The psychological validity of collocation: Evidence from event-related brain potentials

Jennifer Hughes

ESRC Centre for Corpus Approaches to Social Science

Department of Linguistics and English Language

Lancaster University

A thesis submitted to Lancaster University for the degree of Doctor of Philosophy in Linguistics

July 2018

ABSTRACT

Previous studies have used psycholinguistic techniques such as eye-tracking and self-paced reading in order to investigate the psychological validity of corpus-derived collocations (e.g. Conklin & Schmitt 2008; McDonald & Shillcock 2003a; 2003b; Underwood et al. 2004; Huang et al. 2012). The results of these studies reveal that sequences of words which form collocations are read more quickly and receive fewer fixations than sequences of words which do not form collocations. However, behavioural data and eye-tracking data can only ever provide an indirect measure of what is going on in the brain during language processing. In this thesis, I therefore investigate the psychological validity of corpus-derived collocations using a direct measure of neural activity, namely electroencephalography (EEG). More specifically, I use the event-related potential (ERP) technique of analysing brainwave data.

Very few ERP studies focus on collocation, and those that *do* focus on collocation conceptualize and operationalize the notion differently from how it is conceptualized and operationalized in this thesis, or indeed in most corpus linguistics work. For example, although Molinaro and Carreiras (2010:179-180) use corpus-derived collocations for an ERP study, they explicitly state that they only extract collocations which are “idioms or clichés”. By contrast, in this thesis, collocation is conceptualized as a more fluid phenomenon, as compositional or non-compositional word pairs where the words have a high probability of occurring together.

In Experiment 1, which is the first of four ERP experiments presented in this thesis, I aim to pilot a procedure for determining whether or not there is a neurophysiological difference in the way that the native speaker brain processes collocational adjective-noun bigrams compared to non-collocational adjective-noun bigrams. In Experiment 2, I aim to replicate the results of the pilot study using another group of native English speakers; while, in Experiment 3, I aim to investigate the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in *non*-native speakers of English (specifically, native

speakers of Mandarin Chinese). In Experiment 4, the final experiment of this thesis, I then aim to investigate the gradience of the ERP response as well as the psychological validity of different association measures, namely transition probability, mutual information, log-likelihood, z-score, t-score, Dice-coefficient, MI3, and raw frequency.

The results of these studies reveal that there *is* a neurophysiological difference in the way that the brain processes corpus-derived collocational bigrams compared to matched non-collocational bigrams, suggesting that the phenomenon of collocation *can* be seen as having psychological validity. An important finding of this thesis is the discovery of the ‘Collocational N400’: an ERP component reflecting the increase in cognitive load associated with reading a collocational violation. This increase in cognitive load is greater for non-native speakers compared to native speakers, as non-native speakers have less flexibility than native speakers in their use of (non-)collocational patterns. Moreover, while there is a strong correlation between the amplitude of the collocational N400 and all of the measures of collocation strength that I investigate in Experiment 4, the strongest correlations exist between amplitude and the hybrid association measures, including z-score, MI3, and Dice co-efficient. This suggests that mutual information and log-likelihood, which are two of the most commonly used association measures in corpus linguistics (Gries 2014a:37), are not necessarily always the optimal choice. I discuss these results in relation to prior literature from the fields of corpus linguistics and cognitive neuroscience.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Professor Tony McEnery, as well as my former supervisor Dr Andrew Hardie, for their invaluable advice and feedback throughout the PhD process. I am extremely lucky to have had two incredibly supportive supervisors who have been endlessly patient and understanding, and have always let me work at my own pace. I would like to thank Tony for overseeing the project and for providing me with sole access to the EEG equipment, and the space in which to conduct my experiments. I would also like to thank Andrew for providing detailed feedback on every one of my chapters, and for scheduling weekly supervision meetings to help me stay on track. Andrew has always made time to help me and has always gone above and beyond what would be expected. I am forever grateful for his constant support and for how he has always encouraged me and motivated me to continue. I am certain that I could not have completed (or even started) this thesis without Andrew's encouragement.

Secondly, I would like to express my sincere thanks to the Faculty of Arts and Social Science, Lancaster University, for providing me with the Faculty Research Scholarship to carry out this PhD. I would also like to thank the Department of Linguistics and English Language for awarding me with the Geoffrey Leech Scholarship for my MA. Without this funding, I could not have embarked on postgraduate study.

Thirdly, I would like to thank Professor Vincent Reid from the Department of Psychology for the advice and feedback that he provided as an examiner on my PhD confirmation panel. Vincent's input was invaluable in reassuring me that I was making the correct methodological decisions, and in encouraging me to consider alternative explanations of my pilot study results.

I am also extremely grateful for the advice that I received from other members of the Department of Psychology. First, I would like to thank Dr Eugenio Parise, who provided me with background information on EEG/ERPs as well as initial guidance on the steps involved in conducting an ERP experiment. Without this guidance, it would have been very difficult to navigate the EEG/ERP literature. Secondly, I would like to thank Electronics/Research Technician Dave Gaskell, who helped me with the initial setup of the EEG equipment. Dave not only helped me to get the equipment up and running, and made sure it was all working as it should, but also gave up his time to oversee the first test-run of my pilot study. This was invaluable in increasing my confidence in using the EEG equipment, and in allowing me to solve any additional technical problems on my own. Thirdly, I would like to thank psychology PhD students Han Ke and Pèi Liú, who were conducting ERP experiments at the same time as I was. Their advice was incredibly helpful at different stages of the project, from participant recruitment to the processing and interpreting of the data. I am very grateful for their fast and informative responses, and their reassurance that I was on the right track.

