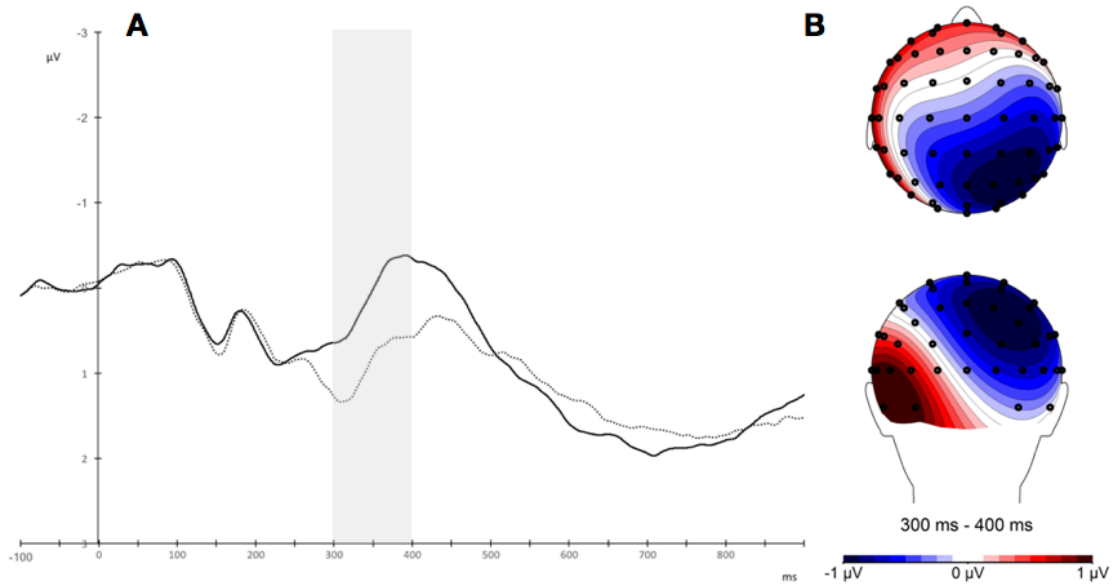


Fig. 27. A: Grand averages for Polish (solid line) and English (dotted line) utterances over right central, centro-parietal, and parietal electrodes, where the early N400 effect was maximal (adapted from Jankowiak et al. 2017: 19).

B: Topographic distribution of Polish and English word dyads in the 300-400 ms time window. Voltage maps were obtained for the averaged value of difference waves (Polish word pairs minus English word pairs; after Jankowiak et al. 2017: 18).

To test whether the N400 peak latency to the non-native and non-dominant language (English) was delayed relative to the native and dominant language (Polish), a peak detection analysis was performed. Since differences in the N400 amplitudes between L1 and L2 were most pronounced over right central, centro-parietal, and parietal electrodes, peak latencies were estimated for the averaged waveforms over these electrodes, separately for Polish and English. The comparison between the N400 peak latencies in response to L1 and L2 utterances confirmed a delayed N400 response to L2 ($M = 424.57$, $SD = 38.93$) compared to L1 word dyads ($M = 406.26$, $SD = 34.02$), $t(22) = -2.15$, $p = .042$, $r = .42$ (Jankowiak et al. 2017: 14-15).

4.4.2.3. Late N400 (400-500 ms)

Within the 400-500 ms time window, an interaction between utterance type and laterality was found, $F(12, 264) = 2.95$, $p = .036$, $\epsilon = .266$, $\eta_p^2 = .118$. To deconstruct the interaction, averaged amplitudes were computed for all left (left and left-medial), all midline, and all

right (right and right-medial) electrode positions. For each of these levels of the left-right axis, a repeated measures *ANOVA* was performed on mean amplitudes for the four utterance types (novel metaphoric vs. conventional metaphoric vs. literal vs. anomalous utterances). The results showed a main effect of utterance type over midline, $F(3, 66) = 4.35$, $p = .007$, $\eta_p^2 = .165$, and right electrode positions, $F(3, 66) = 4.08$, $p = .01$, $\eta_p^2 = .156$ (Fig. 28, Fig. 29, Table 8), but not over left electrodes ($p > .05$). Pairwise comparisons showed that anomalous word pairs evoked more pronounced N400 amplitudes relative to literal utterances over midline electrodes ($p = .014$), and relative to conventional metaphors over both midline ($p = .045$) and right electrode positions ($p = .043$). Furthermore, the difference between novel metaphoric and anomalous word dyads was marginally significant over midline electrodes ($p = .078$). A linear effect was additionally observed, with maximal amplitudes for anomalous, followed by novel metaphoric, conventional metaphoric, and lowest amplitudes for literal word pairs over midline, $F(1, 22) = 13.12$, $p = .002$, $\eta_p^2 = .374$, and right electrode positions, $F(1, 22) = 7.80$, $p = .011$, $\eta_p^2 = .262$, but not over left electrodes, $p > .05$ (Jankowiak et al. 2017: 15-18).

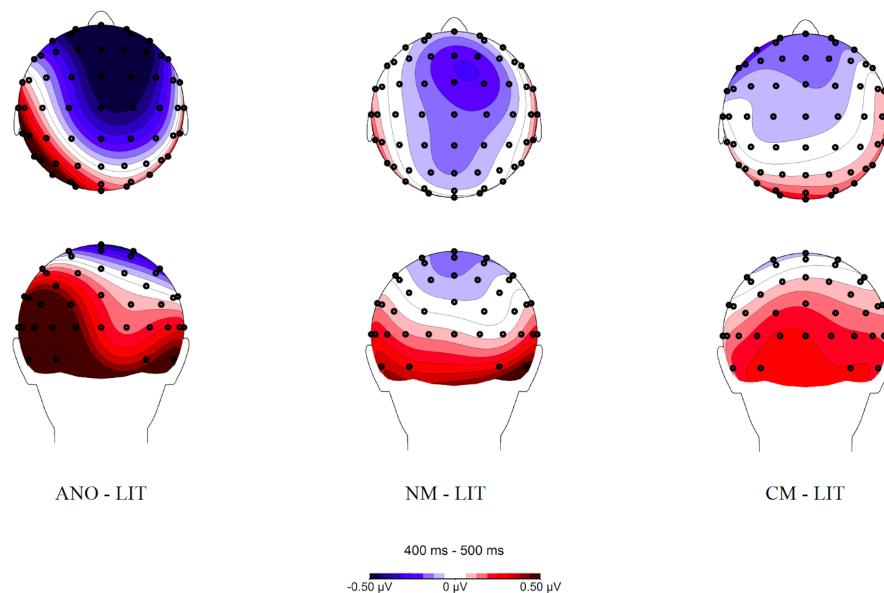


Fig. 28. Topographic distribution of novel metaphoric (NM), conventional metaphoric (CM), literal (LIT), and anomalous (ANO) word dyads within the 400-500 ms time window. Voltage maps were obtained for the averaged value of difference waves (anomalous minus literal word pairs, novel metaphoric minus literal word pairs, and conventional metaphoric minus literal word pairs; after Jankowiak et al. 2017: 20).

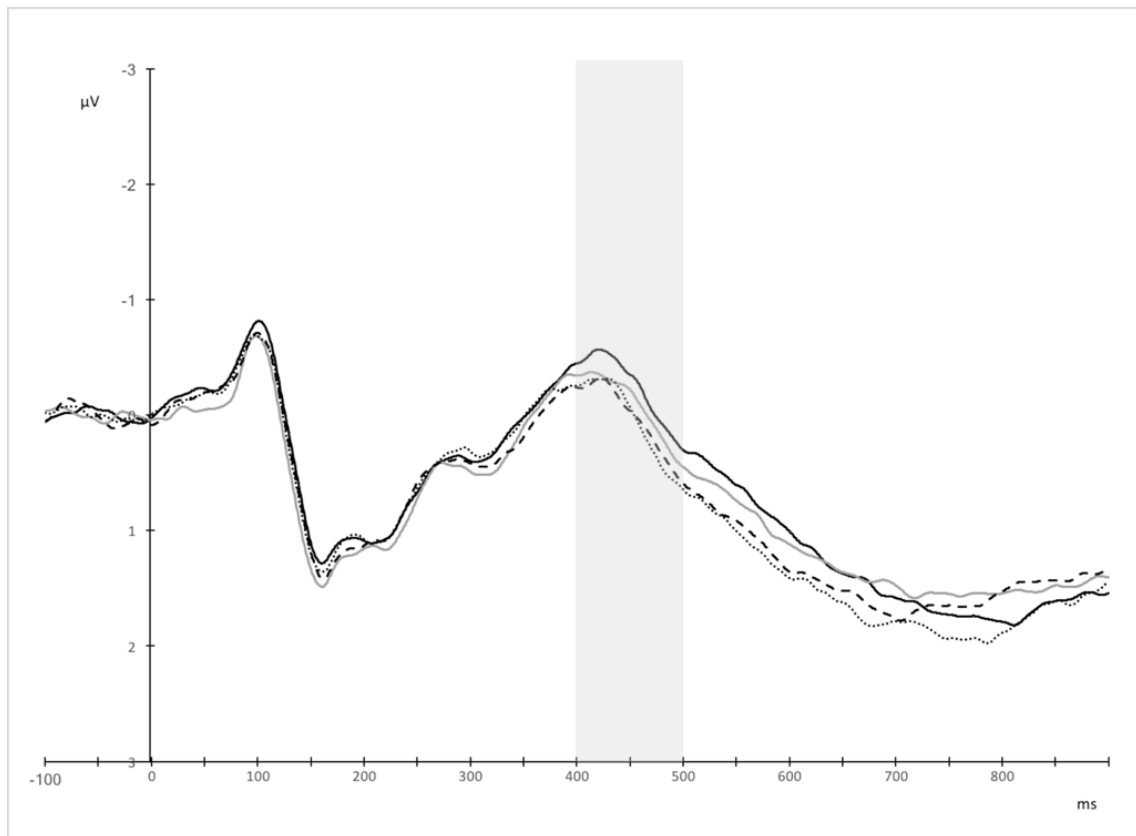


Fig. 29. Grand averages for anomalous (black solid line), novel metaphoric (grey solid line), conventional metaphoric (black dashed line), and literal word dyads (black dotted line) over midline and right electrode positions (adapted from Jankowiak et al. 2017: 21).

Table 8. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous (ANO) word pairs over midline and right electrode positions within the 400-500 ms time window (adapted from Jankowiak et al. 2017: 19).

| Electrode position | Mean amplitude for literal utterances | Mean amplitude for conventional metaphors | Mean amplitude for novel metaphors | Mean amplitude for anomalous utterances | Pairwise comparisons |
|--------------------|---------------------------------------|---|------------------------------------|---|--|
| Midline | $M = -0.01$, $SE = 0.25$ | $M = -0.06$, $SE = 0.26$ | $M = -0.19$, $SE = 0.22$ | $M = -0.35$, $SE = 0.28$ | LIT vs. ANO ($p = .014$); CM vs ANO ($p = .045$); NM vs. ANO ($p = .078$) |
| Right | $M = 0.06$, $SE = 0.19$ | $M = 0.06$, $SE = 0.21$ | $M = -0.05$, $SE = 0.19$ | $M = -0.20$, $SE = 0.23$ | CM vs. ANO ($p = .043$) |

4.4.2.4. Late positivity (500-800 ms)

The analysis of the 500-800 ms time window showed an interaction between anterior-posterior electrode site and utterance type, $F(9, 198) = 6.87, p = .001, \varepsilon = .314, \eta_p^2 = .238$. To deconstruct the interaction, averaged amplitudes were computed separately for all fronto-central, central, centro-parietal, and parietal electrode positions per each utterance type. For each of these levels of the anterior-posterior axis, a repeated measures *ANOVA* was performed on mean amplitudes for the four utterance types (novel metaphoric vs. conventional metaphoric vs. literal vs. anomalous word dyads). The results showed effects of utterance type over fronto-central, $F(3, 66) = 12.28, p < .001, \eta_p^2 = .358$, and central electrode positions, $F(3, 66) = 8.67, p < .001, \eta_p^2 = .283$ (Table 9, Fig. 30), but not over centro-parietal and parietal electrodes ($ps > .05$). Pairwise comparisons showed that over fronto-central electrodes, literal and conventional metaphoric word dyads evoked larger LPC amplitudes than novel metaphors ($p < .001, p = .002$, respectively). Similarly, literal and conventional metaphoric utterances evoked increased positivity relative to anomalous word pairs ($p = .002, p = .03$, respectively). Over central electrodes, literal utterances elicited larger LPC amplitudes compared to novel metaphoric and anomalous utterances ($p = .001, p = .002$, respectively), and conventional metaphors evoked increased positivity compared to novel metaphoric word dyads, $p = .031$. A linear effect was additionally observed over both fronto-central, $F(1, 22) = 23.13, p < .001, \eta_p^2 = .513$, and central electrodes, $F(1, 22) = 22.55, p < .001, \eta_p^2 = .506$, with the most pronounced LPC response evoked by literal, followed by conventional metaphoric, novel metaphoric, and finally anomalous word dyads. The effect was not observed over centro-parietal and parietal electrode sites ($ps > .05$; Jankowiak et al. 2017: 17).

Table 9. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous (ANO) word pairs over fronto-central and central electrode positions within the 500-800 ms time window (adapted from Jankowiak et al. 2017: 23).

| Electrode position | Mean amplitude for literal utterances | Mean amplitude for conventional metaphors | Mean amplitude for novel metaphors | Mean amplitude for anomalous utterances | Pairwise comparisons |
|--------------------|---------------------------------------|---|------------------------------------|---|---|
| Fronto-central | $M = 0.77,$ $SE = 0.25$ | $M = 0.63,$ $SE = 0.23$ | $M = 0.33,$ $SE = 0.26$ | $M = 0.21,$ $SE = 0.28$ | LIT vs. NM ($p < .001$); LIT vs. ANO ($p = .002$); |

| | | | | | |
|---------|----------------------------|----------------------------|----------------------------|----------------------------|--|
| | | | | | CM vs. NM ($p = .002$); CM vs ANO ($p = .03$) |
| Central | $M = 1.36,$ $SE = 0.20$ | $M = 1.21,$ $SE = 0.20$ | $M = 1.03,$ $SE = 0.21$ | $M = 1.02,$ $SE = 0.24$ | LIT vs. NM ($p = .001$); LIT vs. ANO ($p = .002$); CM vs. NM ($p = .031$) |

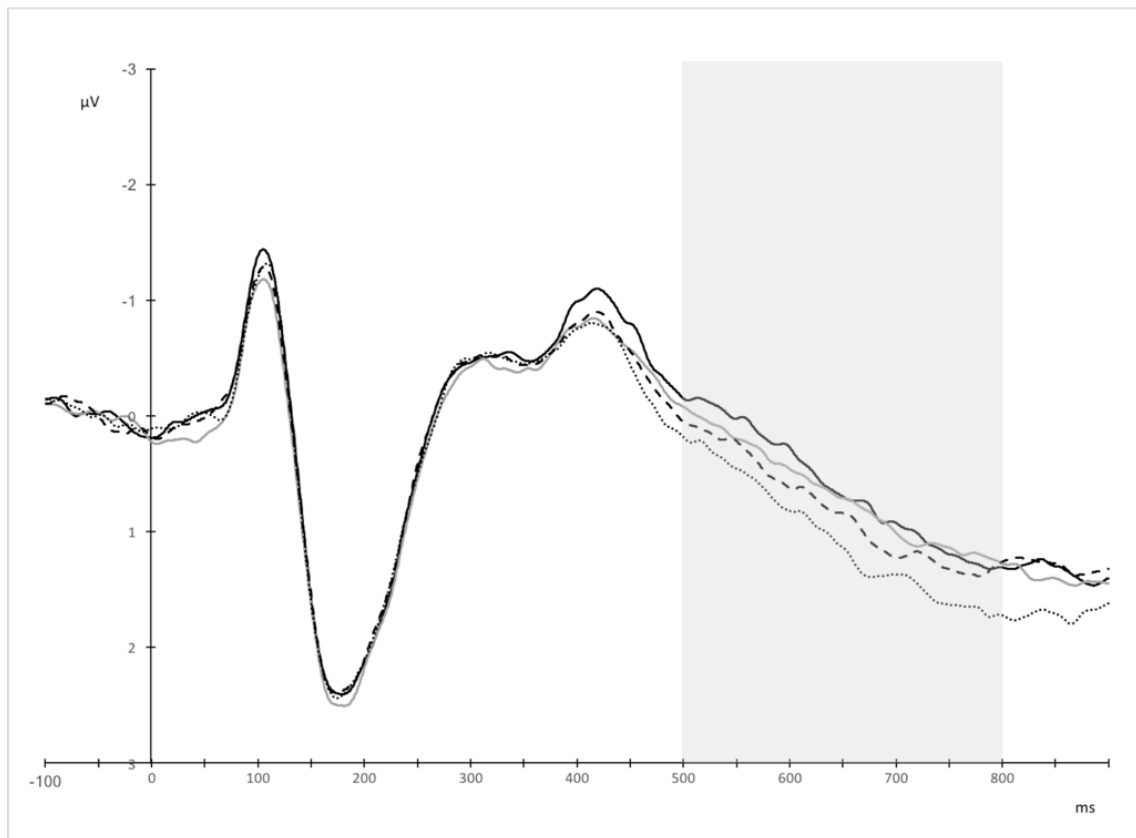


Fig. 30. Grand averages for anomalous (black solid line), novel metaphoric (grey solid line), conventional metaphoric (black dashed line), and literal word dyads (black dotted line) over fronto-central and central electrodes (adapted from Jankowiak et al. 2017: 22).

Additionally, an interaction was found between language and utterance type, $F(3, 66) = 4.32, p = .008, \eta_p^2 = .164$. Follow up analyses were carried out for each language separately. A repeated measures *ANOVA* performed on mean amplitudes for Polish utterances revealed a main effect of utterance type, $F(3, 66) = 4.65, p = .013, \varepsilon = .711, \eta_p^2 = .175$. Pairwise comparisons further showed that Polish novel metaphors evoked a reduced LPC response compared to both conventional metaphors, $p = .002$, and literal word pairs, $p = .021$. No statistically significant differences were found between literal

and conventional metaphoric, between anomalous and novel metaphoric, between anomalous and literal, and between anomalous and conventional metaphoric word dyads, $ps > .05$. Additionally, a linear effect across the utterance types was found, $F(1, 22) = 4.53, p = .045, \eta_p^2 = .171$, with maximal amplitudes evoked by conventional metaphoric, followed by literal, anomalous, and novel metaphoric word pairs.

Similarly, a repeated measures *ANOVA* performed on mean amplitudes for English utterances revealed a main effect of utterance type, $F(3, 66) = 5.73, p = .002, \eta_p^2 = .207$. Pairwise comparisons further showed that English literal utterances elicited more positive amplitudes than both conventional metaphoric meanings, $p = .019$, and anomalous word dyads, $p = .033$. No statistically significant differences were found between conventional metaphoric and novel metaphoric, between conventional metaphoric and anomalous, between novel metaphoric and anomalous, and between novel metaphoric and literal word dyads, $ps > .05$. Additionally, a linear effect across the utterance types was found, $F(3, 66) = 5.44, p = .013, \eta_p^2 = .251$ (Jankowiak et al. 2017: 17-20), with maximal amplitudes elicited by literal, followed by anomalous, novel metaphoric, and conventional metaphoric word dyads (Fig. 31). Mean amplitudes for each utterance type in Polish and English are reported in Table 10.

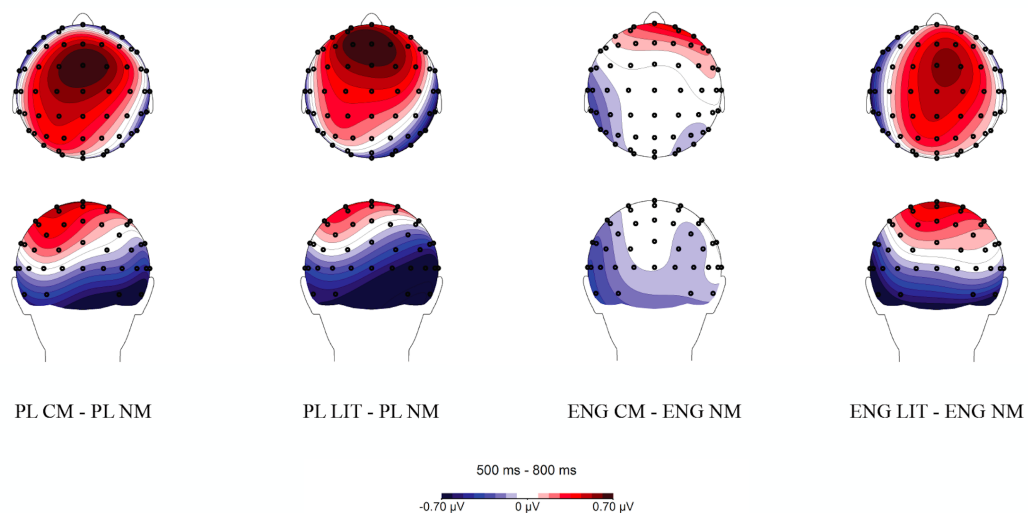


Fig. 31. Topographic distribution of novel metaphoric (NM), conventional metaphoric (CM), literal (LIT), and anomalous (ANO) word dyads in Polish (PL) and in English (ENG) within the 500-800 ms time window. Voltage maps were obtained for the averaged value of difference waves (Polish conventional metaphoric minus Polish novel metaphoric word pairs, Polish literal minus Polish novel metaphoric word pairs, English conventional metaphoric minus English novel metaphoric word pairs, and English literal minus English novel metaphoric word pairs; after Jankowiak et al. 2017: 23).

Table 10. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous word pairs (ANO) in Polish and English within the 500-800 ms time window (adapted from Jankowiak et al. 2017: 23).

| Language | Mean amplitude for literal utterances | Mean amplitude for conventional metaphors | Mean amplitude for novel metaphors | Mean amplitude for anomalous utterances | Pairwise comparisons |
|----------|---------------------------------------|---|------------------------------------|---|--|
| Polish | $M = 1.64$, $SE = 0.19$ | $M = 1.76$, $SE = 0.19$ | $M = 1.38$, $SE = 0.16$ | $M = 1.50$, $SE = 0.23$ | LIT vs. NM ($p = .021$); CM vs. NM ($p = .002$) |
| English | $M = 1.59$, $SE = 0.20$ | $M = 1.25$, $SE = 0.19$ | $M = 1.29$, $SE = 0.18$ | $M = 1.32$, $SE = 0.22$ | LIT vs. CM ($p = .019$); LIT vs. ANO ($p = .033$) |

In addition to the interactions, a main effect of utterance type was observed, $F(3, 66) = 5.89$, $p = .001$, $\eta_p^2 = .211$. Pairwise comparisons showed that literal utterances evoked more positive LPC amplitudes than novel metaphoric, $p = .003$, and conventional metaphoric elicited a more positive LPC response relative to novel metaphoric word dyads, $p = .015$. The difference between literal and anomalous word dyads was marginally significant, $p = .066$. In addition to the main effect, a linear effect across the utterance types was observed, with maximal amplitudes evoked by literal utterances, followed by conventional metaphors, followed by anomalous word pairs, and lowest amplitudes for novel metaphors, $F(1, 22) = 10.56$, $p = .004$, $\eta_p^2 = .324$ (Table 11, Fig. 32; Jankowiak et al. 2017: 20).

Table 11. Mean amplitudes for literal (LIT), conventional metaphoric (CM), novel metaphoric (NM), and anomalous (ANO) word pairs within the 500-800 ms time window (adapted from Jankowiak et al. 2017: 24).

| Mean amplitude for literal utterances | Mean amplitude for conventional metaphors | Mean amplitude for novel metaphors | Mean amplitude for anomalous utterances | Pairwise comparisons |
|---------------------------------------|---|------------------------------------|---|--|
| $M = 1.62$, $SE = 0.17$ | $M = 1.51$, $SE = 0.15$ | $M = 1.33$, $SE = 0.15$ | $M = 1.41$, $SE = 0.20$ | LIT vs. NM ($p = .003$); CM vs. NM ($p = .015$) |

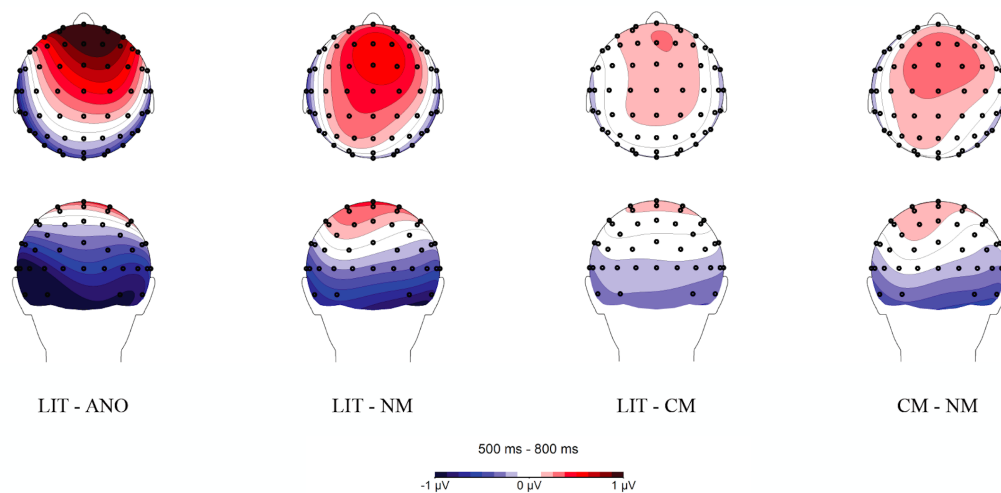


Fig. 32. Topographic distribution of novel metaphoric (NM), conventional metaphoric (CM), literal (LIT), and anomalous (ANO) word dyads within the 500-800 ms time window. Voltage maps were obtained for the averaged value of difference waves (literal minus anomalous word pairs, literal minus novel metaphoric word pairs, literal minus conventional metaphoric word pairs, and conventional metaphoric minus novel metaphoric word pairs; after Jankowiak et al. 2017: 24).

4.5. Discussion and conclusion

The present experiment was aimed to test whether metaphor conventionality modulates metaphoric meaning comprehension in both the native and non-native language. To this end, late proficient unbalanced bilingual speakers performed a two-choice semantic decision task in response to novel metaphoric, conventional metaphoric, literal, and anomalous word pairs presented in the native (Polish) and non-native language (English). The analyses were based on both behavioral data as well as event-related brain potentials in order to provide the exact time-course of novel and conventional metaphoric meaning processing in bilingualism.

Between-language utterance-specific effects were observed in both behavioral and event-related potential results. Namely, conventional metaphoric word pairs elicited longer reaction times along with lower accuracy rates in the non-native relative to the native tongue, thus indicating that they were more difficult to comprehend in L2. Such results were corroborated by the ERP patterns observed within the LPC time window, where conventional metaphors evoked sustained negativity in L2, but not in L1. Such

prolonged negativity might have overlapped with the LPC response, and might be interpreted as indicative of either access to the non-literal route during L2 conventional metaphor comprehension or an ongoing difficulty in comprehending conventional metaphoric word dyads in the non-native language (Arzouan et al. 2007; Goldstein et al. 2012; Mashal et al. 2015). Importantly, no between-language differences were observed for novel metaphors, which thus seemed to be similarly cognitively taxing in both languages.

In line with previous research, differences between languages that were independent of utterance type were observed in both behavioral and ERP results. Namely, L2 utterances elicited longer response time and lower accuracy ratings compared to L1 word pairs. ERP results further revealed an attenuated N400 response to the non-native relative to the native language, suggesting a decreased interconnectivity for L2 items within the semantic memory network (Midgley et al. 2009b). Furthermore, a delay in the N400 peak latency was observed in response to L2, indicating a delayed activation of semantic information in L2 compared to L1 (Weber-Fox and Neville 1996; Phillips et al. 2004; Moreno and Kutas 2005; Braunstein et al. 2012), potentially due to differences in language dominance (Dijkstra and van Heuven 2002). In addition, a more pronounced P200 response was observed to L2 compared to L1, which might be accounted for by a lower subjective frequency of L2 than L1 lexical items in L2 non-dominant bilingual speakers (Dambacher et al. 2006). A more elaborate discussion on the obtained results together with the implications and possible limitations of the present study is presented in Chapter 6.

One of the most intriguing results observed in the present experiment is that while conventional metaphor comprehension seemed to be modulated by language nativeness, no between-language differences were observed with regard to novel metaphor, thus pointing to potentially similar mechanisms engaged when comprehending novel metaphoric meanings in both languages. In order to further investigate that issue, Experiment 2 (see Chapter 5) was aimed to confirm whether and to what extent cognitive mechanisms involved in novel metaphor comprehension are language-independent. To this end, in Experiment 2, L1 and L2 novel metaphors were presented as either novel nominal metaphors (*A is B*) or novel similes (*A is like B*). In line with the Career of Metaphor Model (Bowdle and Gentner 2005), since similes automatically initiate comparison processes, their linguistic form is likely to ease novel metaphor comprehension. The main objective of Experiment 2 was therefore to examine whether such a facilitatory effect would be

observed in both languages, which would confirm that mechanisms involved in novel meaning comprehension are independent of language nativeness.

Chapter 5: Experiment 2: Novel nominal metaphor and novel simile comprehension in bilingualism

5.1. Introductory remarks

Experiment 1 (Chapter 4) revealed that while conventional metaphors are more taxing to comprehend in the non-native compared to the native tongue, novel metaphor comprehension seems to be similarly resource intensive in the two languages (Jankowiak et al. 2017). The experiment reported in this chapter aimed to further show whether and to what extent mechanisms engaged in novel meaning processing are similar in L1 and L2 by examining the comprehension of novel meanings presented in two different linguistic forms: as nominal metaphors (*A is B*) and as similes (*A is like B*). Even though theoretical accounts as well as previous monolingual studies have suggested that novel metaphor comprehension is facilitated when a novel meaning is presented as a simile, which initiates comparison mechanisms between target and base concepts, this assumption has thus far not been tested in the context of bilingualism. Therefore, the main objective of the present experiment was to investigate whether the comparison form (i.e., a simile) eases novel metaphor comprehension in both the native and non-native language. To this end, late proficient unbalanced Polish (L1) – English (L2) bilinguals performed a two-choice semantic decision task in response to novel nominal metaphors, novel similes, as well as literal and anomalous sentences.

5.2. Aims, research questions and hypotheses

Previous research on nominal metaphor and simile comprehension has been mostly conducted in the monolingual context, and has frequently pointed to metaphor conventionality as one of factors modulating figurative meaning processing (e.g., Bowdle and Gentner 2005; Arzouan et al. 2007; Lai et al. 2009; Shibata et al. 2010; Lai and Curran 2013). As postulated in the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), novel (unfamiliar) metaphors are more difficult to comprehend due to the fact that they require meaning construction processes, which are based on comparison mechanisms between target and base concepts. Following this assumption, novel metaphor comprehension should be facilitated when novel metaphors are presented as similes, whose comparison form “like” initiates comparison mechanisms. Such a hypothesis has previously been supported in behavioral studies (Bowdle and Gentner 2005), computational simulations (Utsumi 2011), as well as electrophysiological research (Lai and Curran 2013). Nonetheless, previous experiments have not examined whether comparison mechanisms initiated when processing similes could facilitate novel meaning comprehension also in the bilingual context. Therefore, the main objective of the current experiment was to investigate whether novel meanings are easier to comprehend when they are presented as similes compared to nominal metaphors in both the native and non-native language.

In addition, similarly to Experiment 1, the present study examined the automaticity of bilingual language comprehension, as reflected in behavioral measures. As previously observed in a number of behavioral experiments (e.g., Dijkstra et al. 1998a; Dijkstra et al. 1999; de Groot et al. 2002), and in line with the results of Experiment 1 (see section 4.4.1.2), longer response times for the non-native compared to the native language might point to a slower and less automatic processing of the non-dominant tongue (Dijkstra and van Heuven 2002). The next objective of the present study was therefore to corroborate the previously observed research findings.

Two independent variables were tested in the present experiment: language (Polish: native and dominant language; English: non-native and non-dominant language) and sentence type (novel nominal metaphors, novel similes, literal, and anomalous sentences). Dependent variables included reaction times and accuracy rates. Two main research questions were addressed in the study:

- 1) Does a comparison form present in similes modulate novel metaphor comprehension in both the native and non-native tongue?
- 2) Does non-native and non-dominant language comprehension take more time relative to the native and dominant tongue?

Three research hypotheses have been formulated in order to address the two aforementioned research questions:

Hypothesis 1: *In the native language, novel similes will elicit shorter reaction times and higher accuracy rates compared to novel nominal metaphors, both of which will evoke longer response times and lower accuracy ratings compared to literal sentences.*

Previous monolingual research into novel simile and novel nominal metaphor comprehension has shown that novel similes, whose linguistic form initiates comparison mechanisms, are faster and easier to comprehend. For instance, Bowdle and Gentner (2005) found a higher preference along with shorter RTs for novel similes (e.g., *A mind is like a kitchen*) compared to novel nominal metaphors (*A mind is a kitchen*). Furthermore, using computational simulations based on the semantic space model, Utsumi (2011) found that novel meanings, which are semantically more diverse than conventional utterances, are preferentially comprehended as comparisons. Such results show that comparison mechanisms initiated when processing similes might ease novel meaning comprehension. This in turn provides support for the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), according to which novel metaphors are comprehended as comparisons. Faster RTs along with higher accuracy rates elicited in response to novel similes compared to novel nominal metaphors will therefore corroborate the previously observed results, and will suggest that comparison mechanisms evoked by similes ease novel meaning comprehension.

Furthermore, both novel nominal metaphors and novel similes are expected to elicit longer response latencies and lower accuracy rates compared to literal utterances. Such results would be in line with previous research, which has repeatedly shown that cognitive mechanisms engaged in novel meaning construction are more difficult and more time consuming relative to retrieval mechanisms involved in literal meaning comprehension (e.g., Arzouan et al. 2007; Lai et al. 2009; Rutter et al. 2012; Tang et al. 2017a; Rataj et al. 2018).

Hypothesis 2: *In the non-native language, novel similes will elicit shorter reaction times and higher accuracy rates compared to novel nominal metaphors, both of which will evoke longer response times and lower accuracy ratings compared to literal sentences.*

Previous research has not been devoted to investigating the role of metaphor conventionality in bilingual metaphor comprehension, and thus the question whether comparison mechanisms facilitate novel metaphor processing in the non-native language remains under-investigated. In line with the results observed in Experiment 1 (Chapter 4), while conventional metaphor comprehension seems to be language-dependent, the comprehension of novel metaphoric meanings might be independent of language nativeness (Jankowiak et al. 2017). Therefore, a similar pattern of results regarding novel nominal metaphor and novel simile comprehension in both the native and non-native language will confirm that, since novel metaphors are of low frequency in both languages, their comprehension is similarly taxing in L1 and L2. Consequently, such results could point to the role of the frequency of nonliteral utterances as a factor that potentially modulates metaphor comprehension in bilingualism. Additionally, it is expected that both novel nominal metaphors and novel similes will evoke longer response times and lower accuracy ratings relative to literal sentences also in the non-native tongue, for the same reasons as those already discussed in Hypothesis 1.

Hypothesis 3: *Overall, all sentence types presented in the non-native language will elicit longer reaction times compared to stimuli presented in the native tongue.*

As argued by the temporal delay assumption postulated within the Bilingual Interactive Activation Plus Model (BIA+; Dijkstra and van Heuven 2002; see section 3.2.2.2), and as previously discussed in Hypothesis 2 for Experiment 1 (see section 4.2), the activation of semantic representations is likely to be delayed in the non-dominant relative to the dominant tongue. This point has been supported in previous behavioral research on bilingual language processing, showing longer RTs for the non-native than native language (e.g., Dijkstra et al. 1998a; Dijkstra et al. 1999; de Groot et al. 2002), and has also been observed in Experiment 1 (Jankowiak et al. 2017). Overall longer reaction times for experimental sentences presented in L2 compared to L1 observed in the current study will therefore corroborate the results observed in Experiment 1 (Chapter 4), and will suggest that non-native and non-dominant language processing is slower and less automatic compared to the native and dominant tongue.

5.3. Method

5.3.1. Participants

The original sample included 31 participants, but four of them had to be excluded from analyses owing to low accuracy rates on literal or anomalous trials (below 70%). This resulted in a final sample of 27 participants (19 women, $M_{age} = 23.96$, $SD = 1.17$), who were MA students of the Faculty of English (Adam Mickiewicz University in Poznań), and who participated in the experiment for course credits. All of them were right handed ($M_{right\ hand\ preference} = 90\%$, $SD = 14.43$), as indicated by scores obtained from an online Handedness Questionnaire (Cohen 2008) based on the Edinburgh Inventory (Oldfield 1971). Participants were all native speakers of Polish (L1), and were late proficient unbalanced learners of English as their second language (L2), as they had acquired their L2 after their L1 ($M_{age\ of\ L2\ acquisition} = 9.05$, $SD = 2.91$). They used their native tongue more frequently than their non-native language, and they were thus L1-dominant. Although their L2 proficiency level was not directly tested during the experimental procedures, all participants were students of English Studies, and had passed their Practical English Language Exam, which was equivalent to the Cambridge Proficiency Examination that is administered by Cambridge University. All of the participants had normal or corrected to normal vision, and they did not suffer from any language or neurological disorder.

5.3.2. Materials

Materials used in the study consisted of 304 Polish and 304 English sentences: 76 novel similes (e.g., *Amnesia is like a rubber*), 76 novel nominal metaphoric sentences (*Amnesia is a rubber*), as well as 76 literal (*This small object is a rubber*), and 76 anomalous utterances (*A small bug is a rubber*) in each language. Each set shared the same critical word (a noun), 7 of which were Polish-English cognate words and 1 was an interlingual homograph, which, however, constituted only 5% of all critical words. The criteria under which the critical words were selected are provided in Table 12. Information regarding

mean sentence length of Polish and English experimental materials is presented in Table 13. English sentences had a larger number of words compared to Polish utterances owing to the presence of articles in English, which are not used in Polish. The experimental materials used in the study are provided in Appendix E (Polish stimuli; 263) and Appendix F (English stimuli; p. 266).

Table 12. Polish and English critical words' characteristics, including their frequency per million, number of syllables, and number of letters.

| | Frequency per million | Number of syllables | Number of letters |
|-----------------------|--|--|---------------------------------------|
| Polish critical words | 15-35/million $M = 22.65, SD = 5.68$ (National Corpus of Polish Language) | 2-3 syllables $M = 2.30, SD = 0.46$ | 5-10 letters $M = 6.39, SD = 1.33$ |
| | 15-35/million $M = 24.34, SD = 6.25$ (Corpus of Contemporary American English) | 2-3 syllables $M = 2.21, SD = 0.41$ | 4-9 letters $M = 6.71, SD = 1.22$ |

Table 13. Mean sentence length of Polish and English experimental materials.

| | Novel nominal metaphors | Novel similes | Literal sentences | Anomalous sentences |
|---------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Polish | 3-5 words $M = 3.19, SD = 0.42$ | 4-6 words $M = 4.20, SD = 0.43$ | 3-6 words $M = 4.01, SD = 0.26$ | 3 words $M = 3.00, SD = 0.00$ |
| English | 3-8 words $M = 4.76, SD = 0.96$ | 4-9 words $M = 5.79, SD = 0.97$ | 4-7 words $M = 5.35, SD = 0.72$ | 4-8 words $M = 5.87, SD = 0.88$ |

In order to ensure that the stimuli were adequate representatives of the categories ascribed to them, they were pretested using web-based surveys. Critical words were pretested in order to ensure that they were all concrete nouns. Sentences were pretested in four normative studies: cloze probability tests, as well as meaningfulness, familiarity, and metaphoricality ratings. Instructions used in the surveys are provided in Appendix I (Polish surveys; p. 275) and Appendix J (English surveys; p. 278). While Polish experimental stimuli were rated by native speakers of Polish, English materials were assessed by English native speakers. Raters whose scores were more than 3 *SDs* from the mean were

excluded from final analyses. Table 14 provides the number of raters included in the analyses together with their demographic data. The raters who took part in these pretests did not participate in the experiment proper.

Table 14. Demographic information regarding the participants of the five normative studies in each language, including the number of raters, their gender, and mean age.

| Survey type | Language | Number of raters included in the analyses | Mean age |
|------------------------------|----------|---|------------------------------|
| Critical words' concreteness | Polish | N = 20 (13 women) | $M_{age} = 21.27, SD = 3.34$ |
| | English | N = 15 (9 women) | $M_{age} = 23.93, SD = 4.87$ |
| Cloze probability | Polish | N = 105 (98 women) | $M_{age} = 22.16, SD = 1.99$ |
| | English | N = 105 (67 women) | $M_{age} = 22.50, SD = 6.71$ |
| Meaningfulness | Polish | N = 134 (121 women) | $M_{age} = 22.19, SD = 2.04$ |
| | English | N = 137 (87 women) | $M_{age} = 20.71, SD = 2.84$ |
| Familiarity | Polish | N = 103 (90 women) | $M_{age} = 21.58, SD = 2.05$ |
| | English | N = 99 (47 women) | $M_{age} = 23.96, SD = 5.74$ |
| Metaphoricity | Polish | N = 98 (89 women) | $M_{age} = 22.63, SD = 1.94$ |
| | English | N = 99 (62 women) | $M_{age} = 20.46, SD = 3.24$ |

5.3.2.1. Normative studies: Critical words

In order to ensure that all of the critical words were concrete, raters completed a web-based rating scale, in which they rated 76 concrete critical nouns together with 76 abstract filler words on a 7-point scale, where 1 represented abstract words and 7 represented concrete items. Paired samples *t*-tests showed that Polish critical words ($M = 6.06, SD = 0.55$) were assessed as more concrete than abstract filler words ($M = 2.88, SD = 0.51$), $p < .001$. Similarly, English critical words ($M = 6.38, SD = 0.59$) were rated as more concrete as compared to abstract filler words ($M = 2.67, SD = 0.69$), $p < .001$.

5.3.2.2. Normative studies: Sentences

Cloze probability tests were conducted in order to ensure that all of the critical words were not expected due to the preceding context. Raters were provided with the beginning of a sentence and were asked to write a critical word (a noun) which first came to their mind, so that the sentence would be meaningful. The final list of stimuli included only those sentences whose critical words had been elicited less than three times in the cloze probability test.

For the normative studies with rating scales on stimuli meaningfulness, familiarity, and metaphoricity, analyses of variance (*ANOVAs*) were conducted, whose results are reported below. Significance values for pairwise comparisons were corrected for multiple comparisons using the Bonferroni correction. If Mauchly's tests indicated that the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied. In such cases, the original degrees of freedom are reported with the corrected *p* value. While novel similes, novel nominal metaphors, and literal sentences were pretested with regard to their meaningfulness, familiarity, and metaphoricity, anomalous utterances were tested only on their level of meaningfulness.

To evaluate the meaningfulness of the sentences, raters assessed them on a scale from 1 (totally meaningless) to 7 (totally meaningful). The analysis conducted on Polish survey results showed a main effect of sentence type, $F(3, 390) = 641.14, p < .001, \epsilon = .827, \eta_p^2 = .831$. Pairwise comparisons further revealed that literal sentences ($M = 5.49, SE = 0.09$) were rated as more meaningful than novel similes ($M = 4.81, SE = 0.09$), $p < .001$, novel similes were rated as more meaningful than novel nominal metaphors ($M = 4.17, SE = 0.09$), $p < .001$, and novel nominal metaphors were assessed as more meaningful compared to anomalous sentences ($M = 1.84, SE = 0.06$), $p < .001$.

Similarly, the results obtained from rating scales on English stimuli revealed a main effect of sentence type, $F(3, 399) = 919.02, p < .001, \epsilon = .767, \eta_p^2 = .874$. Pairwise comparisons further showed that literal sentences ($M = 5.48, SE = 0.07$) were rated as more meaningful than novel similes ($M = 4.61, SE = 0.07$), $p < .001$, novel similes were rated as more meaningful than novel nominal metaphors ($M = 4.37, SE = 0.08$), $p < .001$, and novel nominal metaphors were assessed as more meaningful compared to anomalous sentences ($M = 1.91, SE = 0.06$), $p < .001$. Fig. 33 presents meaningfulness ratings for Polish and English experimental materials.

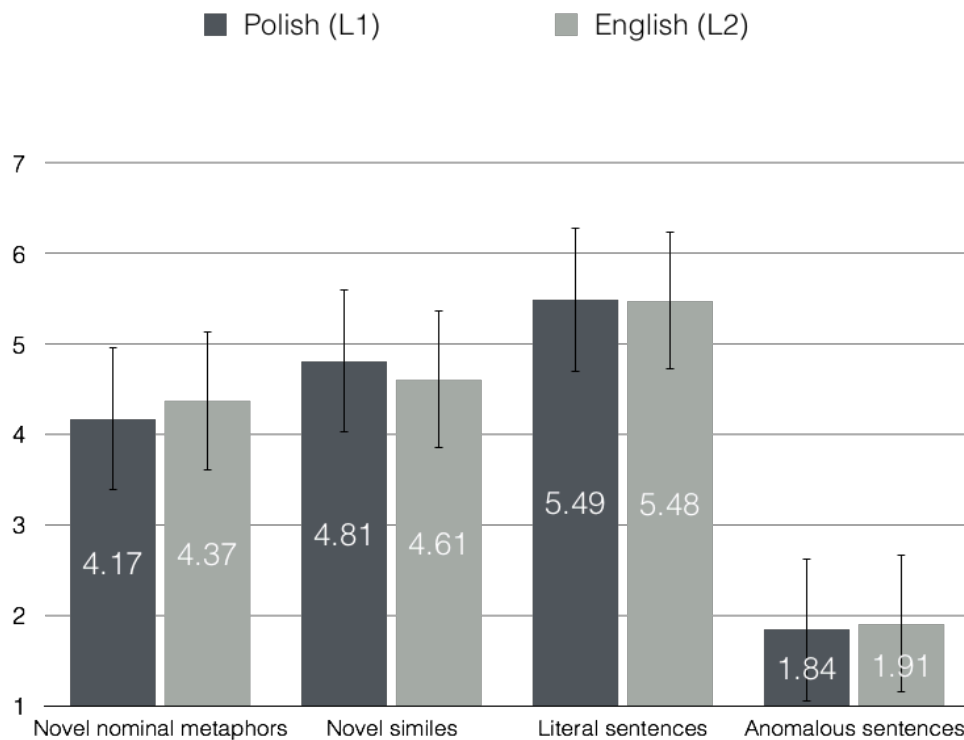


Fig. 33. Meaningfulness ratings for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, literal, and anomalous sentences (1 = totally meaningless, 7 = totally meaningful).

In order to examine the familiarity of the stimuli, raters decided how often they encountered the presented sentences on a scale from 1 (very rarely) to 7 (very frequently). The results obtained from the surveys conducted on Polish stimuli revealed a main effect of sentence type, $F(2, 200) = 40.20, p < .001, \epsilon = .618, \eta_p^2 = .287$. Pairwise comparisons confirmed that novel nominal metaphors ($M = 2.10, SE = 0.09$) were less familiar than both novel similes ($M = 2.21, SE = 0.09$), $p = .041$, and literal sentences ($M = 2.83, SE = 0.13$), $p < .001$. Furthermore, novel similes were less familiar than literal utterances, $p < .001$.

With respect to the rating scales on English stimuli, a main effect of sentence type was found, $F(2, 192) = 33.74, p < .001, \epsilon = .568, \eta_p^2 = .260$. Pairwise comparisons showed that English both novel nominal metaphors ($M = 1.47, SE = 0.05$) and novel similes ($M = 1.47, SE = 0.05$) were rated as less familiar than literal expressions ($M = 1.86, SE = 0.09$; $ps < .001$). There was no statistically significant difference between novel nominal metaphors and novel similes $p > .05$. Fig. 34 presents familiarity ratings for Polish and English experimental materials.

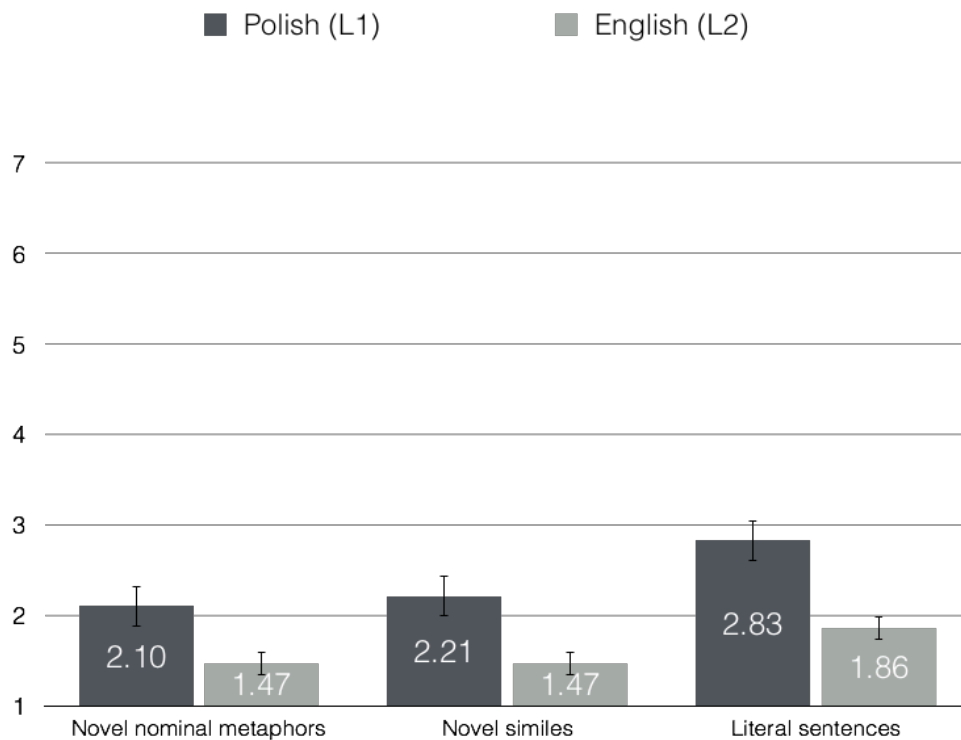


Fig. 34. Familiarity ratings for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, and literal sentences (1 = very rarely, 7 = very frequently).

In order to assess the metaphoricity of the stimuli, raters decided how metaphorical given sentences were on a scale from 1 (very literal) to 7 (very metaphorical). The analysis carried out on the results from Polish surveys showed a main effect of sentence type, $F(2, 190) = 1309.02$, $p < .001$, $\epsilon = .646$, $\eta_p^2 = .932$. Pairwise comparisons further showed that novel nominal metaphors ($M = 5.73$, $SE = 0.07$) and novel similes ($M = 5.81$, $SE = 0.08$) were judged as more metaphorical than literal sentences ($M = 1.64$, $SE = 0.05$; $ps < .001$). There was no statistically significant difference between novel similes and novel nominal metaphors, $p > .05$.

The results obtained from English surveys also revealed a main effect of sentence type, $F(2, 192) = 717.39$, $p < .001$, $\epsilon = .778$, $\eta_p^2 = .882$. Pairwise comparisons confirmed that novel similes ($M = 5.59$, $SE = 0.07$) were rated as more metaphorical than novel nominal metaphors ($M = 5.34$, $SE = 0.07$), $p = .001$, and novel nominal metaphors were rated as more metaphorical than literal sentences ($M = 2.23$, $SE = 0.06$), $p < .001$. Fig. 35 presents metaphoricity ratings for Polish and English experimental materials.

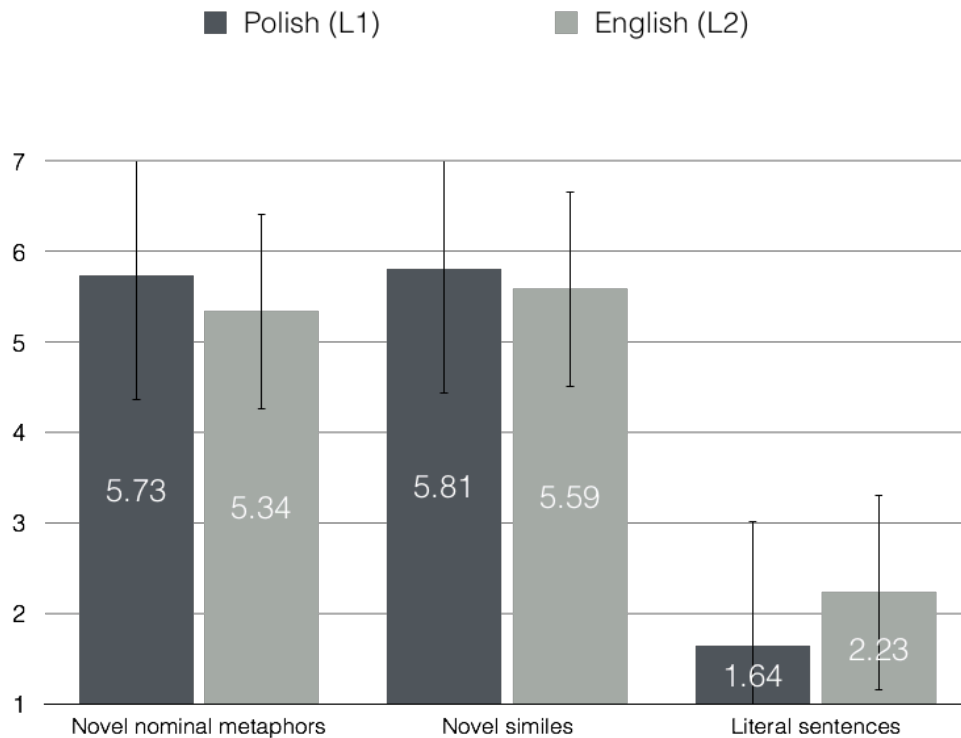


Fig. 35. Metaphoricity ratings for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, and literal sentences (1 = very literal, 7 = very metaphorical).

The aforementioned normative studies resulted in the selection of the final 304 Polish and 304 English sentences that were used in the experiment proper. The results of the normative studies for those sentences that were included in the final list are reported in Table 15. For the experiment, the stimuli were divided into 4 blocks with Polish stimuli and 4 blocks with English stimuli in order to avoid the repetition of critical words within one block. Each participant was presented with one block in Polish and one in English, whose presentation was counterbalanced. Each block consisted of 19 novel nominal metaphors, 19 novel similes, 19 literal, and 19 anomalous sentences. Additionally, 76 filler sentences were added to each block. In order to balance out the number of positive and negative responses to the utterances, 19 of the filler sentences were meaningful, and 57 were meaningless. Altogether, 304 randomized sentences were used in the experiment.

Table 15. Results of the normative tests on the stimuli included in the Experiment 2.

| | Meaningfulness (1 – totally meaningless, 7 – totally meaningful) | | Familiarity (1 – very rarely, 7 – very frequently) | | Metaphoricity (1 – very literal, 7 – very metaphorical) | |
|-------------------------|--|---------------------------|--|---------------------------|---|---------------------------|
| | Polish | English | Polish | English | Polish | English |
| Novel nominal metaphors | $M = 4.17$ $SE = 0.09$ | $M = 4.37$ $SE = 0.08$ | $M = 2.10$ $SE = 0.09$ | $M = 1.47$ $SE = 0.05$ | $M = 5.73$ $SE = 0.07$ | $M = 5.34$ $SE = 0.07$ |
| Novel similes | $M = 4.81$ $SE = 0.09$ | $M = 4.61$ $SE = 0.07$ | $M = 2.21$ $SE = 0.09$ | $M = 1.47$ $SE = 0.08$ | $M = 5.81$ $SE = 0.08$ | $M = 5.59$ $SE = 0.07$ |
| Literal sentences | $M = 5.49$ $SE = 0.09$ | $M = 5.48$ $SE = 0.07$ | $M = 2.83$ $SE = 0.13$ | $M = 1.86$ $SE = 0.09$ | $M = 1.64$ $SE = 0.05$ | $M = 2.23$ $SE = 0.06$ |
| Anomalous sentences | $M = 1.84$ $SE = 0.06$ | $M = 1.91$ $SE = 0.06$ | - | - | - | - |

5.3.3. Procedure

The procedures applied in the experiment followed ethical guidelines for studies with human participants. An informed consent form along with the information about the procedures of the experiment was provided to participants before the experiment began. The informed consent form is provided in Appendix K (p. 281).

The experiment was conducted at the Language and Communication Laboratory, at the Faculty of English, Adam Mickiewicz University in Poznań. Participants were seated in a dim and quiet testing cabin, 50 cm from the computer screen. During the experiment, the sentences were randomly presented on a computer screen using black letters, and were centered on a gray background. The sentences were presented in the following time sequence: context (2000 ms) was followed by a fixation cross (500 ms), and a critical word, which lasted until the response was given. Trials were separated by an intertrial interval (2000 ms), which was presented as a blank screen. The time sequence of stimulus presentation is provided in Fig. 36.

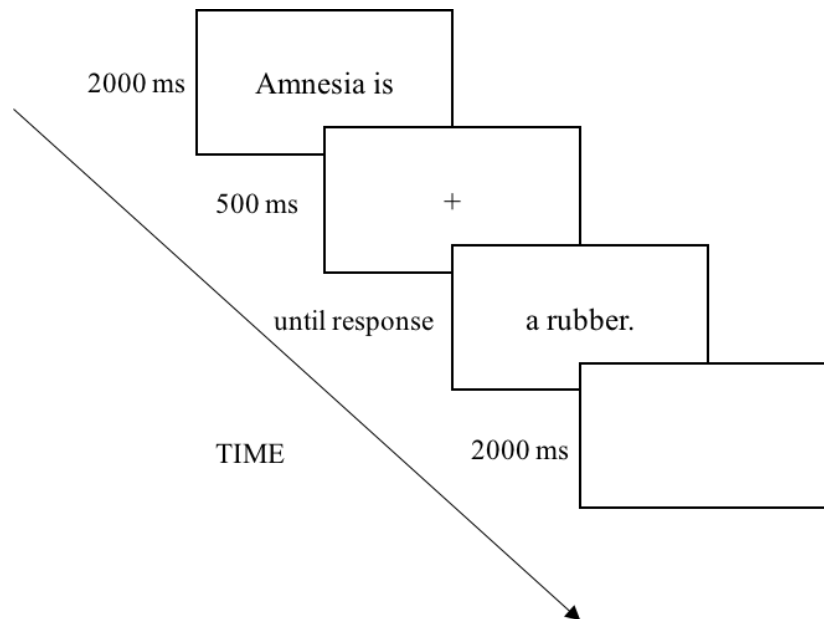


Fig. 36. Time sequence of stimulus presentation.

Participants decided whether the sentence presented on a computer screen was meaningful or meaningless by pressing a corresponding key, whose designation was counterbalanced. Prior to the experimental blocks, participants completed a practice block with 20 sentences not included in the experimental trials, in order to practice the task. Participants who began the experiment with English blocks were instructed in English, and those who started the experiment with Polish blocks were provided with Polish instructions. The aim of such a procedure was to increase the activation of the language of the first block (Grosjean 1998: 140; Marian and Spivey 2003; Canseco-Gonzales et al. 2010).

5.3.3.1. Overview of statistical analyses

Both accuracy ratings and reaction times were analyzed using 2 language (Polish/native dominant vs. English/non-native non-dominant) \times 4 sentence type (novel simile vs. novel nominal metaphor vs. literal vs. anomalous) repeated measures *ANOVAs*. In all analyses, significance values for pairwise comparisons were corrected for multiple comparisons

using the Bonferroni correction. If Mauchly's tests indicated that the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied. In such cases, the original degrees of freedom are reported with the corrected p value.

5.4. Results

5.4.1. Accuracy rates

Accuracy ratings are reported as percentage of correct responses observed in the semantic decision task. For statistical analyses, accuracy rates were arsin-transformed. The results showed an interaction between language and sentence type, $F(3, 78) = 4.56, p = .005, \eta_p^2 = .149$. Follow up analyses were carried out for each language separately. A repeated measures *ANOVA* with sentence type as factor performed on accuracy rates for Polish utterances showed a main effect of sentence type, $F(3, 78) = 51.18, p < .001, \varepsilon = .722, \eta_p^2 = .663$. Pairwise comparisons further showed that novel nominal metaphors ($M = 48.00, SE = 5.31$) elicited lower accuracy rates than novel similes ($M = 61.18, SE = 5.24$), $p = .024$, as well as than literal ($M = 88.11, SE = 1.92$), $p < .001$, and anomalous sentences ($M = 96.48, SE = 1.37$), $p < .001$. Additionally, novel similes elicited lower accuracy rates than literal and anomalous sentences, $ps < .001$, and literal sentences evoked lower accuracy rates compared to anomalous utterances, $p = .002$.

Similarly, a repeated measures *ANOVA* with sentence type as factor performed on accuracy rates for English utterances showed a main effect of sentence type, $F(3, 78) = 32.61, p < .001, \varepsilon = .725, \eta_p^2 = .556$. Pairwise comparisons further showed that novel nominal metaphors ($M = 49.30, SE = 5.48$) elicited lower accuracy rates than novel similes ($M = 69.22, SE = 3.72$), as well as than literal ($M = 84.67, SE = 2.17$), and anomalous sentences ($M = 93.00, SE = 2.03$), $ps < .001$. Additionally, novel similes elicited lower accuracy rates than both literal, $p = .006$, and anomalous utterances, $p < .001$. There was no statistically significant difference between literal and anomalous sentences ($p > .05$).

With regard to between-language differences, post-hoc tests revealed higher accuracy rates for anomalous sentences in Polish ($M = 96.48, SE = 1.37$) than in English

($M = 93.00$, $SE = 2.03$), $p = .005$. There was no statistically significant difference between Polish and English novel nominal metaphors, novel similes, as well as literal sentences ($ps > .05$).¹³

In addition to the interaction, a main effect of utterance type was found, $F(3, 78) = 52.00$, $p < .001$, $\epsilon = .728$, $\eta_p^2 = .667$. Pairwise comparisons further showed that novel nominal metaphors ($M = 48.65$, $SE = 4.91$) elicited lower accuracy rates than novel similes ($M = 65.20$, $SE = 4.21$), as well as than literal ($M = 86.39$, $SE = 1.71$), and anomalous sentences ($M = 94.74$, $SE = 1.60$), $ps < .001$. Additionally, novel similes elicited lower accuracy rates than literal and anomalous sentences, $ps < .001$, and literal sentences evoked lower accuracy rates compared to anomalous sentences, $p = .006$. Mean accuracy rates per each sentence type in each language are provided in Fig. 37.

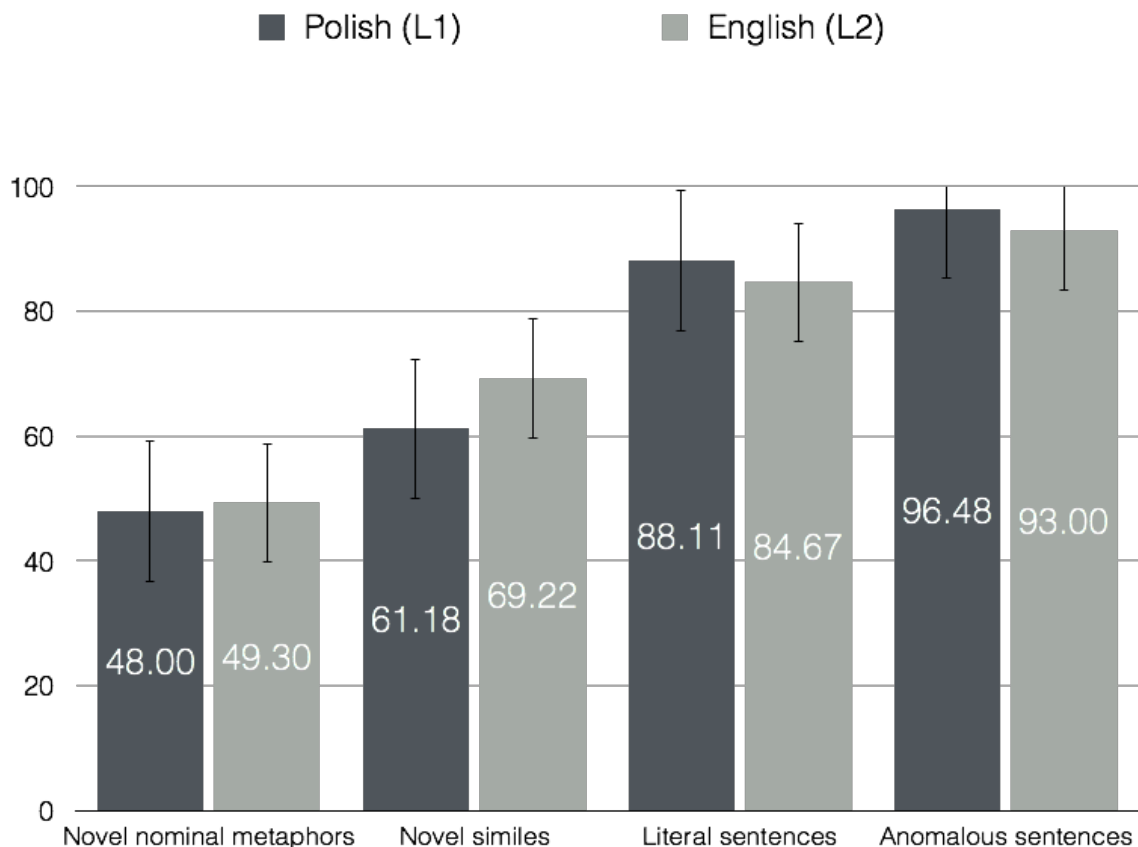


Fig. 37. Accuracy rates (%) for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, literal, and anomalous sentences.

¹³ Correction for multiple comparisons was applied here, and the critical p level for significance was set to .012.

5.4.2. Reaction times

Reaction times (RTs) were measured time-locked to critical word onset. Only correct responses were included in the statistical tests. For statistical analyses, reaction times were log-transformed. The results showed a main effect of language, $F(1, 25) = 6.39$, $p = .018$, $\eta_p^2 = .204$, with longer reaction times for English ($M = 1243.65$ ms, $SE = 83.63$) than Polish sentences ($M = 1132.30$ ms, $SE = 64.36$). Moreover, a main effect of sentence type was found, $F(3, 75) = 14.39$, $p < .001$, $\eta_p^2 = .365$. Pairwise comparisons further showed that novel nominal metaphors ($M = 1365.07$ ms, $SE = 87.84$) elicited longer reaction times than novel similes ($M = 1217.47$ ms, $SE = 78.12$), $p = .029$, as well as than literal ($M = 1104.83$ ms, $SE = 69.81$), $p < .001$, and anomalous sentences ($M = 1064.53$ ms, $SE = 76.25$), $p < .001$. There was no statistically significant difference between literal and anomalous, between novel similes and literal, and between novel similes and anomalous sentences ($p > .05$). Mean reaction times per each sentence type in each language are provided in Fig. 38.

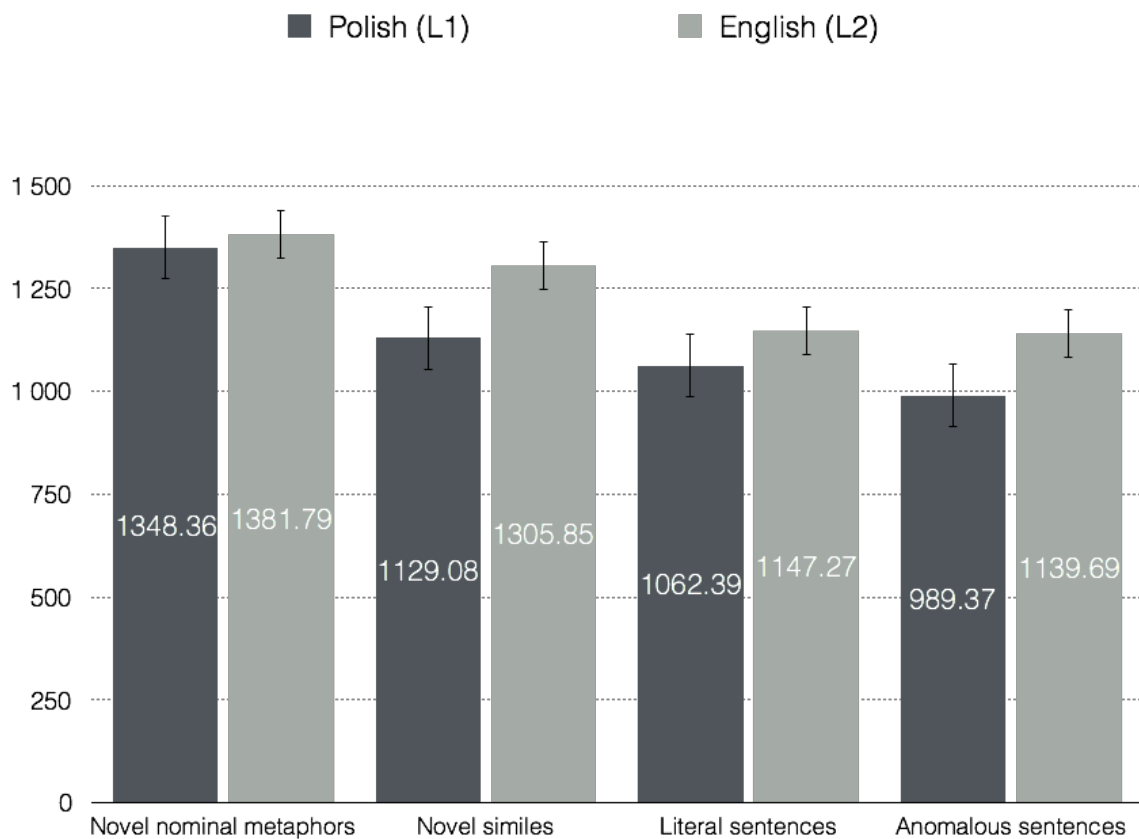


Fig. 38. Reaction times (ms) for Polish (dark grey) and English (light grey) novel nominal metaphors, novel similes, literal, and anomalous sentences.

5.4.3. Correlation analyses

Due to the fact that accuracy rates for Polish stimuli were higher than those for English utterances, correlation analyses were conducted in order to check whether the sentences were evaluated similarly by the normative study participants (native speakers of Polish and English) and the experiment proper participants (native speakers of Polish and L2 learners of English). The correlation analyses were carried out separately for each language, and were conducted on averaged values for all novel nominal metaphors, novel similes, as well as literal sentences. The results showed a strong positive correlation between accuracy rates and meaningfulness ratings for both English, $r(226) = .55, p < .001$, and Polish utterances, $r(226) = .74, p < .001$. Fig. 39 shows scatterplots representing the correlations observed for Polish and English stimuli.

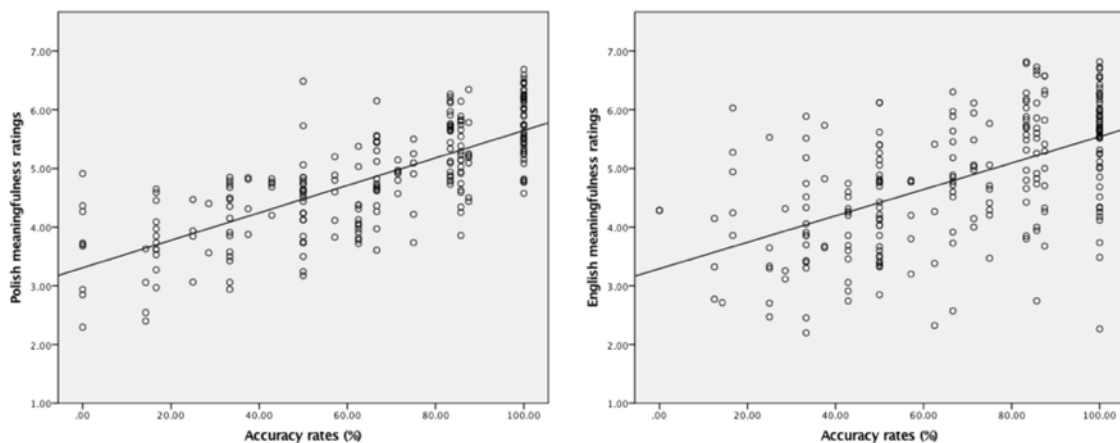


Fig. 39. Scatterplots showing the correlation between meaningfulness ratings (y-axis) and accuracy rates (x-axis) for Polish (left-hand side) and English (right-hand side) novel nominal metaphors, novel similes, and literal sentences.

5.5. Discussion and conclusion

The present experiment investigated the role of comparison form in bilingual novel metaphor comprehension. The study addressed the question whether novel similes, whose linguistic form initiates comparison mechanisms, are easier and faster to comprehend compared to novel nominal metaphors in both the native and non-native tongue. To this

end, late proficient unbalanced Polish (L1) – English (L2) bilingual speakers were presented with novel similes, novel nominal metaphors, literal, and anomalous sentences in their native and non-native language, and performed a two-choice semantic decision task.

The observed results showed shorter response times and higher accuracy rates for novel similes relative to novel nominal metaphors irrespectively of language nativeness, thus suggesting that comparison mechanisms initiated when processing similes facilitate novel metaphor comprehension in both languages. The results therefore lend support to the Career of Metaphor Model (Bowdle and Gentner 2005), and extend its assumptions regarding monolingual metaphor comprehension to the context of bilingualism, thus providing novel insights into how bilingual speakers construct novel meanings. Furthermore, the results showed language-specific effects, as overall longer response times were found for non-native relative to native sentences. Such results are in line with what Experiment 1 (Chapter 4) revealed, in which L2 utterances also evoked longer response times compared to L1 stimuli (see section 4.4.1.2). These results might be interpreted as indicative of slower and less automatic processing of L2 compared to L1, thus lending support to the BIA+ model (Dijkstra and van Heuven 2002). Chapter 6 presents a more elaborate discussion on the results obtained from Experiment 1 and Experiment 2 together with the implications and possible limitations of the two studies.

Chapter 6: Lexico-semantic processing in bilingual figurative language comprehension: General discussion

6.1. Introductory remarks

The two experiments reported in the previous chapters (Chapter 4 and Chapter 5) investigated lexico-semantic processing in bilingual figurative language comprehension. The main objective of Experiment 1 (Chapter 4) was to examine whether metaphor conventionality modulates mechanisms engaged in metaphor comprehension in both the native and non-native language, as reflected in behavioral measures (reaction times and accuracy rates) as well as event-related potentials. Since both behavioral and electrophysiological results observed in Experiment 1 suggested that mechanisms engaged in novel metaphor comprehension might be language-independent, Experiment 2 (Chapter 5) further tested this assumption by examining the role of comparison mechanisms in L1 and L2 novel meaning comprehension. The results obtained from Experiment 2 confirmed that novel meaning construction might be independent of language nativeness, as shown by higher accuracy rates and shorter reaction times for novel similes than novel nominal metaphors in both L1 and L2. This chapter provides a general discussion on the results obtained from three research methods used in the two experiments: web-based rating scales employed in the normative tests, as well as behavioral and electrophysiological measures employed in the experiments proper. The chapter discusses the observed results with reference to the formulated hypotheses, presents possible limitations of the experiments, and suggests future research directions.

6.2. Experiment 1: Novel and conventional metaphor comprehension in bilingualism

In Experiment 1 (Jankowiak et al. 2017), late proficient unbalanced Polish (L1) – English (L2) bilingual speakers performed a two-choice semantic decision task in response to novel metaphoric, conventional metaphoric, literal, and anomalous word pairs presented in their native and non-native tongue. The stimuli used in the study were pretested by means of conducting survey research with rating scales, whose aim was to assess whether the materials were perceived as intended. Section 6.2.1 provides a discussion on the results obtained from the normative tests. The statistical analyses performed on the data obtained from the experiment proper were based on behavioral (i.e., reaction time and accuracy rates) as well as electrophysiological results, which are discussed in section 6.2.2 and 6.2.3, respectively.

6.2.1. Normative studies' results in light of previous research on bilingual lexico-semantic processing

Designing the experimental materials first included the selection of appropriate critical words (Jankowiak et al. 2017). Bearing in mind the fact that visual word processing is modulated by factors such as word frequency (Balota and Spieler 1999), word concreteness (Richards 1976), number of syllables and letters (Balota et al. 2004; New et al. 2006; Chetail 2014), as well as cognate status (Lemhöfer et al. 2008; Temnikova and Nagel 2015), all of these factors were controlled for. As a result, the list of both Polish and English critical words included items which were of low frequency (14-37/million), and were abstract. In addition, the list did not include words that were either monosyllabic or had more than 3 syllables, as well as those with fewer than 4 and more than 13 letters. Out of 152 critical words (76 Polish and 76 English), only 9 items were Polish-English cognates or interlingual homographs (see Appendix C, p. 257, and Appendix D, p. 260).

Once the critical words had been selected, Polish and English word dyads were paired so that for each critical word, four conditions were designed: a novel metaphoric, conventional metaphoric, literal, and anomalous verb-noun word pair. The word dyads

were pretested on native users of a given language; namely, Polish native speakers completed surveys on Polish materials, while English native speakers evaluated English stimuli. Testing the experimental materials on native users of both languages provided a baseline that showed how the raters perceived the word pairs in their mother tongue, as a result of which the results obtained from the normative studies were not influenced by factors such as vocabulary knowledge or general language proficiency. The web-based surveys included rating scales as well as cloze probability tests. All of the rating scales involved 7-point scales with experimental stimuli counterbalanced and presented in a randomized order. The materials were pretested with a view to evaluating four aspects: predictability, meaningfulness, familiarity, and metaphoricity, all of which could modulate mechanisms engaged in meaning comprehension (De Grauwe et al. 2010). The role of meaningfulness ratings was to ensure that all of the experimental conditions, except for anomalous word pairs, were meaningful. This was of special importance with regard to novel metaphors, which are more susceptible to a large variability with respect to whether people perceive them as meaningful or meaningless. Familiarity scales were administered to ensure that the stimuli were of a similar degree of familiarity, while metaphoricity ratings were aimed to evaluate whether novel and conventional metaphors were perceived as metaphoric, while literal word dyads were assessed as literal. Additionally, cloze probability tests were administered in order to ensure that the critical items were not expected due to the context in which they were embedded.

The results obtained from the normative studies showed similar trends for both Polish and English stimuli. The only slight difference between the results obtained from the surveys on Polish and English materials emerged in familiarity ratings, in which Polish conventional metaphors were assessed as more familiar than literal word pairs. In English, on the other hand, this pattern was reversed, as higher familiarity rates were observed for literal than conventional metaphoric word dyads (Jankowiak et al. 2017).

It needs to be noted that previous research on bilingual lexico-semantic processing has not always followed such strict criteria when designing experimental materials, while all of the aforementioned factors might modulate mechanisms engaged in bilingual language comprehension. Namely, whereas some authors do not mention whether and how their stimuli were prepared, other researchers provide information only regarding the frequencies and number of letters of critical items used in the study (e.g., Proverbio et al.

2002; Midgley et al. 2009b; Newman et al. 2012). Additionally, rarely have the experimental materials been thoroughly tested before the experiment proper. For instance, Proverbio and colleagues (2002) carried out a normative test on the complexity of the sentences they used in the ERP study, which was, however, completed by only 20 raters, thus providing a decreased statistical power of the results obtained from their normative study.

Importantly, the significance of conducting normative studies on experimental stimuli is even more profound in research on metaphor comprehension, in which in order to be able to formulate valid conclusions from the observed behavioral and electrophysiological patterns, researchers need to ensure that novel metaphoric, conventional metaphoric, literal, and anomalous utterances are perceived as intended. Nonetheless, in previous studies into monolingual metaphor comprehension, researchers have not always followed such procedures when designing their experimental materials. For example, while De Grauwe et al. (2010) conducted cloze probability tests, as well as familiarity, figurativeness, and sentence plausibility ratings with 5-point rating scales, Lai et al. (2009) carried out only two pretests, in which the authors examined familiarity and interpretability of their stimuli on 3-point rating scales, which might have provided raters with an insufficient number of possible answers to choose from when rating the utterances. Furthermore, Rutter et al. (2012) did not conduct any pretests, and their stimuli selection was based only on the number of letters and frequencies of their critical items. Future studies into bilingual metaphor comprehension should therefore be more scrupulous in stimuli selection, as this might consequently increase the validity of the observed results.

6.2.2. Behavioral results

Behavioral results obtained from Experiment 1 (Jankowiak et al. 2017) included reaction times and accuracy rates from a semantic decision task. Analyses performed on these two dependent variables showed an interaction between language and utterance type, which resulted from between-language differences in response to conventional metaphoric word dyads. Namely, while novel metaphors evoked longer response times and higher accuracy rates relative to literal word pairs in both the native and non-native tongue, conventional metaphors elicited similar RTs as well as accuracy rates to literal utterances, yet only in

the native tongue. In L2, on the other hand, conventional metaphoric word dyads evoked significantly longer reaction times and lower accuracy rates relative to literal word pairs.

Therefore, Hypothesis 1, which addressed the question whether in the native language, novel metaphors would elicit longer reaction times and lower accuracy rates compared to both conventional metaphors and literal utterances, was confirmed. It thus seems that native metaphor comprehension is modulated by metaphor conventionality, as reflected in longer RTs and lower accuracy ratings for novel metaphors than both conventional metaphoric and literal word pairs. This indicates that in the native tongue, meaning construction required when processing novel metaphors is more time consuming than meaning retrieval mechanisms engaged in conventional metaphor comprehension, which is in line with previous behavioral studies on monolingual novel and conventional metaphor comprehension (e.g., Blank 1988; Gentner and Wolff 1997; Bowdle and Gentner 2005). In addition, similarly to the RT results obtained by Arzouan et al. (2007) and Tang et al. (2017a), conventional metaphors did not differ from literal utterances, thus suggesting that conventional metaphoric and literal meaning comprehension requires a similar amount of time.

Hypothesis 2, stating that in the non-native language, novel and conventional metaphors would elicit similar reaction times and accuracy rates, both of which would evoke longer response times than literal word pairs, was, on the other hand, only partially supported. Namely, contrary to what was predicted, L2 novel metaphors elicited significantly lower accuracy rates compared to L2 conventional metaphors, which indicates that participants found it more difficult to categorize L2 novel metaphors as meaningful compared to L2 conventional metaphoric word dyads. Nevertheless, in agreement with what was expected, no differences in reaction times were observed for conventional and novel metaphoric word pairs presented in the non-native language. Similar reaction times for novel and conventional metaphoric word pairs in L2 might therefore indicate that bilingual speakers are less sensitive to the levels of metaphor conventionality in their non-native tongue. Furthermore, L2 novel and conventional metaphor evoked both longer response times and lower accuracy ratings than literal word pairs, thus suggesting that L2 metaphors, irrespectively of their level of conventionality, require more time to comprehend compared to literal meanings.

However, it might be argued that the observed between-language differences to conventional metaphors may have been influenced by the experimental materials. Therefore, in order to ensure that the experimental materials were adequate representatives of categories ascribed to them, correlation analyses were performed, in which meaningfulness ratings for Polish and English novel metaphoric, conventional metaphoric, and literal word dyads were correlated with accuracy rates observed in response to those stimuli. The results showed a strong positive correlation between the two variables, thus confirming that the materials were perceived similarly by the raters from the normative studies and by EEG participants. Such results are of special importance with regard to English stimuli due to the fact that English native speakers assessed their meaningfulness level in the normative tests, but English L2 learners performed a semantic decision task in the experiment proper.

In addition, a follow-up test on English stimuli familiarity was conducted, in which English L2 learners evaluated the level of familiarity of English novel metaphoric, conventional metaphoric, and literal word pairs. A correlation analysis was carried out to see whether the familiarity of English word dyads was evaluated similarly by the raters from the pre-experimental normative study (native speakers of English) and L2 learners of English. The results showed a strong positive correlation between how the two groups of participants rated the familiarity of English word pairs. Thus, the results obtained from the correlation analyses show that both native speakers and L2 learners of English made similar decisions regarding both the meaningfulness and familiarity of the experimental word dyads.

Apart from the differences observed in response to specific utterance types in Polish and English, behavioral results also yielded a general language effect, with overall longer response times and lower accuracy rates for English than Polish stimuli, therefore confirming Hypothesis 3. Previous behavioral studies into bilingual semantic processing have frequently revealed longer response times for non-native relative to native items (e.g., Dijkstra et al. 1998a; Dijkstra et al. 1999; de Groot et al. 2002), and indicated that semantic processing is more time consuming in the non-native tongue. In light of the present results, longer RTs for L2 than L1 might have resulted from the fact that the participants of the present study were all L1-dominant, and thus they used their L2 less frequently than their L1. The observed results thus lend support to the temporal delay assumption postulated within the BIA+ model (Dijkstra and van Heuven 2002; see section

3.2.2.2), according to which L2 semantic information is processed slower and less automatically when bilingual speakers are less dominant in their L2 (Jankowiak et al. 2017).

6.2.3. Electrophysiological results

Although reaction time measures are informative in showing how much time participants need in order to comprehend an utterance, event-related brain potentials can reveal information regarding brain activity that directly follows the presentation of a stimulus, with ERP components reflecting specific cognitive mechanisms engaged in its comprehension. Furthermore, ERP patterns observed within a particular time frame provide information regarding a specific stage of language processing even when participants are not instructed to perform an overt task, whereas RTs reflect all processes underlying language comprehension as well as mechanisms related to decision making (Luck 2005: 21). Previous research has, nonetheless, been rarely devoted to studying bilingual metaphor comprehension by employing an ERP method, and thus the observed electrophysiological results provide novel insights into how the bilingual brain computes novel and conventional metaphors in the native and non-native tongue. This section is devoted to providing interpretations concerning three language-related ERP components that were analyzed in the experiment: the P200, N400, and LPC. The following part of this section first focuses on the N400 and LPC effects, as the main hypotheses were formulated with regard to these two components. Due to the fact that the aims of the present experiment did not include any a priori hypotheses regarding the P200 component, the P200 effect is discussed at the end of this section.

Hypothesis 4 predicted that in both the native and non-native tongue, a linear effect across the utterance types would be observed within the N400 time frame. Such predictions were fully confirmed, as within the late N400 time window (400-500 ms), the most pronounced N400 amplitudes were elicited in response to anomalous, followed by novel metaphoric, conventional metaphoric word dyads, and finally the lowest N400 amplitudes were observed for literal word pairs, irrespectively of language nativeness. Such a graded effect is in agreement with previous monolingual research on novel and conventional metaphor comprehension (Arzouan et al. 2007; Lai et al. 2009). For instance, Arzouan et al. (2007) presented their participants with novel, conventional, semantically

related, and semantically unrelated word pairs, and also found the N400 amplitudes varying as a function of utterance type, with the most robust N400 response to semantically unrelated, followed by novel metaphoric, conventional metaphoric, and finally semantically related word pairs. Such results were interpreted by the researchers as indicative of a larger resource intensity when processing novel metaphoric relative to conventional metaphoric word dyads, therefore providing support for the assumption that mechanisms engaged in metaphor comprehension are influenced by metaphor conventionality (Bowdle and Gentner 2005). Importantly, the present results extend this assumption from the monolingual to the bilingual context, and suggest that metaphor conventionality modulates the amount of information that needs to be retrieved from the semantic memory network, as reflected in the N400 response, in both the native and non-native language. The obtained results might therefore indicate that in both languages, novel metaphor comprehension, which involves sense creation operations, is more resource intensive compared to conventional metaphor comprehension, which requires only sense retrieval mechanisms. Importantly, in both languages, conventional metaphors elicited more pronounced N400 amplitudes than literal utterances, which might suggest that conventional metaphors still require mapping processes that are more taxing compared to literal meaning comprehension (Lai and Curran 2013).

A different pattern of results was, however, observed within the LPC time frame (500-800 ms), where mechanisms engaged in novel and conventional metaphoric meanings comprehension seemed to be modulated by language nativeness, therefore providing only a partial support for Hypothesis 5. Even though it was predicted that in both languages, novel metaphoric and anomalous word pairs would elicit attenuated LPC amplitudes compared to conventional metaphoric and literal word dyads, novel metaphors evoked reduced LPC amplitudes compared to both conventional metaphoric and literal word pairs only in the native tongue. In the non-native language, on the other hand, novel and conventional metaphors converged, and both elicited reduced LPC amplitudes compared to literal utterances.

Attenuated LPC amplitudes for novel metaphors relative to literal utterances are difficult to account for within the functional role of the LPC response, according to which the component indexes processes engaged in meaning re-analyses or additional working memory load (Brouwer et al. 2012: 137), and thus novel metaphoric word pairs should have evoked larger LPC amplitudes than literal word pairs. A reduced LPC response to

novel metaphoric word dyads was, however, previously observed by Arzouan et al. (2007), who interpreted such results as reflecting sustained negativity overlapping with the LPC effect. In studies on monolingual semantic processing, sustained negativity has been linked to processes involved in semantic integration of conceptually challenging stimuli, and to mechanisms required when reinterpreting a meaning after an initial failure in its comprehension (Ruchkin et al. 1988; Anderson et al. 1996; Oberauer et al. 2001; Jiang et al. 2009; Rataj 2014; Tang et al. 2017a; Tang et al. 2017b; Rataj et al. 2018). Thus, a reduced LPC response to novel metaphors relative to literal utterances might be indicative of the ongoing difficulty of final meaning integration or/and access to the non-literal route during novel metaphor comprehension (Arzouan et al. 2007; Goldstein et al. 2012; Tang et al. 2017a; Rataj et al. 2018). Following such interpretations, reduced LPC amplitudes for novel metaphoric word pairs in the native language, and for novel and conventional word dyads in the non-native language observed in the present study might point to the ongoing effort in meaning integration and/or of the access to the non-literal route during novel metaphor comprehension in L1, and during novel and conventional comprehension in L2.¹⁴

Importantly, between-language differences found in response to conventional metaphors within the LPC time window corroborate the behavioral results observed in the present study, showing that conventionalized metaphors might be more difficult to comprehend in the non-native compared to the native language. Longer reaction times together with attenuated LPC amplitudes for conventional metaphors in L2 might be also in line with previous research on the LPC response, showing that the component is sensitive to decision-related factors, including response conflict, post-decision closure, confidence, or/and accuracy (Stuss et al. 1992; Johnson et al. 1998; Finnigan et al. 2002; Wolk et al. 2006; Daltrozzo et al. 2012). For instance, a positive correlation has previously been observed between reaction times and LPC peak latency, thus indicating that the LPC might reflect not only stimulus-related, but also response-related processes (e.g., Ritter et al. 1972; Ford et al. 1973; Kutas et al. 1977; Friedman et al. 1978). Future research is, however, needed in order to further examine the relationship between response latencies and the LPC magnitude in bilingual semantic processing.

¹⁴ A more in-depth discussion on novel and conventional metaphor comprehension in bilingualism is provided in section 6.4.1.

In addition to testing the role of metaphor conventionality in bilingual metaphoric language comprehension, Experiment 1 also addressed the question of the automaticity of L1 and L2 lexico-semantic processing. Partially in line with Hypothesis 6, which predicted between-language differences within the N400 and LPC time frames, materials presented in the non-native language elicited attenuated and delayed effects within the N400 time frame compared to stimuli presented in the native tongue, which was irrespective of utterance type. Reduced N400 amplitudes in response to L2 relative to L1 have been observed in previous ERP studies on semantic processing in bilingualism (e.g., Proverbio et al. 2002; Moreno et al. 2008; Midgley et al. 2009b; Newman et al. 2012), and have been interpreted as indicative of a decreased semantic interconnectivity for L2 compared to L1 lexical items within the memory system. Such a weaker interconnectivity might in turn evoke decreased activity in long-term memory, and consequently a reduced N400 response. This interpretation is therefore in line with the functional role of the N400, according to which the component indexes the amount of information retrieved from long-term memory (Kutas and Federmeier 2000; Kutas and Federmeier 2011; Kotz et al. 2012).

A 20 ms delay observed in the N400 peak latency to L2 relative to L1 is also in agreement with previous research on the N400 effect in bilingual semantic processing (e.g., Weber-Fox and Neville 1996; Phillips et al. 2004; Moreno and Kutas 2005; Braunstein et al. 2012). The obtained results might be interpreted as in line with the temporal delay assumption postulated within the BIA+ model (Dijkstra and van Heuven 2002; see section 3.2.2.2), which holds that the activation of semantic representations is delayed in the non-native and non-dominant language. The BIA+ model further claims that a delayed activation of L2 semantic information results from a lower subjective frequency of L2 items when bilinguals are not dominant in their L2. Nonetheless, the subjective frequency of the materials included in the current study was not tested, and thus more research is needed in order to further show whether a delay in the N400 peak latency might be accounted for by differences in the perceived frequency of L1 and L2 lexical items.

Importantly, no between-language differences were observed with regard to the latency of the LPC response, which might indicate that mechanisms engaged in meaning integration, as indexed by the LPC amplitudes, are not delayed in the non-native language. Overall, such results seem to suggest that even though the activation of the semantic information is slower in L2 relative to L1, as reflected in a delayed N400 peak latency,

meaning integration might follow a similar time course in L1 and L2. It needs, however, to be noted that between-languages differences were observed with regard to the magnitude of the LPC response to conventional metaphors in L1 and L2, thus suggesting that the component is modulated by the level of semantic complexity in both the native and non-native language.

Finally, although no specific hypotheses were formulated with regard to the P200, language-specific differences were observed as soon as between 150-250 ms post stimulus onset, where non-native materials elicited more robust P200 amplitudes compared to native stimuli. In previous monolingual research on language processing, the P200 response has been linked to mechanisms related to attention, short-term memory, early item encoding, as well as lexical expectancy (e.g., Luck and Hillyard 1994; Federmeier and Kutas 2002; Barber et al. 2004; Federmeier et al. 2005; Freuenberg et al. 2007; Wlotko and Federmeier 2007), indicating that the P200 component indexes processes engaged in early lexical access (Sereno et al. 1998). In line with such an assumption, previous studies have reported larger P200 amplitudes for low than high frequency words, which suggests an increased load when processing lexical items of low frequency (Assadollahi and Pulvermüller 2001; Hauk et al. 2004; Dambacher et al. 2006). In light of the observed results, larger P200 amplitudes for the non-native relative to the native language might have resulted from a lower subjective frequency of L2 than L1 lexical items. Following such an interpretation, it might be argued that the effects observed within both the P200 and N400 time frame, where a delayed N400 peak latency was observed for L2 relative to L1, result from similar between-language differences regarding the perceived frequency of native and non-native lexical items.¹⁵

Taken together, the results obtained from Experiment 1 (Jankowiak et al. 2017) indicate that lexico-semantic access, as indexed by the N400 amplitudes, is modulated by metaphor conventionality in both the native and non-native tongue. In contrast, meaning re-analyses as well as semantic integration mechanisms, as reflected in the LPC response, seem to be influenced by language nativeness, as conventional and novel metaphors elicited sustained negativity relative to literal word pairs, yet only in the non-native language. In the native tongue, on the other hand, only novel metaphoric word dyads evoked ongoing negativity. Such results might indicate increased effort related to conventional metaphoric meaning integration in L2 or/and access to the nonliteral route during conventional

¹⁵ A more in-depth discussion on bilingual semantic processing is provided in section 6.4.2.

metaphor comprehension in the non-native, but not in the native tongue. The results observed within the LPC time frame are therefore in accordance with the behavioral results obtained from the present study, as conventional metaphors elicited longer RTs along with lower accuracy rates in the non-native compared to the native tongue. In addition to utterance-specific differences, between-language differences were found within the P200 as well as N400 time windows, indicating a decreased subjective frequency of L2 items, as well as weaker semantic interconnectivity for L2 compared to L1 lexical items within the memory system (Jankowiak et al. 2017).

6.2.4. Limitations of the study

The present experiment poses a few limitations, most of which result from a potentially compromised ecological validity of the study. First of all, due to the fact that the word pairs were embedded in minimal contexts, it might be argued that the study did not provide insights into natural language processing. Namely, human communication is embedded in the context, which can be verbal, non-verbal, or both. Additionally, previous research on monolingual metaphor comprehension has repeatedly shown that context influences mechanisms engaged in metaphoric meaning processing, with highly constraining contexts facilitating metaphor comprehension, and low constraining contexts impairing it (e.g., Fischler and Bloom 1985; Pynte et al. 1996; Katz and Ferretti 2001; Pickering and Frisson 2001; Ivanko and Pexman 2003; Bambini et al. 2016). Presenting participants with word dyads without any additional contextual cues to rely on might have hindered meaning comprehension, and consequently modulated brain activity evoked in response to nonliteral utterances. Word pairs were also used by Arzouan et al. (2007), who tested novel metaphoric, conventional metaphoric, semantically related, and semantically unrelated utterances, and found very similar results to those observed in the present experiment, with sustained negativity elicited in response to novel metaphoric word dyads, thus indicating the ongoing difficulty of meaning integration and access to the nonliteral route during novel metaphor comprehension. Importantly, it needs to be noted that such a minimal context resulted in a reduced working memory load, as working memory storage was limited to holding the meaning of prime and critical words only, and thus the

obtained results were not likely to be influenced by cognitive effort related to extended working memory load.

Furthermore, participants performed a two-choice semantic decision task, and the nature of this particular task might have influenced the obtained results. Being asked to decide whether a presented word dyad was meaningful or not, participants had to attend more closely to the semantic properties of an utterance, and thus they might have looked for a potential meaning of a stimulus more intensively than in natural language comprehension. In addition, a two-choice semantic decision task might seem disadvantageous when employed in studies on novel and conventional metaphor comprehension, mostly due to the fact that novel metaphoric meanings usually fall in the middle of the literal-metaphoric meaning continuum. Consequently, being instructed to make a binary decision regarding the semantic properties of the stimuli, participants might have found it difficult to decide if a novel metaphor was meaningful or not. To address this concern, Lai et al. (2009) as well as Lai and Curran (2013) employed a multiple-choice procedure, in which participants decided whether the presented metaphoric utterances had perfect sense, some sense, little sense, and no sense. As a result of using such a procedure, the authors could investigate the degree of meaningfulness of the presented stimuli. Future research on bilingual metaphor comprehension should thus examine how task requirements modulate metaphor processing in the context of bilingualism.

Additionally, it needs to be noted that though the stimuli had been rigorously pre-tested in the normative studies, they had not been controlled for in terms of whether or not conventional metaphoric utterances included Polish-English translation equivalents. To address this issue, a follow-up analysis was performed on English conventional metaphors, due to the observed differences between L1 and L2 conventional metaphors in both behavioral and ERP patterns. The follow-up analysis first involved translating the English conventional metaphoric word pairs into Polish, after which the National Dictionary of Polish Language was consulted in order to check whether the translated word pairs were listed in the dictionary. The analysis showed that out of 76 English conventional metaphoric word pairs, 14 were English-Polish translation equivalents, which, however, constituted only 18.42% of all English conventional metaphors. Therefore, it is unlikely that the obtained results were influenced by the presence of Polish-English translation equivalents in the list of experimental materials used in the study.

Finally, one might argue that the technical requirements of any EEG experiment pose a threat to the ecological validity of the study. Due to the fact that an ERP study requires a vast number of trials when averaging, the present experiment might have been quite wearying for participants, as it took around 90 minutes. Additionally, before the actual experiment began, each participant underwent the procedure of gel application, and thus the experiment together with any required procedures took around 150 minutes. It must also be noted that EEG experiments require participants to restrict their movement, and to control their blinking during the experiment proper, which makes it strenuous and susceptible to participants becoming tired.

Interestingly, the results observed in Experiment 1 seem to suggest that novel metaphor comprehension is independent of language nativeness, and thus Experiment 2 further tested this assumption. To this end, Experiment 2 examined the comprehension of novel meanings in L1 and L2 that were presented in two linguistic forms: as novel nominal metaphors (*A is B*) and novel similes (*A is like B*). In line with the Career of Metaphor Model (Bowdle and Gentner 2005), novel meanings are claimed to be easier to comprehend when presented as a simile (e.g., *Amnesia is like a rubber*) than as a nominal metaphor (*Amnesia is a rubber*). Experiment 2 was conducted to test this hypothesis in the context of bilingualism, and thus to further show whether and to what extent the comprehension of semantically complex meanings is modulated by language nativeness. Additionally, instead of using word dyads embedded in minimal contexts, the materials used in Experiment 2 involved sentences, thus providing better insights into natural language processing.

6.3. Experiment 2: Novel nominal metaphor and novel simile comprehension in bilingualism

In Experiment 2, late proficient unbalanced Polish (L1) – English (L2) bilingual speakers performed a two-choice semantic decision task in response to novel nominal metaphoric sentences, novel similes, as well as literal and anomalous sentences presented in their native and non-native tongue. The stimuli used in the study were pretested by means of conducting survey research with rating scales, whose aim was to assess whether the materials were perceived as intended. Section 6.3.1 provides a discussion on the results from

the normative tests. The statistical analyses performed on the data obtained from the experiment proper were based on behavioral results (i.e., accuracy ratings and reaction times), which are discussed in section 6.3.2.

6.3.1. Normative studies' results in light of previous research on bilingual lexico-semantic processing

Similarly to Experiment 1, designing the experimental materials first included the selection of appropriate critical words, which were controlled for in terms of their frequency, level of concreteness, number of syllables and letters, as well as cognate status. As a result, the list of both Polish and English critical words involved items which were of relatively low frequency (15-35/million), and were all concrete. Furthermore, the list did not include words that were either monosyllabic or had more than 3 syllables, as well as those with fewer than 4 and more than 10 letters. Out of 152 critical words (76 Polish and 76 English), only 8 items were Polish-English cognates or interlingual homographs (see Appendix E, p. 263, and Appendix F, p. 266).

Once the critical words had been selected, Polish and English sentences were prepared so that for each critical word, four conditions were designed: novel nominal metaphors, novel similes, as well as literal and anomalous sentences. Importantly, nominal metaphors and similes shared the same target and source domain, and they differed only in their linguistic form (i.e., *A is B* vs. *A is like B*). The sentences were pretested on native users of a given language, and thus Polish native speakers completed surveys on Polish materials, while English native speakers assessed English stimuli. The web-based surveys involved rating scales as well as cloze probability tests, all of which were distributed online. All of the rating scales included 7-point rating scales, in which the experimental stimuli were counterbalanced and presented in a randomized order. The materials were pretested with a view to evaluating four aspects: predictability, meaningfulness, familiarity, and metaphoricity, all of which could modulate mechanisms engaged in meaning comprehension (De Grauwe et al. 2010). Meaningfulness ratings were aimed to ensure that all of the experimental conditions, except for anomalous utterances, were meaningful. This was of special importance with regard to novel nominal metaphors and novel similes, which are more susceptible to a large variability with regard to whether they are perceived

as meaningful or meaningless. Familiarity scales were administered to ensure that all of the stimuli were of a similar degree of familiarity, whereas metaphoricity ratings were aimed to evaluate whether novel similes and novel nominal metaphors were both perceived as highly metaphoric, and literal sentences were assessed as literal. Additionally, cloze probability tests were administered to ensure that the critical items were not expected due to the preceding context. The results obtained from the normative studies showed similar trends for both Polish and English stimuli.

6.3.2. Behavioral results

Behavioral results obtained from Experiment 2 included reaction times as well as accuracy rates. Analyses performed on these two dependent variables showed a main effect of sentence type and a main effect of language. Namely, the results revealed that novel nominal metaphors elicited longer response times along with lower accuracy rates compared to novel similes in both the native and non-native tongue. At the same time, the results showed a general language effect, with longer RTs and lower accuracy rates for English (L2) than Polish (L1) stimuli, irrespectively of sentence type.

Therefore, the present results provide support for Hypothesis 1, as novel similes presented in the native language elicited shorter reaction times and higher accuracy rates compared to novel nominal metaphors. Such results indicate that comparison mechanisms initiated by similes facilitate novel metaphor comprehension, and thus a simile form might ease novel meaning construction. This in turn provides support for the Career of Metaphor Model (Bowdle and Gentner 2005; see section 2.3.6), which claims that novel metaphors are understood as comparisons. In line with the model, a vehicle of a novel metaphor involves only a literal reference, and thus novel metaphor comprehension requires the process of comparison between the target and the literal meaning of the base term (Bowdle and Gentner 2005; Zharikov and Gentner 2002; Utsumi 2011). Such results are in line with a monolingual study conducted by Bowdle and Gentner (2005), who employed a self-paced reading task, and found shorter response latencies for novel comparisons compared to novel categorizations.

Importantly, the present results showed the same pattern of results in both the native and non-native language, therefore providing support for Hypothesis 2, and suggesting that mechanisms engaged in novel metaphor processing are similar in both the native and non-native tongue. Consequently, the observed results seem to confirm the conclusions drawn from Experiment 1 (Jankowiak et al. 2017), and indicate that similar cognitive mechanisms are engaged in novel meaning comprehension in both languages.¹⁶

Interestingly, contrary to the hypotheses, there were no statistically significant differences between RTs for novel similes and literal sentences in both L1 and L2. Such results are difficult to account for, as previous research has repeatedly shown longer response times for novel compared to literal meanings (e.g., Arzouan et al. 2007; Lai et al. 2009; Rutter et al. 2012; Tang et al. 2017a; Rataj et al. 2018). Taking into account the fact that RT patterns are argued to reflect linguistic complexity, with longer RTs indicating an increased processing difficulty (Kaiser 2013: 137; see section 1.3), it might be argued that similar reaction times for novel similes and literal sentences point to a similar processing difficulty. Nonetheless, differences between novel similes and literal utterances were found in accuracy rates, which showed that novel similes were more challenging to classify as meaningful compared to literal sentences. Therefore, since response times reflect all processes underlying language comprehension together with mechanisms related to decision making, future electrophysiological research might turn out advantageous when further examining potential differences between novel simile and literal meaning comprehension.

Finally, Hypothesis 3 was supported, as the experimental materials presented in the non-native language elicited overall longer reaction times along with lower accuracy rates compared to the stimuli presented in the native tongue. Such results are in accordance with those observed in Experiment 1 (see section 6.2.2), and might be interpreted as in agreement with the temporal delay assumption postulated within the Bilingual Interactive Activation Plus Model (BIA+; Dijkstra and van Heuven 2002; see section 3.2.2.2), according to which the activation of semantic representation is delayed in the non-native and non-dominant relative to the native and dominant language. Such a temporal delay is argued to result from a lower subjective frequency of L2 lexical items, possibly due to differences in language dominance. Even though the subjective frequency of the stimuli

¹⁶ A more in-depth discussion on novel metaphor comprehension in bilingualism is provided in section 6.4.1.

was not examined in the present experiment, participants were L1-dominant, and they used their L2 less frequently than their L1. As a result, longer response times for L2 compared to L1 materials might have been due to such a lower perceived frequency of lexical items presented in L2.¹⁷

6.3.3. Limitations of the study

It needs to be noted that the present study suffers from some limitations, most of which are related to the nature of the experimental stimuli. Namely, since novel similes involved the comparison form “*like*”, this might have influenced lexico-semantic expectations that participants generated. Consequently, faster response times along with higher accuracy rates for novel similes relative to novel nominal metaphors might have resulted from the fact that participants found it easier to generate accurate expectations with regard to the critical word of a simile. Although meaning expectancy was pretested in cloze probability tests, which showed that the critical words of both novel nominal metaphors and novel similes were not expected due to the preceding contexts, it remains unclear whether anticipation mechanisms to the critical items of novel similes could be somewhat facilitated by the comparison form. Future studies should therefore investigate whether and to what extent a linguistic form can impact meaning expectancy.

In addition, except for similes, the stimuli included categorical statements, which, however, slightly differed in terms of their morphosyntactic forms in Polish and English (i.e., Polish: *A to B*; English: *A is B*). Namely, although in both languages, predicational copular clauses were used, English statements involved a verbal copula *to be*, which was explicitly stated, while Polish utterances included a pronominal copula, which was dropped (Bondaruk 2014: 333). If the verbal copula had been used in Polish, the statements would have differed morphosyntactically from English sentences to a larger extent due to the fact that sentences with a verbal copula *to be* (i.e., *A jest B*) mark the predicate for instrumental in Polish, but for nominative in English (i.e., *A is B*). Importantly, Polish is a highly inflected language, and instrumental is by far more structurally complex compared to nominative. Therefore, in order for the critical words in Polish to be shared across

¹⁷ A more in-depth discussion on bilingual semantic processing is provided in section 6.4.2.

the conditions with no change in how they were inflected, the predicates needed to be marked for nominative by means of using a pronominal copula (i.e., *A to B*).

Furthermore, the results obtained from Experiment 2 might have been confounded by the fact that the level of meaningfulness of the experimental stimuli was evaluated by English native speakers in the norming tests, and by English L2 learners in the experiment proper. In order to address this potential confound, correlation analyses were performed between the meaningfulness ratings for Polish and English novel nominal metaphors, novel similes, and literal sentences and accuracy rates observed in response to those stimuli, which showed a strong positive correlation between the two dependent variables. Such results confirm that the participants of the normative studies and those of the experiment proper made similar decisions regarding the meaningfulness of the English stimuli.

Importantly, Experiment 2 involved experimental materials that were embedded in sentence context, and thus meaning comprehension could resemble natural language processing to a larger extent than in Experiment 1. Nonetheless, sentence context might have involved cognitive mechanisms that were more effortful due to an extended working memory load necessary when comprehending nonliteral meanings embedded in sentences. Namely, sentence comprehension requires parsing mechanisms, which involves assigning semantic and syntactic properties of individual words of a sentence as well as establishing the syntactic relationship between them, which in turn enables meaning comprehension (Kempen 1998: 213). The parsing mechanism operates automatically; yet, it might be hindered when a sentence is relatively long, as working memory load becomes higher with an introduction of any new sentence constituent. Thus, since the sentences used in the present experiment had 3-9 words, they might have brought about an extended storage load compared to word pairs that were used in Experiment 1.

Finally, it might be argued that the experimental nature of the study influenced language processing, as it took place in a laboratory setting, and might have impacted participants' behavior, which consequently might have not reflected natural language processing. Also, a two-choice semantic decision task and its requirement might have influenced the obtained results similarly as in Experiment 1, as participants might have attended more closely to the semantic properties of an utterance than in natural language comprehension. Additionally, they might have found it difficult to perform a binary decision to novel meanings, which may have sometimes been perceived as neither fully meaningful nor meaningless.

6.4. Figurative language comprehension in bilingualism: Insights from Experiment 1 and Experiment 2

The main objective of the two experiments described in this thesis was to examine mechanisms engaged in the comprehension of novel and conventional metaphoric meanings in the native and non-native tongue. Experiment 1 (Jankowiak et al. 2017) investigated whether metaphor conventionality modulates metaphor comprehension in both languages, and showed that while conventional metaphoric meanings might be more difficult to comprehend in L2 than in L1, novel metaphor comprehension seems to be similarly resource intensive in both languages. Experiment 2 further addressed the question whether and to what extent novel metaphor comprehension might be language-independent, and confirmed that similar mechanisms might be engaged in novel meaning construction in L1 and L2. This section aims to elaborate on the results observed in the two experiments. The following subsections focus on how bilingual speakers compute novel and conventional metaphors in their native and non-native language (section 6.4.1), and present conclusions regarding bilingual lexico-semantic access and meaning integration (section 6.4.2).

6.4.1. Novel and conventional metaphor comprehension in bilingualism

The results obtained from the two experiments described in this dissertation show that metaphor conventionality modulates metaphoric meaning comprehension in both the native and non-native language, yet to a different degree. Namely, conventional metaphor processing seems to be dependent on language nativeness, as Experiment 1 (Jankowiak et al. 2017) showed that conventional metaphors elicited longer response times, lower accuracy rates as well as attenuated LPC amplitudes in L2 relative to L1. The processing of novel metaphoric meanings might, on the other hand, be language-independent, as reflected in no between-language differences in reaction times, accuracy rates, as well as ERP results to L1 and L2 novel metaphoric word dyads (Jankowiak et al. 2017). Importantly, such an assumption was supported in Experiment 2, where the comparison form facilitated novel meaning comprehension in both the native and non-native language, as

reflected in no between-language differences in RT patterns for novel similes and novel nominal metaphors.

The results observed in Experiment 1, which showed between-language differences in response to conventional metaphors, are difficult to account for due to the fact that previous research has not been devoted to studying novel and conventional metaphor comprehension in the context of bilingualism. Previous behavioral studies into the comprehension of other types of figurative utterances in bilingualism have generally revealed a decreased sensitivity to nonliteral meanings presented in L2 (Danessi 1992; Littlemore and Low 2006; Cieślicka 2006). A number of studies have, however, suggested that bilingual figurative language comprehension is positively correlated with L2 proficiency level and language dominance, as no between-language differences to L1 and L2 nonliteral utterances were observed when bilingual speakers were balanced and/or highly proficient in L2 (Matlock and Heredia 2002; Heredia and Cieślicka 2014; Zhao et al. 2014; Vaid et al. 2015; Bromberek-Dyzman and Rataj 2016; Heredia and Cieślicka 2016). However, in light of the present findings, it seems difficult to explain the between-language differences to conventional metaphoric word pairs as modulated by language dominance or/and proficiency level. Even though one might argue that the participants of Experiment 1 were highly proficient yet non-dominant in L2, and therefore differences in language dominance might have modulated the observed results, following such an assumption, a decreased sensitivity would have been observed not only in response to conventional, but also to novel metaphors presented in L2 relative to L1.

Thus far only one experiment has examined the comprehension of conventional metaphors in bilingualism. Mashal et al. (2015) tested native speakers of Hebrew and balanced English-Hebrew bilingual speakers in a divided visual field paradigm with a semantic decision task in response to Hebrew conventional metaphoric, literal, and unrelated word dyads. The authors observed between-language differences in hemispheric involvement to conventional metaphors, with a left hemisphere advantage for L1 and right hemisphere advantage for L2 conventional metaphoric word pairs. Mashal and colleagues (2015) interpreted such results as in agreement with the Graded Salience Hypothesis (Giora 2002; see section 2.3.4), which postulates a right hemisphere preference for less salient meanings. Following this hypothesis, conventional metaphors might be less salient in L2 relative to L1, and might therefore be processed somewhat similarly to novel metaphors. Such an interpretation is also in line with the Literal Salience Model (Cieślicka

2006), according to which figurative meaning is of lower salience in the non-native compared to the native tongue. The results observed in Experiment 1 might thus be interpreted as partially in line with what Mashal et al. (2015) found, as longer response latencies, lower accuracy rates, and sustained negativity for conventional metaphors in L2 relative to L1 might point to a decreased sensitivity to the levels of metaphor conventionality in L2. Consequently, access to the nonliteral route, as reflected in ongoing negativity (Arzouan et al. 2007; Goldstein et al. 2012; Tang et al. 2017a), might be required during conventional metaphor integration only in the non-native tongue (Jankowiak et al. 2017).

Importantly, the results obtained from Experiment 1 suggest that, unlike conventional metaphor, novel metaphor comprehension is not modulated by language nativeness, as novel metaphoric word pairs evoked similar response times, accuracy rates as well as sustained negativity within the LPC time frame in both the native and non-native language. In order to further examine the extent to which mechanisms engaged in novel metaphor comprehension might be independent of language nativeness, Experiment 2 investigated the comprehension of two types of novel meanings: novel similes (*A is like B*) and novel nominal metaphors (*A is B*). As argued in the Career of Metaphor Model (Bowdle and Gentner 2005), monolingual novel metaphor comprehension involves comparison processes between a metaphoric target concept and a literal base concept. Therefore, a comparison form present in similes should facilitate such comparison mechanisms, and consequently ease the process of novel metaphoric meaning comprehension. In support of this assumption, the results obtained from Experiment 2 showed that novel similes were easier and faster to comprehend, as indicated by shorter response times and higher accuracy rates for novel similes compared to novel nominal metaphors, which instead of initiating comparison processes, promote categorization mechanisms. Importantly, the observed effect was irrespective of language, and was observed for both the native and non-native tongue, therefore extending the assumptions postulated in the Career of Metaphor Model (Bowdle and Gentner 2005) from the monolingual to the bilingual context. Such findings suggest that a comparison form present in similes might initiate comparison processes, which might in turn ease novel metaphoric meaning comprehension in both the native and non-native language. Consequently, the results obtained from the two experiments suggest that mechanisms involved in novel meaning comprehension are independent of language nativeness.

Since metaphor conventionality reflects how frequently a metaphor vehicle is used in its nonliteral sense (Bowdle and Gentner 2005), the results obtained from the two experiments might point to a lower subjective frequency of conventional metaphoric meanings in L2 relative to L1. In contrast, novel metaphors might be similarly taxing in both the native and non-native language due to the fact that they are of low frequency in both languages. Consequently, between-language differences in response to conventional metaphoric meanings might be modulated by their frequencies of usage. A lower perceived frequency of L2 conventional metaphors might stem from the fact that figurative language is often neglected and undermined when acquiring L2 in a formal school setting, and thus L2 conventional metaphors might be processed similarly to novel metaphors due to the fact that conventional metaphoric meanings are usually frequent in everyday language, but not necessarily in the L2 classroom setting. On the other hand, novel metaphors are generally non-frequent in both native and non-native everyday language. The subjective frequency of novel and conventional metaphoric word dyads in L1 and L2 was, nonetheless, not measured in the present experiments, and thus such an interpretation requires confirmation from future studies.

Overall, the results observed in the two experiments seem to suggest that late proficient unbalanced bilingual speakers are sensitive to metaphor conventionality, but mostly in their native tongue. In their non-native language, on the other hand, both conventional and novel metaphors are likely to be of low frequency, and thus they are both non-conventionalized. As a result, the comprehension of conventional metaphoric meanings, which are of high perceived frequency for native speakers, yet of low subjective frequency for L2 learners of a given language, might be more resource intensive in L2 relative to L1. Novel metaphors, which are, on the other hand, rarely used in both L1 and L2, might be similarly challenging to comprehend in both languages.

6.4.2. Lexico-semantic access and meaning integration in bilingualism

The next objective of the experiments was to investigate the automaticity of lexico-semantic access and meaning integration mechanisms in bilingual language processing. As argued in the temporal delay assumption postulated within the BIA+ model (Dijkstra and

van Heuven 2002), L2 semantic representations are activated slower and less automatically in L1-dominant bilinguals. In favor of this hypothesis, longer response latencies (Experiment 1 and Experiment 2) and a delay in the N400 peak latency (Experiment 1) in response to the non-native compared to the native language were observed. Such results are in line with previous behavioral (e.g., Dijkstra et al. 1998a; Dijkstra et al. 1999; de Groot et al. 2002) and electrophysiological research on bilingual lexico-semantic processing (e.g., Weber-Fox and Neville 1996; Phillips et al. 2004; Moreno and Kutas 2005; Braunstein et al. 2012). The observed results might be indicative of a lower perceived frequency of L2 lexical items when bilingual speakers are not balanced, and they use their native tongue more frequently than their non-native language. Nevertheless, further research is needed in order to investigate whether a delay in the N400 peak latency along with longer response times for the non-native language might be due to differences in the subjective frequency of L1 and L2 words, as this aspect was directly measured neither in the present studies nor in previous experiments on bilingual lexico-semantic processing.

In addition to between-language differences in reaction times and N400 peak latency, Experiment 1 (Jankowiak et al. 2017) showed that L2 lexical items elicited attenuated N400 amplitudes relative to L1 words. A reduced N400 response to L2 compared to L1 has previously been observed (e.g., Proverbio et al. 2002; Moreno et al. 2008; Midgley et al. 2009b; Newman et al. 2012), and has been interpreted as in agreement with the functional role of the N400, according to which the N400 response indexes information retrieval from long-term memory (Kutas and Federmeier 2000; Kutas and Federmeier 2011; Kotz et al. 2012). In line with this assumption, attenuated N400 amplitudes for L2 relative to L1 might indicate a decreased activity in long-term memory, possibly due to a weaker interconnectivity between L2 lexical items within the semantic network when bilingual speakers are not dominant in their non-native tongue (Midgley et al. 2009b).

Interestingly, unlike within the N400 time frame, no general between-language differences were observed in the LPC latency. Instead, the LPC magnitude to L1 and L2 was modulated by the level of semantic complexity, as attenuated LPC amplitudes were observed for semantically complex meanings (i.e., for novel metaphors in L1, and for both novel and conventional metaphors in L2) compared to meanings that were semantically simple (i.e., literal utterances). The fact that no between-language differences were found with regard to the latency of the LPC response might point to a similar time course

of meaning integration mechanisms in L1 and L2. However, it needs to be noted that previous electrophysiological studies into bilingual language processing have rarely examined between-language differences in meaning integration mechanisms, as indexed by the LPC response, and therefore more research is needed in order to show how language nativeness modulates ERP patterns within the LPC time frame (Jankowiak et al. 2017).

Finally, although no hypotheses were formulated with regard to the P200 response, general between-language differences were observed as soon as between 150-250 ms post stimulus onset, where reduced P200 amplitudes were found in response to L1 compared to L2 lexical items (Jankowiak et al. 2017). Owing to the fact that previous monolingual studies into lexical processing have linked the P200 response with mechanisms engaged in early item encoding, and have shown a negative correlation between lexical frequency and P200 magnitude (Assadollahi and Pulvermüller 2001; Hauk et al. 2004; Dambacher et al. 2006), the observed results might be indicative of a lower perceived frequency of L2 lexical items. It therefore seems that differences in the subjective frequency of L1 and L2 words might have modulated cognitive mechanisms engaged in lexical processing, as reflected in both P200 and N400 results. As already mentioned, subjective frequency was, however, not directly measured in the present study, and thus such an interpretation would need to find support in future studies into bilingual lexico-semantic processing.

All in all, the experiments reported in this dissertation indicate that while novel metaphor comprehension is cognitively taxing in both the native and non-native language, conventional metaphors seem to be effortful only in L2, which might result from a between-language difference in the frequency of conventional metaphoric meanings. Additionally, the current results suggest that whereas lexico-semantic access is less automatic in L2 relative to L1, irrespectively of utterance type, meaning integration mechanisms are modulated by the semantic complexity of the stimuli in both the native and non-native tongue.

6.5. The importance of methodology triangulation

Methodology triangulation is defined as involving more than one research method in order to study a particular phenomenon (Bekhet and Zauszniewski 2012: 40). The studies

reported in the present dissertation employed three quantitative research methods. Survey research, as an off-line measure, was used pre-experimentally in order to evaluate the perception of the experimental stimuli. Behavioral measures (i.e., reaction times and accuracy rates), as on-line measures, provided information regarding the time participants needed in order to perform a semantic decision task, and showed how correct they were when performing the task. In both experiments, only RTs for correct responses were analyzed, thus limiting a potential confound stemming from the speed-accuracy tradeoff. Finally, event-related brain potentials were used to examine specific stages of lexico-semantic processing involved in bilingual novel and conventional metaphor comprehension.

The significance of survey research with rating scales was demonstrated not only when describing particular aspects that the surveys were aimed to evaluate pre-experimentally (i.e., levels of abstractness, predictability, meaningfulness, familiarity, and metaphoricity of the stimuli), but also when conducting follow-up correlation analyses, which showed whether the participants of the normative studies and those of the experiments proper made similar decisions regarding the meaningfulness of the materials. Consequently, the correlation analyses ensured that the stimuli used in the experiments had been properly selected. Furthermore, the correlation analyses showed that English materials were evaluated similarly by both English native speakers and English L2 learners, which increases the validity of the results obtained from the two experiments.

Nonetheless, survey research is an off-line research method, which does not provide information regarding the time-course of language processing. Additionally, it does not reflect effects that might be unnoticeable to language users. Thus, in order to investigate real-time aspects of language processing as a rapid mechanism, on-line methods need to be employed. Reaction times were employed in the two experiments with a view to measuring the elapsed time before the presentation of a stimulus and a response given. RT results could then be interpreted as indicative of linguistic complexity, as longer RTs are assumed to reflect increased processing difficulty (Kaiser 2013: 137).

Despite such advantages, it needs to be noted that reaction times index all processes underlying language comprehension together with mechanisms related to decision making, and thus they cannot be used to examine specific stages of language processing. To investigate the exact time-course of meaning comprehension, on-line electrophysiological methods can be employed, which provide a continuous measure of brain activity

and, unlike reaction time measures, reflect neurobiological responses to a linguistic stimulus with a temporal resolution up to 1 ms (Luck 2005; Cohen 2014). For instance, in Experiment 1, reaction time results showed language- as well as utterance-specific effects; however, ERP results showed how specific stages of figurative language comprehension were modulated by language nativeness. Namely, between-language differences were observed within the phase of early item encoding (i.e., the P200 response), as well as lexico-semantic access (i.e., the N400 response). Utterance-specific effects were also found within the time window of meaning access, but additionally within the time frame of meaning integration mechanisms (i.e., the LPC response). Consequently, employing an ERP method enabled the analyses of how particular stages of meaning processing were influenced by language nativeness as well as the semantic complexity of the experimental stimuli.

All in all, bearing in mind the advantages that each of the aforementioned method has to offer, their triangulation might provide reliable insights into the subject under investigation, and might increase the validity of experimental results. Namely, off-line methods (e.g., survey research) and on-line techniques (e.g., response times, electroencephalography) can complement each other, with off-line measures providing the outcome of meaning comprehension, and on-line methods tapping into its time course (Garrod 2006: 251). As a result, the same patterns of results obtained from different research techniques might minimize any inadequacies potentially found in one of these sources, and might make it easier to draw final conclusions.

6.6. Future research directions

The present experiments provide novel insights into lexico-semantic processes engaged when comprehending novel and conventional metaphoric utterances in L1 and L2, and their results provide a window into further research that could be conducted in the field of bilingual nonliteral language processing.

First of all, since both Experiment 1 (Jankowiak et al. 2017) and Experiment 2 employed a binary semantic decision task, replicating the experiments using different tasks could show whether the obtained results might have been influenced by the nature

of the task, and whether metaphor comprehension in bilinguals is modulated by task requirements. Task specificity seems to be especially relevant due to the fact that it has been argued to modulate mechanisms engaged in bilingual lexico-semantic processing. Namely, in line with the task schema node introduced in the BIA+ model (Dijkstra and van Heuven 2002; see section 3.2.2.2), nonlinguistic context effects, including task-related effects, influence attention-sensitive mechanisms taking place during the process of bilingual word recognition. Consequently, different tasks are likely to require different task schemas, which might in turn impact bilingual lexico-semantic processing.

Secondly, since Experiment 2 was based on behavioral measures, its results cannot account for specific stages of language processing owing to the fact that RTs reflect all processes underlying language comprehension as well as mechanisms related to decision making. Consequently, investigating novel simile and novel nominal metaphor comprehension by means of employing an ERP method could provide relevant information concerning particular stages of novel meaning comprehension in bilingualism. For instance, even with no between-language differences in RT patterns for novel nominal metaphors and novel similes, ERP data could reveal differences in the magnitude of particular language-related ERP components (e.g., the N400, LPC) elicited in response to the stimuli presented in the two languages. Therefore, testing bilingual novel simile and novel nominal metaphor comprehension by employing an ERP method might show potential differences in lexico-semantic access (as reflected in the N400 response) and semantic integration (the LPC response) when processing the two types of nonliteral utterances in the native and non-native language. Additionally, ERP patterns could account for the lack of differences in RTs for novel similes and literal sentences observed in Experiment 2 by showing potentially different brain responses to novel similes and literal utterances at particular stages of language comprehension.

Importantly, future research should also take into account individual differences that might impact how bilingual speakers comprehend metaphoric meanings in their native and non-native language. For instance, previous monolingual research into metaphor comprehension has suggested that individual differences in executive control modulate metaphoric language processing, as participants with better inhibitory skills have been observed to outperform those with lower skills on metaphor comprehension tasks (e.g., Chiappe and Chiappe 2007; Mashal et al. 2007; Mashal et al. 2013; Columbus et al. 2015).

Additionally, high-IQ participants were observed to produce more accurate interpretations of metaphoric utterances (Kazmerski et al. 2003). In a recent study, Lü et al. (2017) examined the role of executive control in bilingual familiar and unfamiliar metaphor comprehension. To assess executive control ability, Chinese-English bilinguals performed a Stroop task, in which color names were placed in colored frames, and participants pressed a corresponding key depending on the color of the frame. In negative trials, the color of the frame did not correspond to the name of the color, and thus participants had to inhibit the task-irrelevant information (i.e., the color name). The results revealed that bilinguals with higher executive control ability were faster to comprehend familiar metaphors compared to participants with lower executive control ability. Interestingly, this effect was not observed for unfamiliar metaphors as well as literal utterances. Since neither IQ nor inhibitory and executive control skills were tested in the present experiments, it remains uninvestigated whether such individual differences of participants tested modulated the obtained results. Further research taking such aspects into account could thus provide more thorough insights into metaphoric meaning comprehension in bilingualism.

6.7. Concluding remarks

The two experiments described in this thesis examined lexico-semantic processes engaged when comprehending metaphoric utterances in the native and non-native language. The results obtained from the studies suggest that while novel meaning comprehension might be language-independent, conventional metaphor comprehension seems to be more resource intensive in L2 relative to L1. Furthermore, the results indicate that L2 semantic processing is less automatic and slower in the non-native compared to the native tongue due to between-language differences in language dominance.

Conclusion

This dissertation was an attempt to elucidate lexico-semantic processes engaged when comprehending novel and conventional metaphors in the bilingual context. The main objective of the two experiments reported in this thesis was to examine whether and to what extent metaphor conventionality modulates metaphor comprehension in both the native and non-native tongue. In line with the Career of Metaphor Model (Bowdle and Gentner 2005), cognitive mechanisms involved in monolingual metaphor comprehension are highly influenced by metaphor conventionality. Namely, while novel metaphors require sense creation mechanisms that are based on comparison processes between a metaphoric target concept and a literal base concept, conventional metaphors preferentially involve categorization mechanisms related to sense retrieval processes. Consequently, novel metaphor comprehension is argued to be more cognitively taxing compared to conventional metaphors, which has been shown in previous behavioral and electrophysiological research on monolingual speakers (e.g., Bowdle and Gentner 2005; Arzouan et al. 2007; Lai et al. 2009; Goldstein et al. 2012; Tang et al. 2017a). Additionally, monolingual studies into novel metaphor processing have provided support for the crucial role of comparison mechanisms engaged when arriving at a novel metaphoric meaning, suggesting that comparison processes initiated when comprehending similes might ease novel meaning comprehension (Bowdle and Gentner 2005; Utsumi 2011; Goldstein et al. 2012; Lai et al. 2013). Importantly, previous research has not been devoted to extending studies on novel and conventional metaphoric language comprehension from the monolingual to the bilingual context, and thus the present experiments provide novel insights into lexico-semantic processing in bilingual metaphor comprehension.

Experiment 1 (Jankowiak et al. 2017) employed an event-related potential method in order to examine brain responses to novel metaphoric, conventional metaphoric, literal, and anomalous word pairs in L1 and L2. In line with what was hypothesized, within the late N400 (400-500 ms) time window, a graded effect of utterance type was observed in both languages, with the most pronounced N400 amplitudes for anomalous word dyads, followed by novel metaphoric, conventional metaphoric, and finally literal utterances. Such results indicate the increasing difficulty of mapping between concepts from literal to conventional metaphoric, to novel metaphoric, and finally to anomalous word pairs in both the native and non-native language. Between-language differences were, nonetheless, observed within the LPC time frame (500-800 ms), in which novel metaphors evoked attenuated LPC amplitudes relative to literal word pairs in both languages, while conventional metaphors elicited reduced LPC responses only in L2. These results corroborate the behavioral results observed in Experiment 1, which showed longer response times along with lower accuracy rates for conventional metaphoric word pairs in L2 compared to L1. Overall, such results suggest that while novel metaphor integration is similarly effortful in both languages, conventional metaphor comprehension is more taxing in the non-native tongue, which might result from a lower frequency of conventional metaphors in L2 than L1.

Experiment 2 was further aimed at examining whether and to what extent novel metaphor comprehension in bilingualism might be language-independent. To this end, response times and accuracy rates were measured in response to novel nominal metaphors, novel similes, literal, and anomalous sentences in L1 and L2. In agreement with the formulated hypotheses, novel similes, whose linguistic form initiates comparison mechanisms presumably engaged in novel metaphor comprehension, elicited faster RTs and higher accuracy rates compared to novel nominal metaphors, which promote a categorization mechanism. Importantly, the observed effect was independent of language nativeness, thus suggesting that a comparison form initiates comparison mechanisms, and consequently eases novel metaphor comprehension in both the native and non-native tongue. Such results are therefore in line with those observed in Experiment 1, pointing to similar processes involved in novel meaning comprehension in both languages.

The final objective of the two experiments was to examine lexico-semantic processing in bilingualism. As postulated within the Bilingual Interactive Plus Model (BIA+; Dijkstra and van Heuven 2002), the activation of semantic representations is delayed in

L2 compared to L1 when bilingual speakers are L2-non-dominant and use their L2 less frequently than their native tongue. The two experiments reported in the thesis provide support for this assumption, showing a delayed N400 response (Experiment 1) as well as longer RTs for L2 relative to L1 (Experiment 1 and Experiment 2). Such results indicate that semantic processing might be less automatic in the non-native language, which might result from a lower subjective frequency of L2 words in L2-non-dominant speakers. Importantly, between-language differences in the perceived frequency of L1 and L2 lexical items were also reflected within the P200 time window, in which non-native items elicited more pronounced P200 amplitudes compared to L1 words.

In addition, in Experiment 1, attenuated N400 amplitudes were observed for L2 compared to L1, irrespectively of utterance type. These results might be indicative of a weaker interconnectivity between L2 lexical items in the semantic network, which consequently leads to a decreased amount of information activated in long-term memory during L2 lexico-semantic access. Interestingly, unlike the N400, no general between-language differences were observed within the LPC response, which was instead modulated by semantic complexity in both languages. Namely, semantically complex meanings (i.e., novel metaphors in L1 and both novel and conventional metaphors in L2) evoked prolonged negativity, thus pointing to the ongoing difficulty in meaning integration or access to the nonliteral route during the comprehension of semantically complex meanings in both languages.

Overall, the results of the experiments make a valuable contribution to how we understand bilingualism, and show how bilingual speakers process and integrate non-native utterances at different levels of semantic complexity. Future research is, nonetheless, needed in order to further investigate specific factors by which bilingual metaphoric language comprehension might be modulated, including task requirements and the individual differences of participants tested.

Summary

Much research has recently been devoted to examining metaphoric language comprehension in the monolingual context, as this can provide novel insights into the mechanisms engaged in language processing at different levels of semantic complexity. Traditional views of metaphor comprehension claim that the metaphoric meaning of an utterance is available only once its literal interpretation has been activated and then rejected (Grice 1975). More recent approaches, on the other hand, emphasize specific factors that modulate metaphoric meaning comprehension, including metaphor conventionality, salience, contextual constraints, and the linguistic form of a metaphor itself (Frisson and Pickering 2001; Katz and Ferretti 2001; Giora 2002; Bowdle and Getner 2005). In line with the Career of Metaphor Model (Bowdle and Gentner 2005), while the comprehension of novel (unfamiliar) metaphors is based on comparison mechanisms, conventional (familiar) metaphor comprehension involves the process of categorization. As a result of such assumptions, the model further predicts that comparison processes, which are initiated when processing similes, might ease novel metaphor comprehension.

Nonetheless, the aforementioned postulates have been proposed with a view to elucidating the cognitive mechanisms engaged in monolingual metaphor comprehension. Little attention has, on the other hand, been devoted to examining how bilingual speakers comprehend novel and conventional metaphoric language in their native and non-native tongue. This dissertation was therefore aimed to experimentally test how late proficient unbalanced Polish (L1) – English (L2) bilingual speakers process novel and conventional metaphors (Experiment 1; Jankowiak et al. 2017), as well as novel nominal metaphors and novel similes (Experiment 2) in their L1 and L2. To this end, behavioral (Experiment 1 and Experiment 2) and event-related potential (Experiment 1) data were collected from

participants, who performed a binary semantic decision task for the presented utterances. Behavioral as well as event-related potential results indicated that while the comprehension of novel meanings is language-independent, conventional metaphor comprehension seems to be more resource intensive in L2 relative to L1. Additionally, the results pointed to a less automatic and delayed processing of semantic information in the non-native as opposed to the native tongue, which might be modulated by language dominance. Taken together, the observed results indicated that late proficient unbalanced bilingual speakers are less sensitive to the levels of metaphor conventionality in L2 than in L1, and pointed to a less automatic lexico-semantic processing in the non-native compared to the native tongue.

Streszczenie

Na przestrzeni ostatnich lat, przeprowadzono wiele badań, których celem było zbadanie przetwarzania języka metaforycznego w języku rodzimym, jako że umożliwia to skontrastowanie znaczeń semantycznie prostych (tj., dosłownych) ze znaczeniami semantycznie złożonymi (tj., metaforycznymi), dostarczając tym samym nowego wglądu w mechanizmy zaangażowane w rozumienie języka o różnych poziomach semantycznej złożoności. Zgodnie z tradycyjnym podejściem do rozumienia języka figuratywnego, znaczenie metaforyczne jest zawsze dostępne dopiero po wcześniejszej aktywacji oraz odrzuceniu znaczenia dosłownego (Grice 1975). Współczesne teorie podkreślają z kolei rolę czynników modulujących rozumienie znaczeń metaforycznych, do których zalicza się poziom konwencjonalności przedstawionego wyrażenia metaforycznego, kontekst, w którym metafora jest zaprezentowana, jak również jej forma językowa (Frisson and Pickering 2001; Katz and Ferretti 2001; Giora 2002; Bowdle and Gentner 2005). Zgodnie z modelem *Career of Metaphor* (Bowdle and Gentner 2005), metafory nowe są rozumiane jako porównania, podczas gdy rozumienie metafor utartych wymaga procesów kategoryzacyjnych. W rezultacie, rozumienie metafor nowych powinno być ułatwione, kiedy są one przedstawione w formie porównań, które to inicjują mechanizmy porównawcze.

Powyższe postulaty dotyczą jednak procesów zaangażowanych w rozumienie znaczeń metaforycznych w języku rodzimym. Jak dotąd, mało uwagi było poświęcone zbadaniu tego, jak osoby dwujęzyczne rozumieją nowe oraz utarte znaczenia metaforyczne w języku rodzimym (L1) oraz w języku obcym (L2). Niniejsza rozprawa doktorska miała zatem na celu empiryczne zbadanie tego, jak rodzimi użytkownicy języka polskiego, którzy są na wysokim poziomie zaawansowania w języku angielskim jako języku

obcym, rozumieją metafory nowe oraz utarte (Eksperyment 1), jak również nowe znaczenia metaforyczne przedstawione w formie metafor rzeczownikowych oraz w formie porównań (Eksperyment 2). W tym celu, przeanalizowane zostały dane behawioralne (Eksperyment 1 oraz Eksperyment 2), jak również potencjały wywołane (Eksperyment 1), podczas gdy uczestnicy badań podejmowali decyzję semantyczną w odpowiedzi na wyrażenia zaprezentowane w języku polskim oraz angielskim. Zaobserwowane wyniki behawioralne oraz elektrofizjologiczne pokazały, że podczas gdy rozumienie nowych znaczeń metaforycznych jest niezależne od języka, rozumienie metafor utartych jest trudniejsze w języku obcym niż w języku rodzimym. Co więcej, otrzymane wyniki wskazały na mniej automatyczne przetwarzanie informacji semantycznych w L2, co może wynikać z różnic w dominacji języka rodzimego oraz obcego. Wyniki przedstawionych badań pokazały zatem, że osoby biegle posługujące się językiem obcym, których językiem dominującym jest jednak język rodzimy, są mniej wrażliwe na poziom konwencjonalności wyrażen metaforycznych w L2 niż w L1 oraz przetwarzają informacje językowe mniej automatycznie w języku obcym niż w języku ojczystym.

References

- Abutalebi, Jubin. 2008. "Neural aspects of second language representation and language control", *Acta Psychologica* 128: 466–478.
- Aisenman, Ravid A. 1999. "Structure-mapping and the simile-metaphor preference", *Metaphor and Symbol* 14, 1: 45–51.
- Allen, Mark, William Badecker and Lee Osterhout. 2003. "Morphological analysis in sentence processing: An ERP study", *Language and Cognitive Processes* 18: 405–430.
- Alvarez, Ruben P., Phillip J. Holcomb and Jonathan Grainger. 2003. "Accessing word meaning in two languages: An event-related brain potential study of beginning bilinguals", *Brain and Language* 87, 2: 290–304.
- Anderson, John R., Lynne M. Reder and Christian Lebiere. 1996. "Working memory: Activation limitations on retrieval", *Cognitive Psychology* 30: 221–256.
- Andrews, Mark, Gabriella Vigliocco and David Vinson. 2009. "Integrating experiential and distributional data to learn semantic representations", *Psychological Review* 116, 3: 463–498.
- Arzouan, Yossi, Abraham Goldstein and Miriam Faust. 2007. "Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension", *Brain Research* 1160: 69–81.
- Assadollahi, Ramin and Friedemann Pulvermüller. 2001. "Neuromagnetic evidence for early access to cognitive representations", *Neuroreport* 12, 2: 207–213.
- Balconi, Michela and Simona Amenta. 2010. "A fighter is a lion. Neuropsychological indexes in comprehending frozen metaphors", *Journal of Pragmatics* 42, 12: 3246–3257.

- Balota, David A. and Daniel H. Spieler. 1999. "Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency", *Journal of Experimental Psychology: General* 128, 1: 32–55.
- Balota, David A., Susan D. Sergent-Marshall, Michael J. Cortese and Daniel H. Spieler. 2004. "Visual word recognition of single-syllable words", *Journal of Experimental Psychology: General* 133, 2: 283–316.
- Bambini, Valentina, Chiara Bertini, Walter Schaeken, Alessandra Stella and Francesco Di Russo. 2016. "Disentangling metaphor from context: An ERP study", *Frontiers in Psychology* 7: 1–14.
- Bambini, Valentina, Marta Ghio, Andrea Moro and Petra B. Schumacher. 2013. "Differentiating among pragmatic uses of words through timed sensicality judgments", *Frontiers in Psychology* 4: 1–16.
- Barber, Horacio, Marta Vergara and Manuel Carreiras. 2004. "Syllable-frequency effects in visual word recognition: Evidence from ERPs", *Neuroreport* 15, 3: 545–548.
- Beeman, Mark. 1998. "Coarse semantic coding and discourse comprehension", in: Mark Beeman and Christine Chiarello (eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*. New Jersey: Erlbaum, 255–284.
- Bekhet, Abir K. and Jaclene A. Zauszniewski. 2012. "Methodological triangulation: An approach to understanding data", *Nurse Researcher* 20, 2: 40–43.
- Bialystok, Ellen. 2010. "Bilingualism", *Wiley Interdisciplinary Reviews: Cognitive Science* 1, 4: 559–572.
- Blank, Glenn D. 1988. "Metaphors in the lexicon", *Metaphor and Symbolic Activity* 3, 3: 21–36.
- Bondaruk, Anna. 2014. "Characterizing and defining predicational clauses in Polish", in: Ludmila Veselovská and Markéta Janebová (eds.), *Proceedings of the Olomouc Linguistics Colloquium 2014: Language use and language structure*. Olomouc: Olomouc Modern Languages Series (OMLS), 333–348.
- Bourne, Victoria J. 2006. "The divided visual field paradigm: Methodological considerations", *Laterality* 11, 4: 373–393.
- Bowdle, Brian F. and Dedre Gentner. 1999. "Metaphor comprehension: From comparison to categorization", in: Martin Hahn and Scott C. Stoness (eds.), *Proceedings of the twenty-first annual conference of the cognitive science society*. New Jersey: Lawrence Erlbaum Associates, 90–95.

- Bowdle, Brian F. and Dedre Gentner. 2005. "The career of metaphor", *Psychological Review* 112, 1: 193–216.
- Brain Products. 2012. actiCAP 64-channel standard montage. (http://www.brain-products.com/files/public/downloads/actiCAP-64-channel-Standard-2_1201.pdf) (date of access: 15 Dec. 2016).
- Braunstein, Verena, Anja Ischebeck, Clemens Brunner, Roland H. Grabner, Maxim Stamenov and Christa Neuper. 2012. "Investigating the influence of proficiency on semantic processing in bilinguals: An ERP and ERD/S analysis", *Acta neurobiologiae experimentalis* 72, 4: 421–438.
- Bridger, Emma K., Regine Bager, Olga Kriukova, Kerstin Unger and Axel Mecklinger. 2012. "The FN400 is functionally distinct from the N400", *NeuroImage* 63: 1334–1342.
- Bromberek-Dyzman, Katarzyna and Karolina Rataj. 2016. "Irony comprehension in the non-native language comes at a cost", *Psychology of Language and Communication* 20, 3: 336–353.
- Bromberek-Dyzman, Katarzyna, Karolina Rataj and Jacek Dylak. 2010. "Mentalizing in the second language: Is irony online inferencing any different in L1 and L2?", in: Iwona Witczak-Plisiecka (ed.), *Pragmatic perspectives on language and linguistics. Vol.1: Speech actions in theory and applied studies*. Newcastle: Cambridge Scholars Publishing, 197–216.
- Brouwer, Harm and Matthew W. Crocker. 2017. "On the proper treatment of the N400 and P600 in language comprehension", *Frontiers in Psychology* 8: 1–5.
- Brouwer, Harm, Hartmut Fitz and John C.J. Hoeks. 2012. "Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension", *Brain Research* 1446: 127–143.
- Brouwer, Harm, Matthew W. Crocker, Noortje J. Venhuizen and John C.J. Hoeks. 2017. "A neurocomputational model of the N400 and the P600 in language processing", *Cognitive Science* 41: 1318–1352.
- Brown, Colin and Peter Hagoort. 1993. "The processing nature of the N400: Evidence from masked priming", *Journal of Cognitive Neuroscience* 5: 34–44.
- Brybaert, Marc and Ton Dijkstra. 2006. "Changing views on word recognition in bilingualism", in: José Morais and Géry van Outryve d'Ydewalle (eds.), *Bilingualism and second language acquisition*. Brussels: KVAB, 25–37.

- Brysbaert, Marc and Wouter Duyck. 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, Marc, Amy Beth Warriner, and Victor Kuperman. 2014. "Concreteness ratings for 40 thousand generally known English word lemmas", *Behavior Research Methods* 46, 3: 904–911.
- Brysbaert, Marc. 1998. "Word recognition in bilinguals: Evidence against the existence of two separate lexicons", *Psychologica Belgica* 38: 163–175.
- Caldwell-Harris, Catherine L. and Ayşe Ayçiçeği-Dinn. 2009. "Emotion and lying in a non-native language", *International Journal of Psychophysiology* 71, 3: 193–204.
- Camp, Elisabeth. 2006. "Metaphor in the mind: The cognition of metaphor", *Philology Compass* 1: 154–170.
- Canseco-Gonzales, Enriqueta, Laurel Brehm, Cameron A. Brick, Sarah Brown-Schmidt, Kara Fischer and Katie Wagner. 2010. "Carpet or Cárcel: The effect of age of acquisition and language mode on bilingual lexical access", *Language and Cognitive Processes* 25, 5: 669–705.
- Cardillo, Eileen R., Gwenda L. Schmidt, Alexander Kranjec and Anjan Chatterjee. 2010. "Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor", *Behavior Research Methods* 42, 3: 651–664.
- Carston, Robyn and Catherine Wearing. 2011. "Metaphor, hyperbole and simile: A pragmatic approach", *Language and Cognition* 3, 2: 283–312.
- Carston, Robyn. 2010. "Lexical pragmatics, ad hoc concepts and metaphor: A Relevance Theory perspective", *Italian Journal of Linguistics* 22, 1: 153–180.
- Casaponsa, Aina, Manuel Carreiras and Jon Andoni Duñabeitia. 2014. "Discriminating languages in bilingual context: The impact of orthographic markedness", *Frontiers in Psychology* 424: 1–10.
- Casaponsa, Aina, Manuel Carreiras and Jon Andoni Duñabeitia. 2015. "How do bilinguals identify the language of the words they read?", *Brain Research* 1624: 153–166.
- Chen, Hongjun, Xiaoshuang Peng and Yanli Zhao. 2013. "An ERP study on metaphor comprehension in the bilingual brain", *Chinese Journal of Applied Linguistics* 36, 4: 505–517.

- Chetail, Fabienne. 2014. "Effect of number of syllables in visual word recognition: New insights from the lexical decision task", *Language, Cognition and Neuroscience* 29, 10: 1249–1256.
- Chiappe, Dan and Penny Chiappe. 2007. "The role of working memory in metaphor production and comprehension", *Journal of Memory and Language* 56: 172–188.
- Chiappe, Dan, John M. Kennedy and Tim Smykowski. 2003. "Reversibility, aptness, and the conventionality of metaphors and similes", *Metaphor and Symbol* 18: 85–105.
- Chiarello, Christine. 1998. "On codes of meaning and the meaning of codes: Semantic access of retrieval within and between hemispheres", in: Mark Beeman and Christine Chiarello (eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*. New Jersey: Erlbaum, 141–160.
- Choi, Sungmook, Jingu Kim and Kwangmin Ryu. 2014. "Effects of context on implicit and explicit lexical knowledge: An event-related potential study", *Neuropsychologia* 63: 226–234.
- Christensen, Larry B., Burke Johnson and Lisa Anne Turner. 2014. *Research methods, design, and analysis*. Boston: Pearson.
- Cieślicka, Anna B. 2006. "Literal salience in on-line processing of idiomatic expressions by second language learners", *Second Language Research* 22: 115–144.
- Cieślicka, Anna B. 2010. "Formulaic language in L2. Storage, retrieval and production of idioms by second language learners", in: Martin Putz and Laura Sicola (eds.), *Cognitive processing in second language acquisition – inside the learner's mind*. Philadelphia: John Benjamin Publishing, 149–168.
- Cieślicka, Anna B. and Roberto R. Heredia. 2011. "Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context", *Brain and Language* 116, 3: 136–150.
- Cieślicka, Anna B., Roberto R. Heredia and Marc Olivares. 2014. "It's all in the eyes: How language dominance, salience, and context affect eye movements during idiomatic language processing", in: Mirosław Pawlak and Larissa Aronin (eds.), *Essential topics in applied linguistics and multilingualism*. Cham: Springer International Publishing, 21–41.
- Cieślicka, Anna B., Roberto R. Heredia and Tanya Garcia. 2017. "Task effects in bilingual idiom comprehension", *Poznań Studies in Contemporary Linguistics* 53, 1: 95–117.

- Cohen, Mark S. 2008. Handedness questionnaire. (<http://www.brainmapping.org/shared/Edinburgh.php>) (date of access: 19 Jan. 2015).
- Cohen, Mike X. 2014. *Analyzing neural time series data*. Cambridge: MIT Press.
- Columbus, Georgie, Naveed A. Sheikh, Marilena Côté-Lecaldare, Katja Häuser, Shari R. Baum and Debra Titone. 2015. “Individual differences in executive control relate to metaphor processing: An eye movement study of sentence reading”, *Frontiers in Human Neuroscience* 8: 1–12.
- Comesaña, Montserrat, Rosa Sánchez-Casas, Ana Paula Soares, Ana P. Pinheiro, Andreia Rauber, Sofia Frade and Isabel Fraga. 2012. “The interplay of phonology and orthography in visual cognate word recognition: An ERP study”, *Neuroscience Letters* 529, 1: 75–79.
- Coney, Jeffrey and Kimberly D. Evans. 2000. “Hemispheric asymmetries in the resolution of lexical ambiguity”, *Neuropsychologia* 38, 3: 272–282.
- Connine, Cynthia, John Mullenix, Eve Shernoff and Jennifer Yelen. 1990. “Word familiarity and frequency in visual and auditory word recognition”, *Journal of Experimental Psychology: Learning, Memory and Cognition* 16, 6: 1084–1096.
- Content, Alain and Ulrich H. Frauenfelder. 1996. “On the need for computer modeling: The case of language processing”, *Psychologica Belgica* 36: 113–144.
- Cooper, Thomas C. 1999. “Processing of idioms by L2 learners of English”, *TESOL Quarterly* 33: 233–262.
- Costa, Albert, Mikel Santesteban and Agnès Caño. 2005. “On the facilitatory effects of cognates in bilingual speech production”, *Brain and Language* 94: 94–103.
- Coulson, Seana and Cyma Van Petten. 2002. “Conceptual integration and metaphor: An event-related potential study”, *Memory and Cognition* 30, 6: 958–968.
- Coulson, Seana and Marta Kutas. 2001. “Getting it: Human event-related brain response to jokes in good and poor comprehenders”, *Neuroscience Letters* 316, 2: 71–74.
- Curran, Tim. 2000. “Brain potentials of recollection and familiarity”, *Memory and Cognition* 28, 6: 923–938.
- Daltrozzo, Jérôme, Norma Wioland and Boris Kotchoubey. 2012. “The N400 and late positive complex (LPC) effects reflect controlled rather than automatic mechanisms of sentence processing”, *Brain Sciences* 2: 267–297.

- Dambacher, Michael, Reinhold Kliegl, Markus Hofmann and Arthur M. Jacobs. 2006. "Frequency and predictability effects on event-related potentials during reading", *Brain Research* 1084: 89–103.
- Danesi, Marcella. 1992. "Metaphorical competence in second language acquisition and second language teaching: The neglected dimension", in: James E. Alatis (ed.), *Georgetown University round table on languages and linguistics*. Washington, DC: Georgetown University Press, 489–500.
- De Angelis, Gessica. 2007. *Third or additional language acquisition*. Toronto: Multilingual Matters LTD.
- De Grauwe, Sophie, Abigail Swain, Phillip J. Holcomb, Tali Ditman and Gina R. Kuperberg. 2010. "Electrophysiological insights into the processing of nominal metaphors", *Neuropsychologia* 48, 7: 1965–1984.
- De Groot, Annette M.B. 1992. "Bilingual lexical representation: A closer look at conceptual representation", in: Ram Frost and Leonard Katz (eds.), *Orthography, phonology, morphology and meaning*. Amsterdam: Elsevier Science, 389–412.
- De Groot, Annette M.B., Susanne Borgwaldt, Mieke Bos and Ellen van den Eijnden. 2002. "Lexical decision and word naming in bilinguals: Language effects and task effects", *Journal of Memory and Language* 47: 91–124.
- Deacon, Diana L., Anna Dynowska, Walter Ritter and Jill Grose-Fifer. 2004. "Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition", *Psychophysiology* 28: 185–200.
- DeLong, Katherine A., Thomas P. Urbach, David M. Groppe and Marta Kutas. 2011. "Overlapping dual ERP responses to low cloze probability sentence continuations", *Psychophysiology* 48: 1203–1207.
- Dijkstra, Ton and Koenraad de Smedt. 1996. "Computer models in psycholinguistics: An introduction", in: Ton Dijkstra and Koenraad de Smedt (eds.), *Computational psycholinguistics: AI and connectionist models of human language processing*. London: Taylor & Francis, 3–23.
- Dijkstra, Ton and Walter J.B. van Heuven. 1998. "The BIA model and bilingual visual word recognition", in: Jonathan Grainger and Arthur M. Jacobs (eds.), *Localist connectionist approaches to human cognition*. Mahwah, NJ: Erlbaum Associates, 189–225.

- Dijkstra, Ton and Walter J.B. van Heuven. 2002. "The architecture of the bilingual word recognition system: From identification to decision", *Bilingualism: Language and Cognition* 5: 175–197.
- Dijkstra, Ton, Henk van Jaarsveld and Sjoerd Ten Brinke. 1998a. "Interlingual homograph recognition: Effects of task demands and language intermixing", *Bilingualism: Language and Cognition* 1, 1: 51–66.
- Dijkstra, Ton, Jonathan Grainger and Walter J.B. van Heuven. 1999. "Recognition of cognates and interlingual homographs: The neglected role of phonology", *Journal of Memory and Language* 41: 496–518.
- Dijkstra, Ton, Mark Timmermans and Herbert Schriefers. 2000. "On being blinded by your other language: Effects of task demands on interlingual homograph recognition", *Journal of Memory and Language* 42: 445–464.
- Dijkstra, Ton, Walter J.B. van Heuven and Jonathan Grainger. 1998b. "Simulating crosslanguage competition with the bilingual interactive activation model", *Psychologica Belgica* 38: 177–196.
- Dijkstra, Ton. 2007. "The multilingual lexicon", in: Gareth Gaskell (ed.), *Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, 251–263.
- Dronkers, Nina F., Brenda B. Redfern and Carl A. Ludy. 1995. "Lesion localization in chronic Wernicke's aphasia", *Brain and Language* 51, 1: 62–65.
- Duñabeitia, Jon Andoni, Maria Dimitropoulou, Oxel Uribe-Etxebarria, Itziar Laka and Manuel Carreiras. 2010. "Electrophysiological correlates of the masked translation priming effect with highly proficient simultaneous bilinguals", *Brain Research* 1359: 142–154.
- Duyck, Wouter, Dieter Vanderelst, Timothy Desmet and Robert J. Hartsuiker. 2008. "The frequency effect in second-language visual word recognition", *Psychonomic Bulletin and Review* 15, 4: 850–855.
- Duyck, Wouter, Eva van Asche, Denis Drieghe and Robert Hartsuiker. 2007. "Visual word recognition by bilinguals in a sentence context: Evidence for non-selective lexical access", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33: 663–679.
- Evans, Vyvyan and Melanie Green. 2006. *Cognitive linguistics: An introduction*. Edinburgh: Edinburgh University Press.

- Faust, Miriam and Allon Kahane. 2002. "Priming summation in the cerebral hemispheres: Evidence from semantically convergent and semantically divergent primes", *Neuropsychologia* 40: 892–901.
- Faust, Miriam and Christine Chiarello. 1998. "Sentence context and lexical resolution by the two hemispheres", *Neuropsychologia* 36: 827–836.
- Fawson, Parker C. and D. Ray Reutzel. 1994. "Comprehending metaphor: Using a salient characteristic analysis technique (SCAT)", *Reading Horizons* 34, 4: 356–368.
- Federmeier, Kara D. and Marta Kutas. 1999. "A rose by any other name: Long term memory structure and sentence processing", *Journal of Memory and Language* 41: 469–495.
- Federmeier, Kara D. and Marta Kutas. 2002. "Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres", *Neuropsychologia* 40, 7: 730–747.
- Federmeier, Kara D., Edward W. Wlotko, Esmeralda De Ochoa-Dewald and Marta Kutas. 2007. "Multiple effects of sentential constraint on word processing", *Brain Research* 1146: 75–84.
- Federmeier, Kara D., Heinke Mai and Marta Kutas. 2005. "Both sides get the point: Hemispheric sensitivities to sentential constraints", *Memory and Cognition* 33, 5: 871–886.
- Ferretti, Todd R., Christopher A. Schwint and Albert N. Katz. 2007. "Electrophysiological and behavioral measures of the influence of literal and figurative contextual constraints on proverb comprehension", *Brain and Language* 101, 1: 38–49.
- Ferstl, Evelyn C. and Detlef Y. von Cramon. 2002. "What does the frontomedian cortex contribute to language processing: Coherence or Theory of Mind?", *NeuroImage* 17: 1599–1612.
- Finnigan, Simon, Michael S. Humphreys, Simon Dennis and Gina Geffen. 2002. "ERP old/new effects: Memory strength and decisional factor(s)", *Neuropsychologia* 40: 2288–2304.
- Fischler, Ira S. and Paul A. Bloom. 1985. "Effects of constraint and validity of sentence contexts on lexical decisions", *Memory and Cognition* 13, 2: 128–139.
- Fogelin, Robert J. 1988. *Figuratively speaking*. New Haven: Yale University Press.

- Ford, Judith M., Walton T. Roth, Stanley J. Dirks and Bert S. Kopell. 1973. "Evoked potentials correlates of signal recognition between and within modalities", *Science* 181, 4098: 465–466.
- Forgács, Bálint, Megan D. Bardolph, Ben D. Amsel, Katherine A. DeLong and Marta Kutas. 2015. "Metaphors are physical and abstract: ERPs to metaphorically modified nouns resemble ERPs to abstract language", *Frontiers in Human Neuroscience* 9: 1–11.
- Foucart, Alice, Clara D. Martin, Eva M. Moreno and Albert Costa. 2014. "Can bilinguals see it coming? Word anticipation in L2 sentence reading", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40, 5: 1461–1469.
- Freunberger, Roman, Wolfgang Klimesch, Michael Doppelmayr and Yvonne Höller. 2007. "Visual P2 component is related to theta phase-locking", *Neuroscience Letters* 426, 3: 181–186.
- Friederici, Angela D., Anja Hahne and Douglas Saddy. 2002. "Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair", *Journal of Psycholinguistic Research* 31: 45–63.
- Friederici, Angela D., D. Yves von Cramon and Sonja A. Kotz. 1999. "Language related brain potentials in patients with cortical and subcortical left hemisphere lesions", *Brain* 122: 1033–1047.
- Friedman, David and Ray Johnson. 2000. "Event-related potential (ERP) studies of memory encoding and retrieval: A selective review", *Microscopy Research and Technique* 51: 6–28.
- Friedman, David, Herbert G. Vaughan and Niki Erlenmeyer-Kimling. 1978. "Stimulus and response related components of the late positive complex in the visual discrimination tasks", *Electroencephalography and Clinical Neurophysiology* 45: 319–330.
- Frisson, Steven and Martin J. Pickering. 2001. "Obtaining a figurative interpretation of a word: Support for underspecification", *Metaphor and Symbol* 16: 149–171.
- Frisson, Steven. 2009. "Semantic underspecification in language processing", *Language and Linguistic Compass* 3: 111–127.
- Garrod, Simon. 2006. "Psycholinguistic research methods", in: Keith Brown (ed.), *Encyclopedia of language and linguistics*. New York: Elsevier Science, 251–257.

- Gentner, Dedre and Arthur B. Markman. 1997. "Structure mapping in analogy and similarity", *American Psychologist* 52, 1: 45–56.
- Gentner, Dedre and Brian F. Bowdle. 2001. "Convention, form, and figurative language processing", *Metaphor and Symbol* 16, 3: 223–247.
- Gentner, Dedre and Phillip Wolff. 1997. "Alignment in the processing of metaphor", *Journal of Memory and Language* 37, 3: 331–355.
- Gentner, Dedre, Brian F. Bowdle, Phillip Wolff and Consuelo Boronat. 2001. "Metaphor is like analogy", in: Dedre Gentner, Keith J. Holyoak and Boicho N. Kokinov (eds.), *The analogical mind: Perspectives from cognitive science*. Cambridge: MIT Press, 199–253.
- Gentner, Dedre. 1983. "Structure-mapping: A theoretical framework for analogy", *Cognitive Science* 7, 2: 155–170.
- Geyer, Alexandra, Phillip J. Holcomb, Katherine J. Midgley and Jonathan Grainger. 2011. "Processing words in two languages: An event-related brain potential study of proficient bilinguals", *Journal of Neurolinguistics* 24, 3: 338–351.
- Gibbs, Raymond W. 1994a. "Figurative thought and figurative language", in: Morton A. Gernsbacher (ed.), *Handbook of psycholinguistics*. San Diego: Academic Press, 411–443.
- Gibbs, Raymond W. 1994b. *The poetics of mind*. Cambridge: Cambridge University Press.
- Gibbs, Raymond W. 2001. "Evaluating contemporary models of figurative language understanding", *Metaphor and Symbol* 16, 3: 317–333.
- Gibbs, Raymond W. 2002. "A new look at literal meaning in understanding what is said and implicated", *Journal of Pragmatics* 34: 457–486.
- Gibbs, Raymond W. and Herbert L. Colston. 2006. "Figurative language", in: Matthew Traxler and Morton A. Gernsbacher (eds.), *Handbook of psycholinguistics: Second Edition*. San Diego: Academic Press, 835–856.
- Gibbs, Raymond W. and Herbert L. Colston. 2012. *Interpreting figurative meaning*. New York: Cambridge University Press.
- Gibbs, Raymond W., Jennifer Bogdonovich, Jeffrey Sykes and Dale Barr. 1997. "Metaphor in idiom comprehension", *Journal of Memory and Language* 37: 141–154.

- Giora, Rachel, Eran Zaidel, Nachum Soroker, Gila Batori and Asa Kasher. 2000. "Differential effects of right and left hemispheric damage on understanding sarcasm and metaphor", *Metaphor and Symbol* 15: 63–83.
- Giora, Rachel. 2002. "Literal vs. figurative language: Different or equal?", *Journal of Pragmatics* 34, 4: 487–506.
- Giora, Rachel. 2003. *On our mind: Salience, context, and figurative language*. New York: Oxford University Press.
- Glucksberg, Sam and Catrinel Haught. 2006. "On the relation between metaphor and simile: When comparison fails", *Mind and Language* 21, 3: 360–378.
- Glucksberg, Sam. 2001. *Understanding figurative language: From metaphors to idioms*. New York: Oxford University Press.
- Glucksberg, Sam. 2003. "The psycholinguistics of metaphor", *Trends in Cognitive Sciences* 7, 2: 92–96.
- Goel, Vinod, Brian Gold, Shitij Kapur and Sylvain Houle. 1997. "The seats of reason? An imaging study of deductive and inductive reasoning", *NeuroReport* 8: 1305–1310.
- Goldstein, Abraham, Yossi Arzouan and Miriam Faust. 2012. "Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization", *Brain and Language* 123, 2: 137–142.
- Gollan, Tamar H., Timothy J. Slattery, Diane Goldenberg, Eva van Assche, Wouter Duyck and Keith Rayner. 2011. "Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis", *Journal of Experimental Psychology: General* 140, 2: 186–209.
- Gollan, Tamar, Kenneth L. Forster and Ram Frost. 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: 1122–1139.
- Gomez, Pablo, Ratcliff Roger and Manuel Perea. 2007. "A model of the go/no-go task", *Journal of Experimental Psychology: General* 136, 3: 389–413.
- Gordon, Barry and Alfonso Caramazza. 1982. "Lexical decision for open- and closed-class words: Failure to replicate differential frequency sensitivity", *Brain and Language* 15: 143–160.

- Górska, Elżbieta. 2009. "On the diversity of linguistic evidence for conceptual metaphor", *Studia Anglica Posnaniensa* 45, 2: 81–106.
- Gravetter, Frederick J. and Lori-Ann B. Forzano. 2012. *Research methods for the behavioral sciences*. Belmont, CA: Wadsworth.
- Green, David W. 2008. "Bilingual aphasia: Adapted language networks and their control", *Annual Review of Applied Linguistics* 28: 1–24.
- Green, David W. and Jubin Abutalebi. 2013. "Language control in bilinguals: The adaptive control hypothesis", *Journal of Cognitive Psychology* 25: 515–530.
- Grice, Paul. 1975. "Logic and conversation", in: Peter Cole and Jerry L. Morgan (eds.), *Syntax and semantics: Speech acts. Vol 3*. New York: Academic Press, 51–58.
- Grosjean, François. 1998. "Studying bilingualism: Methodological and conceptual issues", *Bilingualism: Language and Cognition* 1: 131–149.
- Guerrera-Mosquera, Carlos, Armando Malanda Trigueros and Angela Navia-Vazquez. 2012. "EEG signal processing for epilepsy", in: Dejan Stevanovic (ed.), *Epilepsy – histological, electroencephalographic and psychological aspects*. Rijeka: InTech, 49–74.
- Gunter, Thomas C., Angela D. Friederici and Herbert Schriefers. 2000. "Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction", *Journal of Cognitive Neuroscience* 12, 4: 556–568.
- Guo, Taomei, Maya Misra, Joyce W. Tam and Judith F. Kroll. 2012. "On the time course of accessing meaning in a second language: An electrophysiological and behavioral investigation of translation recognition.", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38, 5: 1165–1186.
- Hagoort, Peter, Colin M. Brown and Jolanda Groothusen. 1993. "The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing", *Language and Cognitive Processes* 8, 4: 439–483.
- Hagoort, Peter. 2003. "Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations", *Journal of Cognitive Neuroscience* 15: 883–899.
- Haught, Catrinel. 2013. "A tale of two tropes: How metaphor and simile differ", *Metaphor and Symbol* 28, 4: 254–274.
- Hauk, Olaf and Friedemann Pulvermüller. 2004. "Effects of word length and frequency on the human event-related potential", *Clinical Neurophysiology* 115: 1090–1103.

- Heidlmayr, Karin, Barbara Hemforth, Sylvain Moutier and Frédéri Isel. 2015. “Neurodynamics of executive control processes in bilinguals: Evidence from ERP and source reconstruction analyses”, *Frontiers in Psychology* 6: 1–17.
- Heredia, Roberto R. 2008. “Mental models of bilingual memory”, in: Jeanette Altarriba and Roberto R. Heredia (eds.), *An introduction to bilingualism: Principles and processes*. New York: Taylor & Francis Group, LLC, 39–70.
- Heredia, Roberto R. and Anna B. Cieślicka. 2014. “Bilingual memory storage: Compound-coordinate and derivatives”, in: Roberto R. Heredia and Jeanette Altarriba (eds.), *Foundations of bilingual memory*. New York, NY: Springer New York, 11–39.
- Heredia, Roberto R. and Anna B. Cieślicka. 2016. “Metaphoric reference: An eye movement analysis of Spanish–English and English–Spanish bilingual readers”, *Frontiers in Psychology* 7: 1–10.
- Heredia, Roberto R. and Jeffrey M. Brown. 2006. “Bilingual memory”, in: Tej K. Bhatia and William C. Ritchie (eds.), *The handbook of bilingualism*. Oxford: Blackwell Publishing, 225–249.
- Holcomb, Philip J., Jonathan Grainger and Tim O’Rourke. 2002. “An electrophysiological study of the effects of orthographic neighborhood size on printed word perception”, *Journal of Cognitive Neuroscience* 14, 6: 938–950.
- Hoshino, Noriko and Guillaume Thierry. 2012. “Do Spanish-English bilinguals have their fingers in two pies - or is it their toes? An electrophysiological investigation of semantic access in bilinguals”, *Frontiers in Psychology* 3, 9: 1–6.
- Hoshino, Noriko, Katherine J. Midgley, Phillip J. Holcomb and Jonathan Grainger. 2010. “An ERP investigation of masked cross-script translation priming”, *Brain Research* 1344: 159–172.
- Hoversten, Liv J., Trevor Brothers, Tamara Y. Swaab and Matthew J. Traxler. 2005. “Language membership identification precedes semantic access: Suppression during bilingual word recognition”, *Journal of Cognitive Neuroscience* 27: 2108–2116.
- Hoversten, Liv J., Trevor Brothers, Tamara Y. Swaab and Matthew J. Traxler. 2017. “Early processing of orthographic language membership information in bilingual visual word recognition: Evidence from ERPs”, *Neuropsychologia* 103: 183–193.

- Huettel, Scott A. and Gregory McCarthy. 2004. "What is odd in the oddball task? Prefrontal cortex is activated by dynamic changes in response strategy", *Neuropsychologia* 42: 379–386.
- Indefrey, Peter. 2006. "A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean?", *Language Learning* 56: 279–304.
- Ivanko, Stacey L. and Penny M. Pexman. 2003. "Context incongruity and irony processing", *Discourse Processes* 35: 241–279.
- Jankowiak, Katarzyna and Karolina Rataj. 2017. "The N400 as a window into lexico-semantic processing in bilingualism", *Poznań Studies in Contemporary Linguistics* 53, 1: 119–156.
- Jankowiak, Katarzyna, Karolina Rataj and Ryszard Naskręcki. 2017. "To electrify bilingualism: Electrophysiological insights into bilingual metaphor comprehension", *PLoS ONE* 12, 4: 1–30.
- Javier, Rafael Art. 2007. *The bilingual mind: Thinking, feeling and speaking in two languages*. New York: Springer.
- Jiang, Nan and Kenneth I. Forster. 2001. "Cross-language priming asymmetries in lexical decision and episodic recognition", *Journal of Memory and Language* 44: 32–51.
- Jiang, Nan. 1999. "Testing processing explanations for the asymmetry in masked cross-language priming", *Bilingualism: Language and Cognition* 2: 59–75.
- Jiang, Xiaoming, Yingying Tan and Xiaolin Zhou. 2009. "Processing the universal quantifier during sentence comprehension: ERP evidence", *Neuropsychologia* 47, 8: 1799–1815.
- Jiang, Zhongqing. 2015. "Event-related potentials", in: Nidal Kamel and Aamir Saeed Malik (eds.), *EEG/ERP analysis: Method and applications*. London: Taylor & Francis Group, 73–89.
- Johnson, Andrew T. 1996. "Comprehension of metaphors and similes: A reaction time study", *Metaphor and Symbol Activity* 11, 2: 145–159.
- Johnson, Ray, Kurt Kreiter, Britt Russo and John Zhu. 1998. "A spatio-temporal analysis of recognition-related event-related brain potentials", *International Journal of Psychophysiology* 29: 83–104.
- Jones, Lara L. and Zachary Estes. 2005. "Metaphor comprehension as attributive categorization", *Journal of Memory and Language* 53: 110–124.

- Jones, Lara L. and Zachary Estes. 2006. "Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension", *Journal of Memory and Language* 55, 1: 18–32.
- Jouravlev, Olessia and Debra Jared. 2014. "Reading Russian–English homographs in sentence contexts: Evidence from ERPs", *Bilingualism: Language and Cognition* 17, 1: 153–168.
- Jung-Beeman, Mark. 2005. "Bilateral brain processes for comprehending natural language", *Trends in Cognitive Sciences* 9, 11: 512–518.
- Just, Marcel A., Patricia A. Carpenter and Jacqueline D. Wooley. 1982. "Paradigms and processes in reading comprehension", *Journal of Experimental Psychology: General* 111: 228–238.
- Kaan, Edith. 2007. "Event-related potentials and language processing: A brief overview", *Language and Linguistics Compass* 1, 6: 571–591.
- Kaiser, Elsi. 2013. "Experimental paradigms in psycholinguistics", in: Robert J. Podesva and Denyani Sharma (eds.), *Research methods in linguistics*. New York: Cambridge University Press, 135–168.
- Kamel, Nidal and Aamir Saeed Malik. 2015. *EEG/ERP analysis: Method and applications*. London: Taylor & Francis Group.
- Karakaş, Sirel and Erol Başar. 2006. "Models and theories of brain function in cognition within a framework of behavioral cognitive psychology", *International Journal of Psychophysiology* 60: 186–193.
- Katz, Albert N. and Todd R. Ferretti. 2001. "Moment-by-moment reading of proverbs in literal and nonliteral contexts", *Metaphor and Symbol* 16: 193–221.
- Kazmerski, Victoria A., Dawn G. Blasko and Banchiamlack G. Dessalegn. 2003. "ERP and behavioral evidence of individual differences in metaphor comprehension", *Memory and Cognition* 31, 5: 673–689.
- Keating, Gregory D. and Jill Jegerski. 2015. "Experimental designs in sentence processing research: A methodological review and user's guide", *Studies in Second Language Acquisition* 37: 1–32.
- Kecskes, Istvan. 2000. "A cognitive-pragmatic approach to situation-bound utterances", *Journal of Pragmatics* 32: 605–625.

- Kempen, Geragrd. 1998. "Sentence parsing", in: Angela D. Friederici (ed.), *Language comprehension: A biological perspective*. Berlin: Springer Berlin Heidelberg, 213–228.
- Kerkhofs, Roel, Ton Dijkstra, Dorothee J. Chwilla and Ellen R.A. de Bruijn. 2006. "Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects", *Brain Research* 1068, 1: 170–183.
- Kintsch, Walter and Anita R. Bowles. 2002. "Metaphor comprehension: What makes a metaphor difficult to understand?", *Metaphor and Symbol* 17, 4: 249–262.
- Kintsch, Walter. 1998. *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, Walter. 2001. "Predication", *Cognitive Science* 25: 173–202.
- Kintsch, Walter. 2008. "Symbol systems and perceptual representations", in: Manuel de Vega, Arthur Glenberg and Arthur Graesser (eds.), *Symbols and embodiment*. Oxford: Oxford University Press, 145–164.
- Kotz, Sonja A. and Kerrie Elston-Güttler. 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: 215–235.
- Kotz, Sonja A., Kathrin Rothermich and Maren Schmidt-Kassow. 2012. "Sentence comprehension in healthy and brain-damaged populations", in: Miriam Faust (ed.), *The handbook of the neuropsychology of language*. Malden: Blackwell Publishing, 760–777.
- Kounios, Josh and Phillip J. Holcomb. 1994. "Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, 4: 804–823.
- Kovelman, Iouia, Stephanie A. Baker and Laura-Ann Petitto. 2008. "Bilingual and monolingual brains compared: A functional Magnetic Resonance Imaging investigation of syntactic processing and a possible neural signature of bilingualism", *Journal of Cognitive Neuroscience* 20, 1: 153–169.
- Kroll, Judith F. and Erika Stewart. 1994. "Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations", *Journal of Memory and Language* 33: 149–174.
- Kroll, Judith F. and Natasha Tokowicz. 2005. "Models of bilingual representation and processing: Looking back and to the future", in: Judith F. Kroll and Annette M.

- B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press, 531–553.
- Kroll, Judith F. and Paole E. Dussias. 2006. “The comprehension of words and sentences in two languages”, in: Tej K. Bhatia and William C. Ritchie (eds.), *The handbook of bilingualism*. Oxford: Blackwell Publishing Ltd, 169–200.
- Kroll, Judith F., Erica Michael, Natasha Tokowicz and Robert Dufour. 2002. “The development of lexical fluency in a second language”, *Second Language Research* 18, 2: 137–171.
- Kroll, Judith F., Janet G. van Hell, Natasha Tokowicz and David W. Green. 2010. “The Revised Hierarchical Model: A critical review and assessment”, *Bilingualism: Language and Cognition* 13, 3: 373–381.
- Kropotov, Juri. D. 2009. *Quantitative EEG, event-related potentials and neurotherapy*. London: Academic Press.
- Kutas, Marta and Kara D. Federmeier. 2000. “Electrophysiology reveals semantic memory use in language comprehension”, *Trends in Cognitive Sciences* 4, 12: 463–470.
- Kutas, Marta and Kara D. Federmeier. 2011. “Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP)”, *Annual Review of Psychology* 62, 1: 621–647.
- Kutas, Marta and Steven A. Hillyard. 1980. “Reading senseless sentences: Brain potentials reflect semantic incongruity”, *Science* 207: 203–205.
- Kutas, Marta and Steven A. Hillyard. 1983. “Event-related brain potentials to grammatical errors and semantic anomalies”, *Memory and Cognition* 11, 5: 539–550.
- Kutas, Marta and Steven A. Hillyard. 1984. “Brain potentials during reading reflect word expectancy and semantic association”, *Nature* 307: 161–163.
- Kutas, Marta, Cyma K. Van Petten and Robert Kluender. 2006. “Psycholinguistics electrified II (1994-2005)”, in: Matthew Traxler and Morton Ann Gernsbacher (eds.), *Handbook of psycholinguistics: 2nd Edition*. San Diego: Academic Press, 659–724.
- Kutas, Marta, Gregory McCarthy and Emanuel Donchin. 1977. “Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time”, *Science* 197, 4305: 792–795.

- Kutas, Marta, Steven A. Hillyard and Michael Gazzaniga. 1988. "Processing of semantic anomaly by right and left hemispheres of commissurotomy patients", *Brain* 111: 553–576.
- La Heij, Wido, Andre Hooglander, Robert Kerling and Esther van der Velden. 1996. "Nonverbal context effects in forward and backward word translation: Evidence for concept mediation", *Journal of Memory and Language* 35, 5: 648–665.
- Lai, Vicky Tzuyin and Tim Curran. 2013. "ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors", *Brain and Language* 127, 3: 484–496.
- Lai, Vicky Tzuyin, Tim Curran and Lise Menn. 2009. "Comprehending conventional and novel metaphors: An ERP study", *Brain Research* 1284: 145–155.
- Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago: The University of Chicago Press.
- Lalor, Erin and Kim Kirsner. 2000. "Cross-lingual transfer effects between English and Italian cognates and noncognates", *International Journal of Bilingualism* 4: 385–398.
- Landauer, Thomas and Susan Dumais. 1997. "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge", *Psychological Review* 104: 211–240.
- Lemhöfer, Kristin, Ton Dijkstra, Herbert Schriefers, R. Harald Baayen, Jonathan Grainger and Pienie Zwitserlood. 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.
- Lenneberg, Eric H. 1967. *Biological foundations of language*. New York: Wiley.
- Leynes, Andrew P., Heather Bruett, Jenna Krizan and Ana Veloso. 2017. "What psychological process is reflected in the FN400 even-related potential component?", *Brain and Cognition* 113: 142–154.
- Lijewska, Agnieszka and Agnieszka Chmiel. 2015. "Cognate facilitation in sentence context – translation production by interpreting trainees and non-interpreting trilinguals", *International Journal of Multilingualism* 12, 3: 358–375.
- Likert, Rensis. 1932. "A technique for the measurement of attitudes", *Archives of Psychology* 140: 1–55.

- Littlemore, Jeannette and Graham Low. 2006. *Figurative thinking and foreign language learning*. New York: Palgrave Macmillan.
- Lopes da Silva, Fernando. 2010. "EEG: Origin and measurement", in: Christoph Mulert and Louis Lemieux (eds.), *EEG-fMRI: Psychological basis, technique, and applications*. Berlin: Springer, 19–38.
- Lopes da Silva, Fernando. 2013. "EEG and MEG: Relevance to Neuroscience", *Neuron* 80: 1112–1128.
- Lü, Junmei, Lijuan Liang and Bagouo Chen. 2017. "The effect of executive control ability on the comprehension of second language metaphors", *International Journal of Bilingualism*: 1–15.
- Luck, Steven J. 2005. *An introduction to the event-related potential technique*. Massachusetts: Massachusetts Institute of Technology.
- Luck, Steven J. and Steven A. Hillyard. 1994. "Electrophysiological correlates of feature analysis during visual search", *Psychophysiology* 31, 3: 291–308.
- Marian, Viorica and Michael Spivey. 2003. "Competing activation in bilingual language processing: Within- and between-language competition", *Bilingualism: Language and Cognition* 6, 2: 97–115.
- Marian, Viorica and Ulric Neisser. 2000. "Language-dependent recall of autobiographical memories", *Journal of Experimental Psychology: General* 129, 3: 361–368.
- Martin, Clara D., Albert Costa, Benjamin Dering, Noriko Hoshino, Yan Jing Wu and Guillaume Thierry. 2012. "Effects of speed of word processing on semantic access: The case of bilingualism", *Brain and Language* 120, 1: 61–65.
- Mashal, Nira, Katy Borodkin, Omer Maliniak and Miriam Faust. 2015. "Hemispheric involvement in native and non-native comprehension of conventional metaphors", *Journal of Neurolinguistics* 35: 96–108.
- Mashal, Nira, Miriam Faust, Talma Hendler and Mark Jung-Beeman. 2007. "An fMRI investigation of the neural correlates underlying the processing of novel metaphorical expressions", *Brain and Language* 100, 2: 115–126.
- Mashal, Nira, Miriam Faust, Talma Hendler and Mark Jung-Beeman. 2008. "Hemispheric differences in processing the literal interpretation of idioms: Converging evidence from behavioral and fMRI studies", *Cortex* 44, 7: 848–860.
- Mashal, Nira. 2013. "The role of working memory in the comprehension of unfamiliar and familiar metaphors", *Language and Cognition* 5, 4: 409–436.

- Matlock, Teene and Roberto R. Heredia. 2002. "Understanding phrasal verbs in monolinguals and bilinguals", in: Roberto R. Heredia and Jeanette Altarriba (eds.), *Bilingual sentence processing*. Amsterdam: Elsevier, 251–274.
- Matsukawa, Junko, Joan Gay Snodgrass and Glen M. Doniger. 2005. "Conceptual versus perceptual priming in incomplete picture identification", *Journal of Psycholinguistic Research* 34, 6: 515–540.
- McClelland, James L. and David E. Rumelhart. 1981. "An interactive activation model of context effects in letter perception: Part 1. An account of basic findings", *Psychological Review*: 88, 5: 375–407.
- McElree, Brian and Johanna Nordlie. 1999. "Literal and figurative interpretations are computed in equal time", *Psychonomic Bulletin & Review* 6, 3: 486–494.
- McGlone, Matthew S. 1996. "Conceptual metaphors and figurative language interpretation: Food for thought?", *Journal of Memory and Language* 35, 4: 544–565.
- Mechelli, Andrea, Jenny T. Crinion, Uta Noppeney, John O'Doherty, Josh Ashburner, Richard S. Frackowiak and Cathy J. Price. 2004. "Neurolinguistics: Structural plasticity in the bilingual brain", *Nature* 431: 757.
- Mecklinger, Alex. 2000. "Interfacing mind and brain: A neurocognitive model of recognition memory", *Psychophysiology* 37: 565–582.
- Meyer, David E. and Roger W. Schvaneveldt. 1971. "Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations", *Journal of Experimental Psychology* 90, 2: 227–234.
- Midgley, Katherine J., Phillip J. Holcomb and Jonathan Grainger. 2009a. "Masked repetition and translation priming in second language learners: A window on the time-course of form and meaning activation using ERPs", *Psychophysiology* 46, 3: 551–565.
- Midgley, Katherine J., Phillip J. Holcomb and Jonathan Grainger. 2009b. "Language effects in second language learners and proficient bilinguals investigated with event-related potentials", *Journal of Neurolinguistics* 22: 281–300.
- Midgley, Katherine J., Phillip J. Holcomb, Walter J.B. van Heuven and Jonathan Grainger. 2008. "An electrophysiological investigation of cross-language effects of orthographic neighborhood", *Brain Research* 1246: 123–135.
- Mitchell, Mark L. and Janina M. Jolley. 2010. *Research design explained*. Belmont: Wadsworth, Cengage Learning.

- Moldovan, Cornelia D., Josep Demestre, Pilar Ferré and Rosa Sánchez-Casas. 2016. “The role of meaning and form similarity in translation recognition in highly proficient balanced bilinguals: A behavioral and ERP study”, *Journal of Neurolinguistics* 37: 1–11.
- Monzó, Adelina Estévez and Manuel G. Calvo. 2002. “Context constraints, prior vocabulary knowledge and on-line inferences in reading”, *Psicothema* 14, 2: 357–362.
- Moreno, Eva M. and Marta Kutas. 2005. “Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals”, *Cognitive Brain Research* 22, 2: 205–220.
- Moreno, Eva M., Antoni Rodríguez-Fornells and Matti Laine. 2008. “Event-related potentials (ERPs) in the study of bilingual language processing”, *Journal of Neurolinguistics* 21, 6: 477–508.
- Mummery, Cath J., Karalyn Patterson, Cathy J. Price, Josh Ashburner, Richard S. Frackowiak and Josh R. Hodges. 2000. “A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory”, *Annals of Neurology* 47: 36–45.
- Münte, Thomas F., Hans-Jochen Heinze, Mike Matzke, Bernardina M. Wieringa and Sönke Johannes. 1998. “Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift”, *Neuropsychologia* 36: 217–226.
- New, Boris, Ludovic Ferrand, Christophe Pallier and Marc Brysbaert. 2006. “Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon project”, *Psychonomic Bulletin & Review* 13, 1: 45–52.
- Newman, Aaron J., Antoine Tremblay, Emily S. Nichols, Helen J. Neville and Michael T. Ullman. 2012. “The influence of language proficiency on lexical semantic processing in native and late learners of English”, *Journal of Cognitive Neuroscience* 24, 5: 1205–1223.
- Newman, Sharlene D., Donghoon Lee and Kristen L. Ratliff. 2009. “Off-line sentence processing: What is involved in answering a comprehension probe?”, *Human Brain Mapping* 30, 8: 2499–2511.
- Norris, Dennis. 2013. “Models of visual word recognition”, *Trends in Cognitive Sciences* 17, 10: 517–524.

- Oberauer, Klaus, Anke Demmrich, Ulrich Mayr and Reinhold Kliegl. 2001. "Dissociating retention and access in working memory: An age-comparative study of mental arithmetic", *Memory and Cognition* 29, 1: 18–33.
- Oganian, Yulia, Markus Conrad, Arash Aryani, Hauke R. Heekeren and Katharina Spalek. 2016. "Interplay of bigram frequency and orthographic neighborhood statistics in language membership decision", *Bilingualism: Language and Cognition* 19: 578–596.
- Oldfield, Richard C. 1971. "The assessment and analysis of handedness: The Edinburgh inventory", *Neuropsychologia* 9, 1: 97–113.
- Onysko, Alexander. 2016. "Enhanced creativity in bilinguals? Evidence from meaning interpretations of novel compounds", *International Journal of Bilingualism* 20, 3: 315–334.
- Ortony, Andrew. 1979. "Beyond literal similarity", *Psychological Review* 86: 161–180.
- Osterhout, Lee and Janet Nicol. 1999. "On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies", *Language and Cognitive Processes* 14, 3: 283–317.
- Paivio, Allan and Alain Desrochers. 1980. "A dual-coding approach to bilingual memory", *Canadian Journal of Psychology* 34, 4: 388–399.
- Paivio, Allan. 1991. "Dual coding theory: Retrospect and current status", *Canadian Journal of Psychology* 45: 255–287.
- Paivio, Allan. 2014. "Bilingual dual-coding theory and memory", in: Roberto R. Heredia and Jeanette Altarriba (eds.), *Foundations of bilingual memory*. New York: Springer, 41–62.
- Palinkas, Istvan. 2013. "Irony and the Standard Pragmatic Model", *International Journal of English Language* 3, 5: 14–19.
- Paller, Ken A., Joel L. Voss and Stephan G. Boehm. 2007. "Validating neural correlates of familiarity", *Trends in Cognitive Sciences* 11, 6: 243–250.
- Paller, Ken A., Marta Kutas and Andrew R. Mayes. 1987. "Neural correlates of encoding in an incidental learning paradigm", *Electroencephalography and Clinical Neurophysiology* 67: 360–371.

- Patterson, Karalyn and John R. Hodges. 2000. "Semantic dementia: One window on the structure and organization of semantic memory", in: Laird S. Cermak (ed.), *Revised handbook of neuropsychology: Memory and its disorders*. Amsterdam: Elsevier Science B.V., 313–335.
- Patterson, Karalyn and Josh R. Hodges. 2000. "Semantic dementia: One window on the structure and organisation of semantic memory", in: Laird S. Cermak (ed.), *Handbook of neuropsychology*. Amsterdam: Elsevier, 313–333.
- Paulmann, Silke, Zainab Ghareeb-Ali and Claudia Felser. 2015. "Neuropsychological markers of phrasal verbs processing: Evidence from L1 and L2 speakers", in: Roberto R. Heredia and Anna B. Cieřlicka (eds.), *Bilingual figurative language processing*. Cambridge: Cambridge University Press, 197–213.
- Peeters, David, Ton Dijkstra and Jonathan Grainger. 2013. "The representation and processing of identical cognates by late bilinguals: RT and ERP effects", *Journal of Memory and Language* 68, 4: 315–332.
- Pfaff, Kerry, Raymond W. Gibbs and Michael Johnson. 1997. "Metaphor in using and understanding euphemisms and dysphemisms", *Applied Psycholinguistics* 18: 59–83.
- Phillips, Natalie A., Norman Segalowitz, Irena O'Brien and Naomi Yamasaki. 2004. "Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials", *Journal of Neurolinguistics* 17, 2: 237–262.
- Pickering, Martin J. and Steven Frisson. 2001. "Processing ambiguous verbs: Evidence from eye movements", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27, 2: 556–573.
- Pierce, Russell S., Rick MacLaren and Dan L. Chiappe. 2010. "The role of working memory in the metaphor interference effect", *Psychonomic Bulletin & Review* 17, 3: 400–404.
- Potter, Mary C., Kwok-Fai So, Barbara von Eckardt and Laurie B. Feldman. 1984. "Lexical and conceptual representation in beginning and proficient bilinguals", *Journal of Verbal Learning and Verbal Behavior* 23, 1: 23–38.
- Proverbio, Alice Mado, Barbara Āok and Alberto Zani. 2002. "Electrophysiological measures of language processing in bilinguals", *Journal of Cognitive Neuroscience* 14, 7: 994–1017.

- Pynte, Joel, Mireille Besson, Fabrice-Henri Robichon and Jézabel Poli. 1996. "The time-course of metaphor comprehension: An event-related potential study", *Brain and Language* 55, 3: 293–316.
- Randall, Mick. 2007. *Memory, psychology and second language learning*. Philadelphia: John Benjamins Publishing Company.
- Ransdell, Sarah E. and Ira Fischler. 1987. "Memory in monolingual mode: When are bilinguals at a disadvantage?", *Journal of Memory and Language* 26: 392–405.
- Rataj, Karolina, Anna Przekoracka-Krawczyk and Rob H.J. van der Lubbe. 2018. "On understanding creative language: The late positive complex and novel metaphor comprehension", *Brain Research* 1678: 231–244.
- Rataj, Karolina. 2014. "Surfing the brainwaves of metaphor comprehension", *Poznań Studies in Contemporary Linguistics* 50, 1: 55–73.
- Récanati, François. 1995. "The alleged priority of literal interpretation", *Cognitive Science* 19: 207–232.
- Regel, Stefanie, Seana Coulson and Thomas C. Gunter. 2010. "The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony", *Brain Research* 1311: 121–135.
- Regel, Stefanie, Thomas C. Gunter and Angela D. Friederici. 2011. "Isn't it ironic? An electrophysiological exploration of figurative language processing", *Journal of Cognitive Neuroscience* 23, 2: 277–293.
- Ribot, Théodule. 1882. *Diseases of memory: An essay in the positive psychology*. London: Paul.
- Richards, Larry G. 1976. "Concreteness as a variable in word recognition", *American Journal of Psychology* 89, 4: 707–718.
- Ritter, Walter, Richard Simson and Herbert G. Vaughan. 1972. "Association cortex potentials and reaction time in auditory discrimination", *Electroencephalography and Clinical Neurophysiology* 33, 6: 547–555.
- Roberts, Leah and Claudia Felser. 2011. "Plausibility and recovery from garden paths in L2 sentence processing", *Applied Psycholinguistics* 32: 299–331.
- Rogers, Timothy T. 2009. "Connectionist models", in: Marc D. Binder, Nobutaka Hirokawa and Uwe Windhorst (eds.), *Encyclopedia of neuroscience*. Berlin: Springer-Verlag Berlin Heidelberg, 75–82.

- Ruchkin, Daniel R., Ray Johnson Jr., David Mahaffey and Samuel Sutton. 1988. "Toward a functional categorization of slow waves", *Psychophysiology* 25, 3: 339–353.
- Rugg, Michael D. and Tim Curran. 2007. "Event-related potentials and recognition memory", *Trends in Cognitive Sciences* 11: 251–257.
- Rutter, Barbara, Sören Kröger, Holger Hill, Sabine Windmann, Christiane Hermann and Anna Abraham. 2012. "Can clouds dance? Part 2: An ERP investigation of passive conceptual expansion", *Brain and Cognition* 80, 3: 301–310.
- Sahlin, Brooke H., Matthew G. Harding and John G. Seamon. 2005. "When do false memories cross language boundaries in English-Spanish bilinguals?", *Memory and Cognition* 33, 8: 1414–1421.
- Salthouse, Timothy A. and Trey Hedden. 2002. "Interpreting reaction time measures in between-group comparisons", *Journal of Clinical and Experimental Neuropsychology* 24, 7: 858–872.
- Sanei, Saeid and Jonathon Chambers. 2007. *EEG signal processing*. Chichester: John Wiley & Sons Ltd.
- Sanford, Anthony J., Harmut Leuthold, Jason Bohan and Alison J. S. Sanford. 2011. "Anomalies at the borderline of awareness: An ERP study", *Journal of Cognitive Neuroscience* 23, 3: 514–523.
- Schacht, Annkathrin, Werner Sommer, Olga Shmuilovich, Pilar Casado Martinez and Manuel Martin-Loeches. 2014. "Differential task effects on N400 and P600 elicited by semantic and syntactic violations", *PLoS ONE* 9, 3: 1–7.
- Schmitz, John R. 2002. "Humor as a pedagogical tool in foreign language and translation courses", *Humor: International Journal of Humor Research* 15: 89–113.
- Schneider, Sabrina, Alexander M. Rapp, Florian B. Haeußinger, Lena H. Ernst, Friedrich Hamm, Andreas J. Fallgatter and Ann-Christine Ehlis. 2014. "Beyond the N400: Complementary access to early neural correlates of novel metaphor comprehension using combined electrophysiological and haemodynamic measurements", *Cortex* 53: 45–59.
- Schwanenflugel, Paula J., Katherine K. Harnishfeger and Randall W. Stowe. 1988. "Context availability and lexical decisions for abstract and concrete words", *Journal of Memory and Language* 27, 5: 499–520.

- Sereno, Sara C., Keith Rayner and Michael I. Posner. 1998. "Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials", *Neuroreport* 9: 2195–2200.
- Severens, Els and Robert J. Hartsuiker. 2009. "Is there a lexical bias effect in comprehension monitoring?", *Language and Cognitive Processes* 24, 6: 910–927.
- Shelton, Jose and Gideon Praveen Kumar. 2010. "Comparison between auditory and visual simple reaction times", *Neuroscience and Medicine* 1, 1: 30–32.
- Shibata, Midori, Akira Toyomura, Hiroki Motoyama, Hiroaki Itoh, Yasuhiro Kawabata and Junichi Abe. 2012. "Does simile comprehension differ from metaphor comprehension? A functional MRI study", *Brain and Language* 121, 3: 254–260.
- Sperber, Dan and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.
- Spivey, Michael J. and Viorica Marian. 1999. "Cross talk between native and second languages: Partial activation of an irrelevant lexicon", *Psychological Science* 10: 281–284.
- Spotorno, Nicola, Anne Cheylus, Jean-Baptiste Van Der Henst and Ira A. Noveck. 2013. "What's behind a P600? Integration operations during irony processing", *PLoS ONE* 8, 6: 1–10.
- Spruyt, Adriaan, Jan De Houwer and Dirk Hermans. 2009. "Modulation of automatic semantic priming by feature-specific attention allocation", *Journal of Memory and Language* 61: 37–54.
- Sternberg, Robert J. 1996. *Cognitive psychology*. Orlando: Hartcourt Brace College Publishers.
- Stuss, Donald T., Terence Picton, Anna M. Cerri, Edward E. Leech and Lee L. Stethem. 1992. "Perceptual closure and object identification: Electrophysiological responses to incomplete pictures", *Brain and Cognition* 19: 253–266.
- Su, Chang, Shuman Huang and Yijiang Chen. 2017. "Automatic detection and interpretation of nominal metaphor based on the theory of meaning", *Neurocomputing* 219: 300–311.
- Sunderman, Gretchen and Judith F. Kroll. 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.

- Swick, Diane, Marta Kutas and Robert T. Knight. 1998. "Prefrontal lesions eliminate the LPC but do not affect the N400 during sentence reading", *Journal of Cognitive Neuroscience Supplement* 29.
- Szewczyk, Jakub M. and Herbert Schriefers. 2011. "Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing", *Brain Research* 1368: 208–221.
- Talamas, Adrienne, Judith F. Kroll and Robert Dufour. 1999. "From form to meaning: Stages in the acquisition of second-language vocabulary", *Bilingualism: Language and Cognition* 2, 1: 45–58.
- Tang, Xuemei, Senqing Qi, Xiaojuan Jia, Botao Wang and Wei Ren. 2017a. "Comprehension of scientific metaphors: Complementary processes revealed by ERP", *Journal of Neurolinguistics* 42: 12–22.
- Tang, Xuemei, Senqing Qi, Xiaojuan Jia, Botao Wang and Wei Ren. 2017b. "The temporal dynamics underlying the comprehension of scientific metaphors and poetic metaphors", *Brain Research* 1655: 33–40.
- Tarter, Vivien C., Hilary Gomes, Boris Dubrovsky, Sophie Molholm and Rosemarie Vala Stewart. 2002. "Novel metaphors appear anomalous at least momentarily: Evidence from N400", *Brain and Language* 80, 3: 488–509.
- Tatum, William O., Aatif M. Husain, Selim R. Benbadis and Peter W. Kaplan. 2008. *Handbook of EEG interpretation*. USA: Demos Medical Publishing.
- Temnikova, Irina G. and Olga V. Nagel. 2015. "Effects of cognate and relatedness status on word recognition in Russian-English bilinguals of upper-intermediate and advanced levels", *Procedia – Social and Behavioral Sciences* 200: 381–386.
- Thierry, Guillaume and Yan J. Wu. 2007. "Brain potentials reveal unconscious translation during foreign-language comprehension", *Proceedings of the National Academy of Sciences* 104, 30: 12530–12535.
- Thornhill, Dianne E. and Cyma Van Petten. 2012. "Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components", *International Journal of Psychophysiology* 83, 3: 382–392.
- Utsumi, Akira and Maki Sakamoto. 2007. "Computational evidence for two-stage categorization as a process of adjective metaphor comprehension", *Proceedings of the Second European Cognitive Science Conference (EuroCogSci2007)*: 77–82.

- Utsumi, Akira and Maki Sakamoto. 2011. "Indirect categorization as a process of predicative metaphor comprehension", *Metaphor and Symbol* 26, 4: 299–313.
- Utsumi, Akira. 2007. "Interpretive diversity explains metaphor–simile distinction", *Metaphor and Symbol* 22, 4: 291–312.
- Utsumi, Akira. 2011. "Computational exploration of metaphor comprehension processes using a semantic space model", *Cognitive Science* 35, 2: 251–296.
- Vaid, Jyotsna and Cheryl French-Mestre. 2002. "Do orthographic cues aid language recognition? A laterality study with French-English bilinguals", *Brain and Language* 82: 47–53.
- Vaid, Jyotsna, Belen G. López and Francisco F. Martínez. 2015. "Linking the figurative to the creative: Bilingual's comprehension of metaphors, jokes, and remote associates", in: Roberto R. Heredia and Anna B. Cieślicka (eds.), *Bilingual figurative language processing*. New York: Cambridge University Press, 43–68.
- van Berkum, Jos J.A. 2009. "The 'neuropragmatics' of simple utterance comprehension: An ERP review", in: Uli Sauerland and Kazuko Yatsushiro (eds.), *Semantics and pragmatics: From experiment to theory*. New York: Palgrave Macmillan, 276–316.
- Van de Moerendonk, Nan, Herman H.J. Kolk, Constance Th. W.M. Vissers and Dorothee J. Chwilla. 2010. "Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns", *Journal of Cognitive Neuroscience* 22: 67–82.
- Van der Meij, Maartje, Fernando Cuetos, Manuel Carreiras and Horacio A. Barber. 2011. "Electrophysiological correlates of language switching in second language learners: Electrophysiological correlates of language switching", *Psychophysiology* 48, 1: 44–54.
- van Hell, Janet G. and Ton Dijkstra. 2002. "Foreign language knowledge can influence native language performance in exclusively native contexts", *Psychonomic Bulletin & Review* 9: 780–789.
- van Herten, Marieke, Herman H. J. Kolk and Dorothee J. Chwilla. 2005. "An ERP study of P600 effects elicited by semantic anomalies", *Cognitive Brain Research* 22, 2: 241–255.
- van Heuven, Walter J.B. and Ton Dijkstra. 2010. "Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models", *Brain Research Review* 64: 104–122.

- van Heuven, Walter J.B., Herbert Schriefers, Ton Dijkstra and Peter Hagoort. 2008. "Language conflict in the bilingual brain", *Cerebral Cortex* 18: 2706–2716.
- van Heuven, Walter J.B., Ton Dijkstra and Jonathan Grainger. 1998. "Orthographic neighborhood effects in bilingual word recognition", *Journal of Memory and Language* 39: 458–483.
- van Kesteren, Ron, Ton Dijkstra and Koenraad de Smedt. 2012. "Markedness effects in Norwegian-English bilinguals: Task-dependent use of language-specific letters and bigrams", *Quarterly Journal of Experimental Psychology* 65, 11: 2129–2154.
- Van Petten, Cyma and Marta Kutas. 1991. "Influences of semantic and syntactic context on open and closed class words", *Memory and Cognition* 19: 95–112.
- VanRullen, Rufin. 2011. "Four common conceptual fallacies in mapping the time course of recognition", *Frontiers in Psychology* 2, 365: 1–6.
- Voss, Joel L. and Kara D. Federmeier. 2011. "FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing", *Psychophysiology* 48, 4: 532–546.
- Weber-Fox, Christine M. and Helen J. Neville. 1996. "Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers", *Journal of Cognitive Neuroscience* 8, 3: 231–156.
- Weiland, Hanna, Valentina Bambini and Petra B. Schumacher. 2014. "The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP", *Frontiers in Human Neuroscience* 8: 1–17.
- Weinreich, Uriel. 1953. *Languages in contact*. The Hague: Mouton.
- Whaley, Charles P. 1978. "Word-nonword classification time", *Journal of Verbal Learning and Verbal Behaviour* 17: 143–154.
- Wlotko, Edward D. and Kara D. Federmeier. 2007. "Finding the right word: Hemispheric asymmetries in the use of sentence context information", *Neuropsychologia* 45, 13: 3001–3014.
- Wolk, David A., Daniel L. Schacter, Melissa Lygizos, N. Mandu Sen, Phillip J. Holcomb, Kirk R. Daffner and Andrew E. Budson. 2006. "ERP correlates of recognition memory: Effects of retention interval and false alarms", *Brain Research* 1096: 148–162.

- Yang, Fan-Pei Gloria, Kailyn Bradley, Madiha Huq, Dai-Lin Wu and Daniel C. Krawczyk. 2013. "Contextual effects on conceptual blending in metaphors: An event-related potential study", *Journal of Neurolinguistics* 26, 2: 312–326.
- Zhao, Qian, Liang Yu and Yanqing Yang. 2014. "Correlation between receptive metaphorical competence and reading proficiency", *English Language Teaching* 7, 11: 168–181.
- Zharikov, Sergey and Dedre Gentner. 2002. "Why do metaphors seem deeper than similes", *Proceedings of the twenty-fourth Annual Conference of the Cognitive Science Society*: 976–981.
- Zysset, Stefan, Oswald Huber, Andrea Samson, Evelyn C. Ferstl and Detlef Y. von Cramon. 2003. "Functional specialization within the anterior medial prefrontal cortex: A functional Magnetic Resonance Imaging study with human subjects", *Neuroscience Letters* 335, 3: 183–186.

Appendix A: Electrophysiological experiments into the N400 and LPC responses in monolingual metaphoric language comprehension

| Authors | Year | Title | Materials | Task | Electrophysiological and behavioral ¹⁸ results |
|--------------|------|--|--|---------------------------|--|
| Pynte et al. | 1996 | <i>The time-course of metaphor comprehension: An event-related potential study</i> | Experiment 1: conventional metaphoric and literal sentences; Experiment 2: novel metaphoric, conventional metaphoric, and literal sentences; Experiment 3: novel metaphoric sentences embedded in irrelevant contexts, conventional metaphoric sentences embedded in relevant contexts; Experiment 4: novel metaphoric sentences embedded in relevant contexts, | a self-paced reading task | ERP results: Experiment 1: a larger N400 response to conventional metaphoric than literal sentences; Experiment 2: a larger N400 response to both novel and conventional metaphoric sentences compared to literal conditions; Experiment 3: a larger N400 response to novel metaphoric sentences embedded in irrelevant contexts than to conventional metaphoric sentences embedded in relevant contexts; |

¹⁸ Behavioral results are presented for those experiments in which behavioral data were reported.

| | | | | | |
|------------------------|------|--|--|--|--|
| | | | conventional metaphoric sentences embedded in irrelevant contexts | | Experiment 4: a larger N400 response to novel metaphoric sentences embedded in relevant contexts than conventional metaphoric sentences embedded in irrelevant contexts. |
| Coulson and Van Petten | 2002 | <i>Conceptual integration and metaphor: An event-related potential study</i> | metaphoric, literal, and literal mapping conditions | a reading task with yes/no comprehension questions | ERP results: A graded N400 response, with the most negative amplitudes for metaphoric utterances, followed by the literal mapping condition, and the least negative N400 amplitudes for literal sentences; A more pronounced LPC response to metaphoric sentences compared to both literal and literal mapping conditions. |
| Tartter et al. | 2002 | <i>Novel metaphors appear anomalous at least momentarily: Evidence from N400</i> | novel metaphoric, literal, and anomalous sentences | a self-paced reading task followed by a recognition task | ERP results: A graded N400 response, with the most pronounced N400 amplitudes for anomalous conditions, followed by novel metaphoric, and finally literal sentences. |
| Arzouan et al. | 2007 | <i>Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension</i> | novel metaphoric, conventional metaphoric, semantically related, and semantically unrelated word dyads | Experiment 1: a semantic decision task; Experiment 2: a self-paced reading task | Behavioral results: Longest RTs for novel metaphors, followed by semantically unrelated utterances, conventional metaphors and semantically related meanings in both experiments. ERP results: A linear effect within the N400 time window, with the smallest N400 amplitudes for semantically related, followed by conventional metaphoric, novel metaphoric, and finally semantically unrelated word pairs in both experiments; |

| | | | | | |
|------------------|------|---|--|--|--|
| | | | | | An attenuated LPC response (sustained negativity) to novel metaphors relative to semantically related and conventional metaphoric word pairs. |
| Lai et al. | 2009 | <i>Comprehending conventional and novel metaphors: An ERP study</i> | novel metaphoric, conventional metaphoric, literal, and anomalous sentences | a multiple-choice semantic decision task | <p>Behavioral results: Longest RTs for novel metaphors, followed by anomalous utterances, conventional metaphors, and shortest RTs for literal sentences.</p> <p>ERP results: A more pronounced N400 response to anomalous, novel metaphoric, and conventional metaphoric sentences compared to literal utterances within the early N400 time frame; A more pronounced N400 response to anomalous and novel metaphoric sentences relative to conventional metaphoric and literal utterances within the late N400 time frame.</p> |
| De Grauwe et al. | 2010 | <i>Electrophysiological insights into the processing of nominal metaphors</i> | <p>Experiment 1: familiar nominal metaphors, literal, and anomalous sentences with their critical words presented at a sentence-final position;</p> <p>Experiment 2: familiar nominal metaphors, literal, and anomalous sentences with their critical words presented at a mid-sentence position</p> | a semantic decision task | <p>ERP results: A more robust N400 response to anomalous conditions and to familiar metaphoric sentences, yet only within the early N400 time frame (in both experiments); A more pronounced LPC response to familiar metaphoric conditions compared to both literal and anomalous sentences in both experiments.</p> |

| | | | | | |
|--------------------|------|---|--|---|---|
| Balconi and Amenta | 2010 | <i>'A fighter is a lion'. Neuropsychological indexes in comprehending frozen metaphors</i> | congruous metaphoric, congruous literal, incongruous metaphoric, and incongruous literal sentences | a self-paced reading task | ERP results: A more pronounced N400 response to incongruous both metaphoric and literal conditions. |
| Goldstein et al. | 2012 | <i>Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization</i> | novel metaphoric, conventional metaphoric, literal, and anomalous word dyads | An exposure phase: a semantic decision task (50% of trials) and an explanation task (50% of trials); An EEG test phase: a semantic decision task | Behavioral results: Longest RTs for novel metaphoric, followed by anomalous, conventional metaphoric, and finally literal word pairs; Shorter RTs for explained compared to unexplained novel metaphors. ERP results: A reduced N400 response to those novel metaphors whose interpretations were provided in the exposure phase; A more pronounced N400 response to those conventional metaphors whose interpretations were provided in the exposure phase. A reduced LPC response (sustained negativity) to novel metaphoric word pairs, which was more pronounced for unexplained compared to explained novel metaphors. |
| Rutter et al. | 2012 | <i>Can clouds dance? Part 2: An ERP investigation of passive conceptual expansion</i> | novel metaphoric, literal, and anomalous sentences | a delayed decision task with two questions: 1). Is the sentence unusual? 2). Is the sentence appropriate? The questions were followed by a subsequent question: for unusual utterances, participants decided whether the sentences were familiar or unfamiliar. For appropriate | Behavioral results: Longest RTs for novel metaphors, followed by anomalous utterances, and finally literal statements. ERP results: A linear effect within the N400 time frame, with the most robust N400 response to anomalous, followed by novel metaphoric, and finally literal statements; |

| | | | | | |
|----------------|------|---|--|---|--|
| | | | | utterances, participants decided if sentences were sensible or nonsensical. | An attenuated LPC response to anomalous and novel metaphoric conditions relative to literal statements. |
| Yang et al. | 2013 | <i>Contextual effects on conceptual blending in metaphors: An event-related potential study</i> | novel metaphors and literal statements preceded by different prime words, so that five conditions were created: congruent target metaphor, congruent source metaphor, incongruent metaphor, congruent literal, and incongruent literal condition | a self-paced reading task | ERP results: No modulations within the N400 response; A more robust LPC response to incongruent metaphoric and literal conditions compared to congruent conditions; A larger LPC response to incongruent metaphoric than incongruent literal conditions; A linear effect across the congruent conditions, with the largest LPC response to congruent target metaphors, followed by congruent literal conditions, and finally congruent source metaphors. |
| Lai and Curran | 2013 | <i>ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors</i> | prime items: sentence-primers, simile-primers; critical items: novel metaphoric, conventional metaphoric, literal, and anomalous sentences | a multiple-choice semantic decision task | Behavioral results: Priming effects observed for conventional metaphors, but not for novel metaphors. ERP results: The N400 priming effect for simile-primers and both novel and conventional metaphors; The N400 priming effect for sentence-primers and conventional metaphors. |

| | | | | |
|---------------------|---|---|--|---|
| Weiland et al. 2014 | <i>The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP</i> | <p>Experiment 1 (without priming): nominal metaphoric and literal sentences;</p> <p>Experiment 2 (with masked priming): prime items: words related to the literal meaning of a presented metaphor; critical items: nominal metaphoric and literal sentences</p> | a color detection task and a word recognition task | <p>Experiment 1: Behavioral results: No differences between literal and nominal metaphoric sentences. ERP results: A biphasic N400-Late Positivity effect, with more pronounced N400 and LPC amplitudes for nominal metaphors relative to control literal sentences.</p> <p>Experiment 2: Behavioral results: Longer RTs for nominal metaphoric relative to literal conditions. ERP results: A biphasic N400-Late Positivity effect, with more pronounced N400 and LPC amplitudes for nominal metaphors relative to control literal sentences; an attenuated N400 response to metaphors preceded by a literal prime word.</p> |
| Bambini et al. 2016 | <i>Disentangling Metaphor from Context: An ERP Study</i> | <p>Experiment 1: familiar and non-familiar metaphoric, as well as literal expressions embedded in minimal contexts;</p> <p>Experiment 2: familiar and non-familiar metaphoric, as well as literal expressions embedded in supportive contexts.</p> | an adjective matching task | <p>Behavioral results: No differences between familiar and non-familiar metaphors.</p> <p>ERP results: More pronounced N400 amplitudes for both familiar and non-familiar metaphoric expressions relative to literal utterances only in the neutral context; A larger LPC response to metaphors compared to literal sentences in both neutral and supportive contexts.</p> |

| | | | | | |
|--------------|-------|---|--|--|--|
| Tang et al. | 2017a | <i>Comprehension of scientific metaphors: Complementary processes revealed by ERP</i> | nominal scientific metaphors, nominal conventional metaphors, and literal statements | a metaphoric judgment task | Behavioral results: Longer RTs for scientific metaphors compared to both conventional metaphors and literal statements. ERP results: A more pronounced N400 response to scientific metaphors than conventional metaphors and literal utterances; Sustained negativity in response to scientific metaphors. |
| Tang et al. | 2017b | <i>Comprehension of scientific metaphors: Complementary processes revealed by ERP</i> | novel scientific metaphors, novel poetic metaphors, conventional metaphors, and literal statements | a metaphoric judgment task | Behavioral results: Longer RTs for both novel scientific and novel poetic metaphors relative to conventional metaphors and literal sentences. ERP results: A more pronounced N400 response to both novel scientific and novel poetic metaphors compared to conventional metaphors and literal utterances; A more robust sustained negativity in response to novel scientific metaphors compared to both novel poetic and conventional metaphors. |
| Rataj et al. | 2018 | <i>On understanding creating language: The late positive complex and novel metaphor comprehension</i> | novel metaphoric, literal, and anomalous sentences | Experiment 1: a semantic decision task; Experiment 2: a reading task with comprehension questions | Behavioral results: Longer RTs for novel metaphors compared to anomalous and literal sentences observed in both experiments. |

ERP results:

A more pronounced N400 response to both anomalous and novel metaphoric sentences compared to literal utterances observed in both experiments;

Attenuated LPC amplitudes in response to novel metaphoric relative to both literal and anomalous sentences over parietal electrode sites, with a broad posterior distribution in Experiment 1, and a left-lateralized distribution in Experiment 2.

Appendix B: Electrophysiological experiments into the N400 and LPC responses in bilingual language processing (based on a review paper by Jankowiak and Rataj 2017: 141-151)

| Authors | Year | Title | Participants | Materials | Task | Electrophysiological and behavioral ¹⁹ results |
|--|------|---|--|---|---------------------------|---|
| THE N400 IN BILINGUAL LEXICO-SEMANTIC PROCESSING | | | | | | |
| Weber-Fox and Neville | 1996 | <i>Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers.</i> | English monolinguals and five groups of Chinese-English bilinguals that differed with respect to their age of L2 acquisition | semantically and syntactically correct as well as incorrect English sentences | a self-paced reading task | ERP results: Larger N400 amplitudes for incorrect compared to correct items in all groups of bilinguals; A delayed N400 peak latency in bilinguals who acquired their L2 after the age of 10. |

¹⁹ Behavioral results are presented for those experiments in which behavioral data were reported.

| | | | | | | |
|------------------|------|---|--|---|---|---|
| Proverbio et al. | 2002 | <i>Electrophysiological measures of language processing in bilinguals.</i> | Italian monolinguals and proficient Italian-Slovenian bilinguals | semantically and syntactically correct and incorrect sentences | a correctness judgment task | ERP results: An attenuated N400 response to semantically and syntactically incorrect stimuli in L2 relative to L1; The N400 response observed over left electrode sites in L2, and over right electrodes in L1. |
| Alvarez et al. | 2003 | <i>Assessing word meaning in two languages: An event-related brain potential study of beginning bilinguals.</i> | beginning and intermediate English-Spanish bilinguals | within-language repetitions, translation equivalents (presented in the backward and forward translation direction) and semantically unrelated words | a mixed-language semantic categorization task | ERP results: A graded N400 effect across conditions, with the most robust N400 amplitudes for semantically unrelated items, followed by translation equivalents presented in the between-language condition, and finally within-language repetitions; No differences in the N400 amplitudes for backward and forward translation directions; A delayed N400 response in the L1-L2 compared to the L2-L1 direction. |

| | | | | | | |
|-------------------------|------|--|--|--|---|---|
| Kotz and Elston-Güttler | 2004 | <i>The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials.</i> | highly proficient and non-proficient German-English bilinguals | English (L2) associatively and categorically related words | an L2 lexical decision task in response to English stimuli presented in an associative and categorical priming paradigm | <p>Behavioral results: Proficient bilinguals: Priming effect for items presented in both categorical and associative priming conditions; Non-proficient bilinguals: Priming effect for items presented in the associative priming condition.</p> <p>ERP results: The N400 associative priming effect observed in both groups of participants; The N400 categorical priming effect observed only in the group of highly proficient bilinguals.</p> |
| Phillips et al. | 2004 | <i>Semantic priming in a first and second language: Evidence from reaction time variability and event-related potentials.</i> | highly proficient and non-proficient English-French bilinguals | semantically related and unrelated items presented in both L1 and L2 | a semantic categorization task | <p>Behavioral results: Longer RTs for L2 in the group of non-proficient bilinguals.</p> <p>ERP results: The N400 semantic priming effect to L2 items observed only in proficient bilinguals; A delay in the N400 peak latency to L2 compared to L1 in both groups of participants.</p> |

| | | | | | | |
|------------------|------|---|---|---|---|--|
| Moreno and Kutas | 2005 | <i>Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish-English bilinguals.</i> | Spanish-English and English-Spanish bilinguals divided into two groups: late non-proficient Spanish-dominant and early proficient English-dominant bilinguals | English and Spanish semantically plausible and implausible sentences | a self-paced reading task | ERP results: A delayed N400 response to the non-dominant language, irrespectively of the condition. |
| Kerkhofs et al. | 2006 | <i>Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects.</i> | late proficient Dutch-English bilinguals | Dutch-English interlingual homographs | an English (L2) lexical decision task | ERP results: More pronounced N400 amplitudes for high-frequency compared to low-frequency items presented in the task-irrelevant language. |
| Thierry and Wu | 2007 | <i>Brain potentials reveal unconscious translation during foreign-language comprehension.</i> | late proficient Chinese-English bilinguals, English monolinguals, and Chinese monolinguals | English words, half of which had a character repetition when translated into Chinese | an English semantic relatedness task | ERP results: Reduced N400 amplitudes for English words with a character repetition when translated into Chinese. |
| Midgley et al. | 2008 | <i>An electrophysiological investigation of cross-language effects of orthographic neighborhood.</i> | late proficient French-English bilinguals | L1 and L2 words with either many or a few cross-language orthographic neighbors in the other language | a go/no-go semantic categorization task | ERP results: Larger N400 amplitudes for words having many orthographic neighbors in the other language compared with lexical items having a few orthographic neighbors in the other tongue. |

| | | | | | | |
|----------------|-------|--|---|--|---|--|
| Midgley et al. | 2009a | <i>Masked repetition and translation priming in second language learners: A window on the time-course of form and meaning activation using ERPs.</i> | late unbalanced intermediate French-English bilinguals | animal names preceded by semantically related or unrelated lexical items presented in the between- and within-language conditions. The between-language conditions were presented in the L1-L2 translation direction (Experiment 1), and L2-L1 translation direction (Experiment 2). | a go/no-go semantic categorization task | ERP results: The N400 semantic priming effect in both L1-L2 and L2-L1 translation directions. |
| Midgley et al. | 2009b | <i>Language effects in second language learners and proficient bilinguals investigated with event-related potentials.</i> | intermediate unbalanced and highly proficient French-English bilinguals | translation equivalents | a go/no-go semantic categorization task for L1 and L2 words | ERP results: A reduced posterior N400 effect to L2 relative to L1 in the group of intermediate bilinguals; A delayed anterior N400 effect to L2 relative to L1 in both intermediate and proficient bilinguals. |

| | | | | | | |
|-------------------|------|---|--|--|--|---|
| Duñabeitia et al. | 2010 | <i>Electrophysiological correlates of the masked translation priming effect with highly proficient simultaneous bilinguals.</i> | balanced proficient Basque-Spanish bilinguals | within-language prime-target repetitions, between-language translation equivalents presented in the L1-L2 and L2-L1 directions, and semantically unrelated words | a lexical decision task | ERP results: A symmetrical N400 semantic priming effect to translation equivalents presented in both L1-L2 and L2-L1 translation directions. |
| Hoshino et al. | 2010 | <i>An ERP investigation of masked cross-script translation priming.</i> | unbalanced intermediate Japanese-English bilinguals | cross-script translation equivalents presented in the L1-L2 and L2-L1 translation directions | a semantic categorization task for L1 and L2 lexical items | ERP results: The N400 semantic priming effect only in response to the L1-L2 translation direction. |
| Geyer et al. | 2011 | <i>Processing words in two languages: An event-related potential study of proficient bilinguals.</i> | proficient Russian-English bilinguals | translation equivalents | a mixed-language lexical decision task | ERP results: A symmetrical N400 semantic priming effect in both L1-L2 and L2-L1 translation directions. |
| Braunstein et al. | 2012 | <i>Investigating the influence of proficiency on semantic processing in bilinguals: An ERPs and ERD/S analysis.</i> | proficient and intermediate German-English bilinguals | English (L2) semantically incongruous items, and words that were high/low in cloze probability | an L2 plausibility judgment task | ERP results: A delayed N400 response in the group of intermediate compared to proficient bilinguals. |
| Newman et al. | 2012 | <i>The influence of language proficiency on lexical semantic processing in native and late learners of English.</i> | English monolinguals and proficient Spanish-English bilinguals | semantically plausible and implausible items | a semantic decision task | ERP results: A reduced and delayed N400 response in the bilingual group, which was more widely distributed in response to |

| | | | | | | |
|---------------------|------|---|---|---|--------------------------------------|--|
| | | | | | | L2 than L1, and more left-lateralized in response to L1 compared to L2. |
| Comesaña et al. | 2012 | <i>The interplay of phonology and orthography in visual cognate word recognition: An ERP study.</i> | proficient Portuguese-English bilinguals | phonological cognates, orthographic cognates, both phonological and orthographic cognates, and non-cognate control words | a reading task | ERP results: A more robust N400 cognate facilitation effect to cognates with high-orthographic and low-phonological overlap compared to cognates having low-orthographic and high-phonological overlap. |
| Hoshino and Thierry | 2012 | <i>Do Spanish-English bilinguals have their finger in two pies – or is it their toes? An electrophysiological investigation of semantic access in bilinguals.</i> | unbalanced Spanish-English bilinguals, and English monolinguals | interlingual homographs related to Spanish or English meanings, and semantically unrelated words | an English semantic relatedness task | ERP results: The N400 semantic priming effect to items that were related to both L1 and L2 words, observed only in the bilingual group. |
| Peeters et al. | 2013 | <i>The representation and processing of identical cognates by late bilinguals: RT and ERP effects.</i> | late advanced French-English bilinguals | identical cognates that were either of high frequency in both L1 and L2, of high frequency in L2 and low frequency in L1, of low frequency to both L1 and L2, or of low frequency to L2 and high frequency in L1. | an L2 lexical decision task | Behavioral results: Shorter RTs for cognates compared to non-cognates. ERP results: More pronounced N400 amplitudes for low-frequency English and high-frequency French cognates compared to high-frequency English and low-frequency French cognate words. |

| | | | | | | |
|---------------------|------|--|--|--|--|--|
| Chen et al. | 2013 | <i>An ERP study on metaphor comprehension in the bilingual brain</i> | proficient Chinese-English bilinguals | Chinese and English metaphoric and literal sentences | a yes/no judgment task, where participants decided whether the sentence had a metaphoric meaning | ERP results: No differences within the N400 time window to metaphorical and literal sentences presented in Chinese (L1); a more pronounced N400 response to English (L2) metaphors as compared to English literal meanings. |
| Jouravlev and Jared | 2014 | <i>Reading Russian–English homographs in sentence contexts: Evidence from ERPs.</i> | English monolinguals, balanced proficient Russian-English bilinguals | sentences with interlingual homographs, English translations of the Russian meaning of a homograph (English Match), and unrelated control words (English Mismatch) | a reading task | ERP results: A graded N400 effect, with the most pronounced N400 amplitudes for English Mismatch, followed by interlingual homographs, and the least negative N400 amplitudes for the English Match condition. |
| Casaponsa et al. | 2015 | <i>How do bilinguals identify the language of the words they read?</i> | early proficient Spanish-Basque bilinguals | L1 and L2 words manipulated in terms of their orthotactic plausibility in L1 | a go/no-go semantic categorization task | ERP results: Larger N400 amplitudes for bigrams that were implausible in L1 compared to those that were plausible in L1. |
| Paulmann et al. | 2015 | <i>Neurophysiological markers of phrasal verb processing: Evidence from L1 and L2 speakers</i> | English native speakers and late Arabic-English bilinguals | English verb-preposition strings, presented in two different contexts biasing towards either a literal or a figurative interpretation. | a reading task with comprehension questions | ERP results: A more robust N400 response to literal compared to figurative interpretations. |

| | | | | | | |
|--|------|--|--|---|--------------------------------|--|
| Moldovan et al. | 2016 | <i>The role of meaning and form similarity in translation recognition in highly proficient balanced bilinguals: A behavioral and ERP study</i> | highly proficient Catalan/Spanish bilinguals | Catalan-Spanish pairs of words that were either highly or moderately related in meaning, as well as pairs of words where the second item had a similar lexical form to the correct translation of the first item (i.e., a translation neighbor) | a translation recognition task | ERP results: A more robust N400 response to translation neighbors compared to words that were highly and moderately related in meaning. |
| THE LPC RESPONSE IN BILINGUAL LEXICO-SEMANTIC PROCESSING | | | | | | |
| van der Meij et al. | 2011 | <i>Electrophysiological correlates of language switching in second language learners</i> | high- and low-proficient Spanish-English bilinguals | English (L2) sentences, half of which contained a Spanish (L1) adjective in the middle of a sentence | a self-paced reading task | ERP results: A robust LPC effect to switching conditions, with a broad fronto-posterior distribution. The effect was, however, maximal over the frontal sites for the low-proficiency group, and over the posterior sites for the high-proficiency group. |
| Hoshino and Thierry | 2012 | <i>Do Spanish-English bilinguals have their finger sin two pies – or is it their toes? An electrophysiological investigation of semantic access in bilinguals.</i> | unbalanced Spanish-English bilinguals and English monolinguals | interlingual homographs | a semantic relatedness task | ERP results: A more pronounced LPC effect to semantically related conditions, which was maximal over the centro-parietal region, and did not vary as a function of language group. |

| | | | | | | |
|------------|------|--|---------------------------------------|---------------------------|--------------------------------|---|
| Guo et al. | 2012 | <i>On the time course of accessing meaning in a second language: An electrophysiological and behavioral investigation of translation recognition</i> | proficient Chinese-English bilinguals | English and Chinese words | a translation recognition task | ERP results: A larger anterior LPC effect to semantically related translation distractors relative to unrelated controls; A more pronounced posterior LPC effect to unrelated conditions. |
|------------|------|--|---------------------------------------|---------------------------|--------------------------------|---|

Appendix C: Experiment 1: Polish experimental stimuli

| No. | Novel metaphors | Conventional metaphors | Literal utterances | Anomalous utterances | Critical words |
|-----|-----------------|------------------------|--------------------|----------------------|----------------|
| 1. | turlać | znosić | wspominać | przeplłynąć | biedę |
| 2. | przyciągać | budować | osiągnąć | posmarować | bogactwo |
| 3. | podeptać | odzyskać | wyrazić | odśnieżyć | chwałę |
| 4. | nakarmić | rozpalić | krytykować | ugnieść | ciekawość |
| 5. | rozerwać | przebić | rozjaśnić | wysuszyć | ciemność |
| 6. | odbierać | nadwyreżać | wykazać | odwiedzić | cierpliwość |
| 7. | ostudzić | podwyższać | badać | aresztować | ciśnienie |
| 8. | osierocić | przekreślić | ubezpieczyć | rozładować | dorobek |
| 9. | ociekać | napelniać | reagować | pomieszać | dumą |
| 10. | zbierać | ciąć | zatwierdzić | wszczepić | etaty |
| 11. | oświecać | zwalczać | zrozumieć | suszyć | głupotę |
| 12. | ukraść | sprzedać | szanować | podlewać | godność |
| 13. | wyzwalać | zbadać | ogłosić | kulać | gotowość |
| 14. | odwracać | oddalić | pojąć | czyścić | groźbę |
| 15. | pielęgnować | rozsiwać | wybaczyć | wynajmować | kłamstwo |
| 16. | splącić | wyrównać | przebaczyć | popłamić | krzywdę |
| 17. | popchnąć | odłożyć | zaplanować | okłamywać | kupno |
| 18. | rozmnożyć | odnaleźć | podziwiać | lakierować | mądrość |
| 19. | reżyserować | przedłużać | kochać | napiąć | młodość |
| 20. | przeskoczyć | uchylić | spisać | zwiedzać | nakaz |
| 21. | przepędzić | śledzić | notować | przebierać | objawy |
| 22. | przełamać | kształtować | zmienić | wyspowiadać | obyczaj |
| 23. | zawiesić | wyłączyć | zapewnić | pomalować | odbiór |
| 24. | gubić | wywołać | nazwać | zjeść | odczucie |
| 25. | przelknąć | złożyć | wyjaśnić | ubierać | odmowę |
| 26. | wyprosić | zebrać | przyrzekać | połaskotać | odwagę |
| 27. | uprawiać | wyłapać | rozpoznać | skleić | oszustwo |
| 28. | podgrzewać | ucinać | ignorować | skaleczyć | plotki |
| 29. | odsuwać | przerwać | rozpocząć | zamieść | pogoń |
| 30. | odkręcić | naprawiać | usprawiedliwić | dźgać | pomyłkę |
| 31. | dusić | nakręcić | sprawdzić | opublikować | popyt |

| | | | | | |
|-----|-------------------|---------------|--------------|------------|---------------|
| 32. | chłonać | zamówić | szeptać | wylizać | porady |
| 33. | klepnać | zagrzać | doradzić | nalać | posadę |
| 34. | zdusić | zdradzać | spełnić | upiec | pragnienie |
| 35. | wylewać | wnosić | wy tłumaczyć | oliwić | pretensje* |
| 36. | wietrzyć | przyjąć | potwierdzić | pokroić | przegraną |
| 37. | wyplenić | ścigać | analizować | kupić | przestępczość |
| 38. | odświeżyć | znieść | oferować | zaparzyć | przywilej |
| 39. | przesłać | sprowadzić | obietcać | szczytać | ratunek |
| 40. | wyrwać | uratować | zanotować | piłować | remis |
| 41. | odpychać | siać | pamiętać | remontować | rozpacz |
| 42. | odkurzyć | znaleźć | organizować | ukłuć | rozrywkę |
| 43. | usypiać | odebrać | docenić | odrobić | rozsądek |
| 44. | strzelić | przeprowadzić | planować | zapraszać | rozwód |
| 45. | odpędzać | pożegnać | wybrać | rozbroić | samotność |
| 46. | wydobyć | odziedziczyć | wykryć | wypić | skłonność |
| 47. | włączyć | ćwiczyć | ułatwić | powiesić | skupienie |
| 48. | zburzyć | powiększyć | zweryfikować | zalać | skuteczność |
| 49. | odsłonić | pokazać | przewidywać | nagrać | słabość |
| 50. | roztrwonić | zdobyć | uwielbiać | nawilżyć | sławę |
| 51. | zamrozić | zagłuszyć | udawać | przybijać | smutek |
| 52. | łapać | obniżyć | polepszyć | pofarbować | sprawność |
| 53. | zacierać | pogodzić | rozważyć | doprawić | sprzecznosci |
| 54. | pokonać | osłodzić | lubić | prać | starość |
| 55. | zasznurować | krępować | wykorzystać | gotować | swobodę |
| 56. | karmić | wyleczyć | akceptować | rozpakować | szaleństwo |
| 57. | złowić | wychwycić | zapamiętać | podrywać | szepty |
| 58. | wyciągnąć | rozwinąć | obliczyć | pustoszyć | szybkość |
| 59. | zdeptać | ukryć | ustalić | dekorować | tożsamość |
| 60. | przechowy- wać | zmniejszyć | stwierdzić | pobić | uraz |
| 61. | zrujnować | zdjąć | przeceniać | czesać | urok |
| 62. | wyhodować | wprowadzić | wymyślić | pościelić | ustrój |
| 63. | rodzić | oddać | czuć | masować | wdzięczność |
| 64. | stępić | wykształcić | okazać | zwijać | wrażliwość |
| 65. | zabić | podnieść | testować | przyciszyć | wydajność |
| 66. | przemycić | odnowić | przypomnieć | zagrabić | wyznanie |
| 67. | wyprzedzić | dośćignąć | uznać | podwieźć | wzorzec |
| 68. | uciszyć | budzić | powodować | holować | wzruszenie |
| 69. | zatrucić | dać | usłyszeć | objąć | zachętę |
| 70. | odgrzać | pogłębiać | poczuć | umyć | zachwyty |
| 71. | uwolnić | ujrzyć | omówić | rozminować | zalety |
| 72. | obciąć | poszerzyć | wyznaczyć | rozmieszać | zasięg |
| 73. | utopić | przejąć | uwzględnić | kołysać | zasługi |
| 74. | odganiać | leczyć | uzasadnić | zginać | zazdrość |
| 75. | przyrządzić | wymierzyć | podejrzewać | skanować | zemstę |

* A Polish-English cognate

76. podbić oszukać posiadać przesadzić zmysły

Appendix D: Experiment 1: English experimental stimuli

| No. | Novel metaphors | Conventional metaphors | Literal utterances | Anomalous utterances | Critical words |
|-----|-----------------|------------------------|--------------------|----------------------|-----------------|
| 1. | to sweep | to weaken | to allow | to melt | an adjustment |
| 2. | to eliminate | to acquire | to want | to shave | admission |
| 3. | to compose | to attract | to continue | to put | adventure |
| 4. | to reflect | to treat | to feel | to bump | anger |
| 5. | to slice | to beat | to tolerate | to eat | anxiety |
| 6. | to draw | to withhold | to require | to tickle | an approval |
| 7. | to upset | to design | to complete | to poke | an arrangement* |
| 8. | to launch | to block | to file | to pick | assault |
| 9. | to hurt | to damage | to note | to torture | attendance |
| 10. | to nurse | to pass | to lack | to fold | awareness |
| 11. | to seal | to strike | to find | to flash | a bargain |
| 12. | to dictate | to take | to advise | to score | caution |
| 13. | to hunt | to practice | to start | to shoot | charity |
| 14. | to break | to halt | to prevent | to suck | a collapse |
| 15. | to shake | to develop | to reject | to plug | conception* |
| 16. | to clean | to heal | to regulate | to duck | conduct |
| 17. | to enter | to recover | to study | to park | consciousness |
| 18. | to capture | to construct | to sign | to retire | consent |
| 19. | to carry | to follow | to await | to tie | conviction |
| 20. | to harvest | to gather | to experience | to move | courage |
| 21. | to store | to trade | to discuss | to irritate | courtesies |
| 22. | to offend | to adopt | to learn | to water | a custom |
| 23. | to hide | to reverse | to report | to wear | decline |
| 24. | to swallow | to challenge | to admit | to frame | defeat |
| 25. | to freeze | to catch | to delay | to dig | departure |
| 26. | to uncover | to boost | to get | to cross | dignity |
| 27. | to heat | to build | to mediate | to hug | dispute |
| 28. | to approach | to achieve | to consider | to clip | divorce |
| 29. | to expand | to shift | to indicate | to knit | emphasis* |

* A Polish-English cognate

| | | | | | |
|-----|--------------|-----------------|----------------|--------------|---------------|
| 30. | to fill | to deliver | to limit | to bite | employment |
| 31. | to loosen | to strengthen | to prepare | to sell | an engagement |
| 32. | to lower | to reveal | to notice | to slap | an exception |
| 33. | to smell | to invent | to believe | to hit | excuses |
| 34. | to stretch | to close | to propose | to pull | an extension |
| 35. | to grasp | to lose | to offer | to consume | a favor |
| 36. | to clear | to judge | to calculate | to press | frequency* |
| 37. | to divide | to steal | to deserve | to climb | glory |
| 38. | to generate | to supply | to provide | to crown | guidance |
| 39. | to collect | to raise | to wish | to electrify | happiness |
| 40. | to open | to leave | to observe | to dust | heaven |
| 41. | to soften | to overpower | to survive | to paint | an illness |
| 42. | to formulate | to enhance | to share | to count | an insight |
| 43. | to digest | to crash | to finish | to burn | a journey |
| 44. | to choke | to crush | to film | to peel | laughter |
| 45. | to repair | to defend | to manage | to steam | legacy |
| 46. | to bury | to exercise | to guarantee | to reprint | liberty |
| 47. | to fight | to exit | to devote | to drown | a lifetime |
| 48. | to renovate | to refresh | to purchase | to snort | a membership |
| 49. | to spread | to bring | to need | to dress | a miracle |
| 50. | to sharpen | to broaden | to change | to push | an outlook |
| 51. | to tax | to characterize | to demonstrate | to cook | patience** |
| 52. | to slip | to introduce | to increase | to swim | a penalty |
| 53. | to win | to secure | to prohibit | to roast | permission |
| 54. | to pierce | to cut | to cause | to brush | pollution |
| 55. | to battle | to combat | to minimize | to hike | poverty |
| 56. | to feed | to shape | to describe | to fly | prediction* |
| 57. | to taste | to extend | to abuse | to drag | privilege* |
| 58. | to absorb | to skip | to risk | to kiss | a punishment |
| 59. | to deceive | to execute | to escape | to tear | pursuit |
| 60. | to inspect | to overcome | to write | to bruise | regret |
| 61. | to craft | to throw | to add | to marry | remarks |
| 62. | to trace | to perform | to attempt | to dye | a rescue |
| 63. | to breed | to silence | to deny | to cry | rumors |
| 64. | to erase | to command | to make | to read | a sacrifice |
| 65. | to fix | to direct | to signal | to bake | a shortage |
| 66. | to chase | to miss | to lengthen | to stick | survival* |
| 67. | to attack | to investigate | to confirm | to annoy | a suspicion |
| 68. | to harden | to destroy | to understand | to scratch | unity |
| 69. | to cultivate | to promote | to explain | to ride | virtue |
| 70. | to produce | to highlight | to identify | to rescue | weakness |
| 71. | to terminate | to display | to announce | to lick | willingness |
| 72. | to hold | to accumulate | to define | to walk | wisdom |
| 73. | to kill | to ruin | to ignore | to smoke | wishes |

* A Polish-English cognate

** A Polish-English interlingual homograph

| | | | | | |
|-----|-----------|------------|-----------|-----------|--------------|
| 74. | to drop | to settle | to permit | to clutch | a withdrawal |
| 75. | to murder | to provoke | to enjoy | to glue | wonder |
| 76. | to kick | to balance | to show | to curl | worry |

Appendix E: Experiment 2: Polish experimental stimuli

| No. | Novel nominal metaphors | Novel similes | Literal sentences | Anomalous sentences | Critical words |
|-----|-------------------------|------------------------|-------------------|---------------------|----------------|
| 1. | Rozsądek to | Rozsądek jest jak | Ten zawód to | Kierownica to | adwokat |
| 2. | Natura to | Natura jest jak | Ten sklep to | Czapka to | apteka |
| 3. | Dziecko to | Dziecko jest jak | Ta postać to | Grzejnik to | badacz |
| 4. | Studia to | Studia są jak | To wydarzenie to | Lodówka to | bitwa |
| 5. | Rozterki to | Rozterki są jak | To zjawisko to | Wazon to | burza |
| 6. | Jej figura to | Jej figura jest jak | Ten mur to | Szklanka to | cegła |
| 7. | Kłopoty to | Kłopoty są jak | Ten widok to | Wstążka to | chmura |
| 8. | Jego mięśnie to | Jego mięśnie są jak | Ten deser to | Bluzka to | ciasto |
| 9. | Zakochanie to | Zakochanie jest jak | Ten towar to | Armia to | cukier |
| 10. | Sumienie to | Sumienie jest jak | Ten instrument to | Chleb to | dzwonek |
| 11. | Krew to | Krew jest jak | Ta ciecz to | Pierścień to | farba |
| 12. | Opinia to | Opinia jest jak | Ten symbol to | Taczka to | flaga* |
| 13. | Maniery to | Maniery są jak | To ubranie to | Pudło to | garnitur |
| 14. | Kołyska to | Kołyska jest jak | Ta konstrukcja to | Zakładka to | gniazdo |
| 15. | Plotka to | Plotka jest jak | Ten blask to | Gałka to | iskra |
| 16. | Ten grubas to | Ten grubas jest jak | Ten posiłek to | Puder to | jajko |
| 17. | Makijażysta to | Makijażysta jest jak | Ten mężczyzna to | Kolumna to | kapłan |
| 18. | Jej garderoba to | Jej garderoba jest jak | Ten budynek to | Kominek to | kaplica |
| 19. | Krytyka to | Krytyka jest jak | Ta broń to | Kabel to | karabin |
| 20. | Pożyczka to | Pożyczka jest jak | Ten samochód to | Ramka to | karetka |
| 21. | Więzienie to | Więzienie jest jak | Ten zabytek to | Latarnia to | klasztor |
| 22. | Koszula to | Koszula jest jak | Ten papier to | Obrus to | koperta |
| 23. | Nieśmiałość to | Nieśmiałość jest jak | Ten zator to | Klamka to | korek |
| 24. | Uroda to | Uroda jest jak | Ta ozdoba to | Linijka to | korona |
| 25. | Rodzina to | Rodzina jest jak | Ten badył to | Podeszwa to | korzeń |
| 26. | Skóra to | Skóra jest jak | To okrycie to | Wiatrak to | koszula |
| 27. | Osobowość to | Osobowość jest jak | Ten malunek to | Sałátka to | krajobraz |
| 28. | Uczelnia to | Uczelnia jest jak | Szwecja to | Cyrkiel to | królestwo |
| 29. | Starość to | Starość jest jak | Ta książka to | Spinka to | kronika* |

* A Polish-English cognate

| | | | | | |
|-----|------------------------|------------------------------|---------------------|--------------|------------|
| 30. | Człowiek to | Człowiek jest jak | Ta łąza to | Brama to | kropla |
| 31. | Obrączka to | Obrączka jest jak | Ten sprzęt to | Kredka to | łańcuch |
| 32. | Zapomoga to | Zapomoga jest jak | Ta podpora to | Kapsel to | laska |
| 33. | Emerytura to | Emerytura jest jak | Ta deska to | Widły to | ławka |
| 34. | Potomek to | Potomek jest jak | Ta powierzchnia to | Gałąź to | lustro |
| 35. | Intelekt to | Intelekt jest jak | To wybrzuszenie to | Strzelba to | mięsień |
| 36. | Jej młodsza siostra to | Jej młodsza siostra jest jak | To stworzenie to | Łopata to | mucha |
| 37. | Prawo to | Prawo jest jak | Ten strój to | Piekarnik to | mundur |
| 38. | Uczucia to | Uczucia są jak | Ten proces to | Rękawice to | oddech |
| 39. | Akcent to | Akcent jest jak | Ten drobiazg to | Woda to | pamiątka |
| 40. | Blizny to | Blizny są jak | Te zapiski to | Młot to | pamiętnik |
| 41. | Życie to | Życie jest jak | Ta nawierzchnia to | Cukierek to | parkiet |
| 42. | Ta dziewczyna to | Ta dziewczyna jest jak | Ten przedmiot to | Kurtka to | pióro |
| 43. | Twarz to | Twarz jest jak | Ten arkusz to | Kołyska to | plakat |
| 44. | Przeszłość to | Przeszłość jest jak | Ten ślad to | Hamulec to | plama |
| 45. | Zatarg to | Zatarg jest jak | To światło to | Klucz to | plomień |
| 46. | Ten docinek to | Ten docinek jest jak | Ta kulka to | Okno to | pocisk |
| 47. | Rozwód to | Rozwód jest jak | Ten rytuał to | Gazeta to | pogrzeb |
| 48. | Rozpacz to | Rozpacz jest jak | Ten kataklizm to | Beczka to | powódź |
| 49. | Umysł to | Umysł jest jak | Ten pokój to | Tama to | pracownia |
| 50. | Błąd to | Błąd jest jak | Ta przestrzeń to | Walizka to | przepaść |
| 51. | Wiedza to | Wiedza jest jak | Ta osoba to | Stoik to | przewodnik |
| 52. | Dorosłość to | Dorosłość jest jak | Ten dół to | Dywan to | pułapka |
| 53. | Kac to | Kac jest jak | Ten obszar to | Pomadka to | pustynia |
| 54. | Ta fryzura to | Ta fryzura jest jak | Ta dekoracja to | Szynka to | rzeźba |
| 55. | Kościół to | Kościół jest jak | Ten dom to | Grabie to | schronisko |
| 56. | Tradycja to | Tradycja jest jak | Ta przeszkoda to | Plaszcz to | skała |
| 57. | Ten naród to | Ten naród jest jak | Te łabędzie to | Zapałka to | stado |
| 58. | Rozum to | Rozum jest jak | Ta budowla to | Klatka to | świątynia |
| 59. | Istnienie to | Istnienie jest jak | To oświetlenie to | Rower to | świeca |
| 60. | Mózg to | Mózg jest jak | Ta wnęka to | Manekin to | szafa |
| 61. | Uśmiech to | Uśmiech jest jak | Ten znak to | Łyżka to | sztandar |
| 62. | Makijaż to | Makijaż jest jak | To uzbrojenie to | Krzesło to | tarcza |
| 63. | Autostrada to | Autostrada jest jak | To wykończenie to | Siodło to | taśma |
| 64. | Pamięć to | Pamięć jest jak | To opakowanie to | Ekran to | torba |
| 65. | Życiorys to | Życiorys jest jak | Ten druczek to | Lakier to | ulotka |
| 66. | Rodzeństwo to | Rodzeństwo jest jak | Kuszetka to | Drzwi to | wagon* |
| 67. | Talent to | Talent jest jak | To pomieszczenie to | Kałuża to | winda |
| 68. | Instynkt to | Instynkt jest jak | Ten człowiek to | Włócznia to | władca |
| 69. | Ciało to | Ciało jest jak | Ten pojemnik to | Szlaban to | worek |
| 70. | Jej świat to | Jej świat jest jak | Ten pojazd to | Chodnik to | wózek |
| 71. | Słowo to | Słowo jest jak | Ten prezent to | Kalafior to | zabawka |

* A Polish-English cognate

| | | | | | |
|-----|----------------|----------------------|-----------------|-------------|----------|
| 72. | Szczęście to | Szczęście jest jak | To zwierzę to | Ręcznik to | zając |
| 73. | Serce to | Serce jest jak | Ten przyrząd to | Nożyczki to | zegar |
| 74. | Jego nos to | Jego nos jest jak | To warzywo to | Balsam to | ziemniak |
| 75. | Piec to | Piec jest jak | Ten organ to | Robak to | żołądek |
| 76. | Ta widownia to | Ta widownia jest jak | Te szczątki to | Dzida to | zwłoki |

Appendix F: Experiment 2: English experimental stimuli

| No. | Novel nominal metaphors | Novel similes | Literal sentences | Anomalous sentences | Critical words |
|-----|-------------------------|-----------------------------|------------------------|----------------------|----------------|
| 1. | My head is | My head is like | The brown thing is | A flying pen is | a basket |
| 2. | Your body is | Your body is like | This warming thing is | A pen in the eyes is | a blanket |
| 3. | Her soul is | Her soul is like | This grey thing is | A big drama is | a boulder |
| 4. | Happiness is | Happiness is like | A flying circle is | These shoes are | a bubble |
| 5. | His decision is | His decision is like | The object he holds is | Her nice tooth is | a bullet |
| 6. | Hope is | Hope is like | That bright dot is | These books are | a candle |
| 7. | My lover is | My lover is like | A sweet thing is | A spare camel is | a candy |
| 8. | My room is | My room is like | This big building is | A small dot is | a castle |
| 9. | A great achievement is | A great achievement is like | The wall above you is | A running diary is | a ceiling |
| 10. | The Internet is | The Internet is like | This corridor is | A blue velvet is | a chamber |
| 11. | Good attitude is | Good attitude is like | This runner is | A warm ice is | a champion |
| 12. | Addiction is | Addiction is like | This shape is | A green cream is | a circuit |
| 13. | Fitness is | Fitness is like | That man is | My red hair is | a companion* |
| 14. | The human mind is | The human mind is like | This box is | These flowers are | a container |
| 15. | Time is | Time is like | That woman is | A period of time is | a convict** |
| 16. | A positive thought is | A positive thought is like | A tasty thing is | A warm blanket is | a cookie |
| 17. | Language is | Language is like | This cash is | Ugly faces are | a currency |
| 18. | A lie is | A lie is like | A long material is | A long dis-course is | a curtain |
| 19. | God is | God is like | A salesman is | The door is | a dealer* |
| 20. | Alcohol is | Alcohol is like | This being is | This shower is | a devil |
| 21. | My heart is | My heart is like | This divider is | A large bug is | a drawer |

* A Polish-English cognate

** A Polish-English interlingual homograph

| | | | | | |
|-----|-------------------------|------------------------------|----------------------|------------------------|-------------|
| 22. | Memory is | Memory is like | This piece of art is | Sweet lemons are | a drawing |
| 23. | A good book's ending is | A good book's ending is like | This celebration is | A blue monkey is | a funeral |
| 24. | Prescription is | Prescription is like | Her word is | A book in the chair is | a guideline |
| 25. | A family house is | A family house is like | This area is | A nice owl is | a harbor |
| 26. | My baby's babble is | My baby's babble is like | This speech is | A brown cow is | a lecture |
| 27. | A favorite book is | A favorite book is like | My husband is | Smelly words are | a lover |
| 28. | The devil is | The devil is like | That seller is | A homeless cup is | a merchant |
| 29. | Apathy is | Apathy is like | A violet being is | A little flower is | a monster |
| 30. | An anorectic is | An anorectic is like | A tiny device is | A pink mother is | a needle |
| 31. | A family is | A family is like | This unit is | A crying device is | a parish |
| 32. | Her shoulder is | Her shoulder is like | This soft thing is | A very small plate is | a pillow |
| 33. | Hatred is | Hatred is like | This liquid is | Bees with a tire are | a poison |
| 34. | My cat is | My cat is like | That woman is | A huge cup is | a princess |
| 35. | Pleasure is | Pleasure is like | This animal is | A small edge is | a rabbit |
| 36. | Laziness is | Laziness is like | That soldier is | Big lips are | a rebel |
| 37. | A mouse in my attic is | A mouse in my attic is like | This rebel is | A yawning skirt is | a refugee |
| 38. | A happy relationship is | A happy relationship is like | This medal is | Drunk cats are | a reward |
| 39. | A redhead is | A redhead is like | That bird is | A pale negro is | a robin |
| 40. | Amnesia is | Amnesia is like | This small object is | A small bug is | a rubber |
| 41. | Youth is | Youth is like | This skinny man is | Her big fridge is | a runner |
| 42. | A medicine is | A medicine is like | This maid is | This blanket is | a servant |
| 43. | Friendship is | Friendship is like | This cave is | A pink dragon is | a shelter |
| 44. | An ugly truth is | An ugly truth is like | A wet place is | Draught in the box is | a shower |
| 45. | My purring cat is | My purring cat is like | That man is | That chair is | a singer |
| 46. | A man's reason is | A man's reason is like | That tall man is | Blue chickens are | a spokesman |
| 47. | Fear is | Fear is like | That individual is | Blue wagons are | a stranger |
| 48. | A flashback is | A flashback is like | This article is | A smelly cat is | a summary |
| 49. | Achievement is | Achievement is like | This peak is | A short bar is | a summit |
| 50. | A sapper is | A sapper is like | The man in white is | My blue velvet is | a surgeon |
| 51. | Hesitation is | Hesitation is like | That tall boy is | A round square is | a teenager |
| 52. | A candy shop is | A candy shop is like | A holy place is | That blue cat is | a temple |
| 53. | Freedom is | Freedom is like | This organism is | My new dress is | a tiger |
| 54. | Integrity is | Integrity is like | This golden box is | Blue wind is | a treasure |

| | | | | | |
|-----|----------------------------|---------------------------------|-------------------------|------------------------|---------------|
| 55. | Marriage is | Marriage is like | This agreement is | A cub in the air is | a treaty |
| 56. | A doctor's office is | A doctor's office is like | This house is | A handsome man is | a workshop |
| 57. | An unpleasant situation is | An unpleasant situation is like | This substance is | A dirty eye is | an acid |
| 58. | An alcoholic beverage is | An alcoholic beverage is like | A good friend is | A man with a claw is | an adviser |
| 59. | Faith is | Faith is like | This steel is | A man in a hat is | an anchor |
| 60. | Calmness is | Calmness is like | This figure is | A round dog is | an angel |
| 61. | A happy life is | A happy life is like | This street is | A soft movie is | an avenue |
| 62. | My mother is | My mother is like | This huge animal is | A pink dress is | an elephant |
| 63. | This love is | This love is like | This small boy is | A huge razor is | an infant |
| 64. | Hate is | Hate is like | This virus is | A sour girl is | an infection* |
| 65. | Peace is | Peace is like | That researcher is | These jeans are | an observer* |
| 66. | A desert is | A desert is like | The hot drawer is | A tasty space is | an oven |
| 67. | Details in a plan are | Details in a plan are like | Small circles are | A meat in the glass is | buttons |
| 68. | A compliment is | A compliment is like | This material is | A fresh mobile is | cotton |
| 69. | Sad words are | Sad words are like | These sticks are | Funny handles are | crutches |
| 70. | Angels are | Angels are like | These black people are | Books in the wind are | fighters |
| 71. | Your words are | Your words are like | A yellow thing is | A curtain in a bar is | honey |
| 72. | Kind words are | Kind words are like | These shiny things are | A storm in north is | jewelry |
| 73. | My thoughts are | My thoughts are like | The thing in a glass is | A hairy monkey is | liquid |
| 74. | Alcoholics are | Alcoholics are like | These villains are | Burning plans are | prisoners |
| 75. | Viruses are | Viruses are like | These people are | White plates are | travelers |
| 76. | Happy thoughts are | Happy thoughts are like | These soldiers are | Bad frogs are | warriors |

* A Polish-English cognate

Appendix G: Experiment 1: Instructions in the normative studies on Polish experimental word dyads

1. Abstractness ratings:

Na skali od 1-7 zaznacz jak bardzo, według Ciebie, abstrakcyjne lub konkretne są poniższe rzeczowniki:

- 1- zdecydowanie abstrakcyjny
- 2- abstrakcyjny
- 3- raczej abstrakcyjny
- 4- ani konkretny ani abstrakcyjny
- 5- raczej konkretny
- 6- konkretny
- 7- zdecydowanie konkretny

Na przykład,

miłość – 1 (zdecydowanie abstrakcyjny)

książka – 7 (zdecydowanie konkretny)

2. Meaningfulness ratings:

Na skali od 1-7 zaznacz jak bardzo, według Ciebie, sensowne lub bezsensowne są poniższe wyrażenia:

- 1- zdecydowanie bezsensowne
- 2- bezsensowne
- 3- raczej bezsensowne

- 4- ani sensowne ani bezsensowne
- 5- raczej sensowne
- 6- sensowne
- 7- zdecydowanie sensowne

Na przykład,

kopać wyrazy – 1 (zdecydowanie bezsensowne)

skreślać wyrazy – 7 (zdecydowanie sensowne)

3. Familiarity ratings:

Na skali od 1-7 zaznacz jak często spotykasz się z poniższymi wyrażeniami:

- 1- bardzo rzadko
- 2- rzadko
- 3- raczej rzadko
- 4- czasem
- 5- raczej często
- 6- często
- 7- bardzo często

Na przykład,

rozbić serce – 1 (bardzo rzadko)

trzymać kciuki – 7 (bardzo często)

4. Metaphoricity ratings:

Na skali od 1-7 zaznacz jak bardzo, według Ciebie, dosłowne lub metaforyczne są poniższe wyrażenia:

- 1- bardzo dosłowne
- 2- dosłowne
- 3- raczej dosłowne
- 4- ani dosłowne ani metaforyczne
- 5- raczej metaforyczne

6- metaforyczne

7- bardzo metaforyczne

Na przykład,

przygotować projekt – 1 (bardzo dosłowne)

zapiąć projekt – 7 (bardzo metaforyczne)

5. Cloze probability tests:

Do podanego czasownika dopisz rzeczownik, który pierwszy przyjdzie Ci na myśl, tak aby powstałe wyrażenie miało sens.

Na przykład,

podlewać kwiaty

zapalić światło

Appendix H: Experiment 1: Instructions in the normative studies on English experimental word dyads

1. Abstractness ratings:

On the scale from 1 to 7, please rate how abstract or concrete you think a particular noun is.

- 1- Definitely abstract
- 2- Mostly abstract
- 3- Rather abstract
- 4- Neither abstract nor concrete
- 5- Rather concrete
- 6- Mostly concrete
- 7- Definitely concrete

For instance,

love – 1 (highly abstract)

computer – 7 (highly concrete)

2. Meaningfulness ratings:

On the scale from 1-7, please rate how meaningful or meaningless, in your opinion, a particular utterance is.

- 1- Totally meaningless
- 2- Meaningless

- 3- Rather meaningless
- 4- Neither meaningless nor meaningful
- 5- Rather meaningful
- 6- Meaningful
- 7- Totally meaningful

For instance,

to drive a potato – 1 (totally meaningless)

to buy food – 7 (totally meaningful)

3. Familiarity ratings:

On the scale from 1-7, decide how often you encounter a particular utterance.

- 1- Very rarely
- 2- Rarely
- 3- Rather rarely
- 4- Sometimes
- 5- Rather frequently
- 6- Frequently
- 7- Very frequently

For instance,

to explore reflections – 1 (very rarely)

to buy food – 7 (very frequently)

4. Metaphoricity ratings:

On the scale from 1-7, please rate how metaphorical or literal, according to you, a particular utterance is.

- 1- Very literal
- 2- Literal
- 3- Slightly literal
- 4- Neither literal nor metaphorical

5- Slightly metaphorical

6- Metaphorical

7- Very metaphorical

For instance,

to buy food – 1 (very literal)

to inject ambitions – 7 (very metaphorical)

5. Cloze probability tests:

To a presented verb add a noun that first comes to your mind, so that the expression makes sense.

For instance,

to water flowers

to wash the dishes

Appendix I: Experiment 2: Instructions in the normative studies on Polish experimental sentences

1. Abstractness ratings:

Na skali od 1-7 zaznacz jak bardzo, według Ciebie, abstrakcyjne lub konkretne są poniższe rzeczowniki:

- 1- zdecydowanie abstrakcyjny
- 2- abstrakcyjny
- 3- raczej abstrakcyjny
- 4- ani konkretny ani abstrakcyjny
- 5- raczej konkretny
- 6- konkretny
- 7- zdecydowanie konkretny

Na przykład,

miłość – 1 (zdecydowanie abstrakcyjny)

książka – 7 (zdecydowanie konkretny)

2. Meaningfulness ratings:

Na skali od 1-7 zaznacz jak bardzo, według Ciebie, sensowne lub bezsensowne są poniższe zdania:

- 1- zdecydowanie bezsensowne
- 2- bezsensowne
- 3- raczej bezsensowne

- 4- ani sensowne ani bezsensowne
- 5- raczej sensowne
- 6- sensowne
- 7- zdecydowanie sensowne

Na przykład,

Ta sytuacja to cebula – 1 (zdecydowanie bezsensowne)

Ten mężczyzna to bohater – 7 (zdecydowanie sensowne)

3. Familiarity ratings:

Na skali od 1-7 zaznacz jak często spotykasz się z poniższymi zdaniami:

- 1- bardzo rzadko
- 2- rzadko
- 3- raczej rzadko
- 4- czasem
- 5- raczej często
- 6- często
- 7- bardzo często

Na przykład,

Ten egzamin to mur – 1 (bardzo rzadko)

Ten mężczyzna to dziennikarz – 7 (bardzo często)

4. Metaphoricity ratings:

Na skali 1-7 zaznacz jak bardzo, według Ciebie, dosłowne lub metaforyczne są poniższe zdania:

- 1- bardzo dosłowne
- 2- dosłowne
- 3- raczej dosłowne
- 4- ani dosłowne ani metaforyczne
- 5- raczej metaforyczne

6- metaforyczne

7- bardzo metaforyczne

Na przykład,

Ten mężczyzna do dziennikarz – 1 (bardzo dosłowne)

To spotkanie to wojna – 7 (bardzo metaforyczne)

Ten egzamin jest jak mur – 7 (bardzo metaforyczne)

5. Cloze probability tests:

Do podanej początkowej części zdania dopisz rzeczownik, który pierwszy przyjdzie Ci na myśl, tak aby powstałe zdanie miało sens.

Na przykład,

Ten mężczyzna to bohater

Ta pogoda jest jak koszmar

Appendix J: Experiment 2: Instructions in the normative studies on English experimental sentences

1. Abstractness ratings:

On the scale from 1 to 7, please rate how abstract or concrete you think a particular noun is.

- 1- Definitely abstract
- 2- Mostly abstract
- 3- Rather abstract
- 4- Neither abstract nor concrete
- 5- Rather concrete
- 6- Mostly concrete
- 7- Definitely concrete

For instance,

love – 1 (highly abstract)

computer – 7 (highly concrete)

2. Meaningfulness ratings:

On the scale from 1-7, please rate how meaningful or meaningless, in your opinion, a particular sentence is.

- 1- Totally meaningless
- 2- Meaningless

- 3- Rather meaningless
- 4- Neither meaningless nor meaningful
- 5- Rather meaningful
- 6- Meaningful
- 7- Totally meaningful

For instance,

This weather is an onion – 1 (totally meaningless)

This man is a hero – 7 (totally meaningful)

3. Familiarity ratings:

On the scale from 1-7, decide how often you encounter a particular sentence.

- 1- Very rarely
- 2- Rarely
- 3- Rather rarely
- 4- Sometimes
- 5- Rather frequently
- 6- Frequently
- 7- Very frequently

For instance,

This situation is like an onion – 1 (very rarely)

This man is a hero – 7 (very frequently)

4. Metaphoricity ratings:

On the scale from 1-7, please rate how metaphorical or literal, according to you, a particular sentence is.

- 1- Very literal
- 2- Literal
- 3- Slightly literal
- 4- Neither literal nor metaphorical

5- Slightly metaphorical

6- Metaphorical

7- Very metaphorical

For instance,

This man is a hero – 1 (very literal)

My knowledge is a well – 7 (very metaphorical)

Amnesia is like a rubber – 7 (very metaphorical)

5. Cloze probability tests:

To the presented beginning of a sentence add a noun that first comes to your mind, so that the sentence makes sense.

For instance,

That man is a hero.

This weather is like an apocalypse.

Appendix K: An informed consent form (in Polish)

ZGODA NA UDZIAŁ W BADANIU NAUKOWYM

Oświadczam, że zaznajomiłem/łam się i zrozumiałem/łam informację dla osoby badanej. Wyrażam dobrowolną i świadomą zgodę na udział w badaniu. Jestem również świadomy/a faktu, iż w każdej chwili mogę odstąpić od udziału w badaniu. Wyrażam zgodę na przetwarzanie moich danych uzyskanych w trakcie eksperymentu, zgodnie z ustawą z dnia 29 sierpnia 1997r. o ochronie danych osobowych, lecz wyłącznie w celach naukowych. Niniejszy dokument, potwierdzający zgodę na udział w badaniach będzie przechowywany zgodnie z zasadami przechowywania dokumentacji poufnej.

.....

imię i nazwisko badanego

.....

podpis badanego

.....

data

Oświadczam, że osoba badana zapoznała się z informacją dla uczestnika badania, a dane uzyskane podczas eksperymentu będą przechowywane oraz przetwarzane zgodnie z ustawą z dnia 29 sierpnia 1997r. o ochronie danych osobowych.

.....

podpis prowadzącego badanie

.....

data