In addition to the help that I received from individual members of the Department of Psychology, I was also able to audit two courses in this department, namely advanced courses in statistics and neuropsychology. These courses were extremely helpful in filling in the gaps in my knowledge of statistics and psychology, and thereby increasing my confidence in my ability to conduct statistical analyses and to critically review the psychology literature.

Next, I would like to express my sincere thanks to the 76 participants who so kindly gave up their time to take part in my experiments. This thesis literally would not exist without them. Similarly, I would like to thanks my friends and fellow CASS PhD students Rosie Harvey, Abi Hawtin, Gillian Smith, Amelia Joulain-Jay, and Ruth Byrne, for kindly volunteering to be test participants when I was testing the equipment and the experimental procedure. They were all incredibly patient when the test-run inevitably revealed technical problems and delays.

The friendships that I made with the CASS PhD students helped to make the PhD process an incredibly enjoyable and exciting experience. I would therefore like to thank the CASS PhD students, not only those mentioned above, but also Robbie Love, Mathew Gillings, Andressa Rodrigues Gomide, and Irene Marin Cervantes, for their warmth and friendship over the last few years. We have shared many milestones together throughout this process, from giving our first conference presentations and teaching our first seminar groups, to passing our upgrade panels and completing drafts of chapters. Socialising with the CASS students has provided me with some amazing memories of my time as a PhD student.

As well as thanking the CASS PhD students, I would also like to thank every other member of the ESRC Centre for Corpus Approaches to Social Science. Being part of such as active research centre has really enhanced my PhD experience by enabling me to assist in other projects, and by giving me a clear insight into academic life. This has really motivated me to work to the best of my ability and to aim to produce work that is of publication quality. Being part of CASS has also enabled me to access funding to attend conferences and workshops, and for that I am incredibly grateful. Attending the Birmingham ERP Bootcamp ran by renowned ERP researchers Professor Steven Luck and Dr Emily Kappenman was particularly useful, as the practical knowledge that I gained at this bootcamp massively increased the efficiency of my data processing, and the theory-based lectures really consolidated my knowledge of ERPs.

Next, I would like to thank my A Level English Language teacher Frank Pearson, for instilling me with my love for English language and linguistics, and for encouraging me to follow this passion at Lancaster University. I feel very lucky to have been able to undertake a PhD here at Lancaster at the same time as Frank. Our discussions about language and collocation always remind me of why I am interested in my topic, and in language and linguistics more generally.

I would also like to thank my amazing friends for their support and encouragement over the last few years. I would especially like to thank Megan Woodburn, who has been my best friend for as long as I can remember, for encouraging me in everything that I do. Special thanks also go to Rebecca Wilson and Matthew Howard for always making me laugh, and for attempting to motivate me by periodically asking how many words and references I’ve got (and for the mathematical knowledge shared by Rebecca). I would also like to thank the fabulous dancers and teachers at KNT Danceworks Liverpool, especially Adele Peaty, Jennifer Hynes, Eloise Shiner, Aimée Harris, Holly Blue, Sophia Komninou, Charlotte Weatherstone, Mark Hindle, and Sophie Taylor. These individuals motivate me to constantly strive to improve in every area of my life.

Finally, I would like to thank my Mum and Dad for supporting me in every possible way. I am incredibly lucky to have such amazing parents who provide endless love and encouragement, and I am deeply grateful for everything that they do. I would also like to thank Ralph, Harriet, and Charlotte for the memories that we have made together, and for being the best possible companions on the many nights spent working on this thesis.

TABLE OF CONTENTS

ABSTRACT .... **i**
ACKNOWLEDGEMENTS .... **iii**
TABLE OF CONTENTS .... **vii**
LIST OF FIGURES .... **xiii**
LIST OF TABLES .... **xvii**

CHAPTER 1: INTRODUCTION **1**
1.1 Overview of the chapter 1
1.2 Introduction to corpus linguistics 1
1.3 Introduction to cognitive neuroscience 4
1.4 Combining corpus linguistics and cognitive neuroscience 8
1.5 Aims of the thesis 10
1.6 Outline of the thesis 15

CHAPTER 2: A REVIEW OF THEORETICAL AND PRACTICAL ISSUES IN THE STUDY OF COLLOCATION **16**
2.1 Overview of the chapter 16
2.2 Definitions 16
2.3 Overview of different approaches to collocation 19
2.4 The neo-Firthian approach 22
2.4.1 Sinclairian theory 22
2.4.1.1 The Idiom Principle 22
2.4.1.2 Problems with Sinclairian theory 23
2.4.2 Lexical Priming 26
2.4.3 Pattern Grammar 30
2.5 The syntax-based approach 34
2.6 Formulaicity 37
2.7 The usage-based approach 39
2.8 How the different approaches represent collocation at a psychological level: Introducing the network model of language processing 44
2.9 Main parameters of difference between the different approaches 50
2.10 Defining and measuring psychological validity 52
2.11 Association measures and psychological validity 53

2.12 Chapter summary and conclusion ... 56

CHAPTER 3: A REVIEW OF PRIOR WORK ON THE PSYCHOLOGY OF COLLOCATION ... 57

3.1 Overview of the chapter ... 57

3.2 Self-paced reading studies ... 58

3.3 Eye-tracking studies ... 62

3.4 Electrophysiological studies: Electroencephalography and event-related potentials ... 64

3.4.1 Background information on EEG ... 64

3.4.1.1 Overview ... 64

3.4.1.2 Source of electrical activity ... 66

3.4.2 Background information on ERPs and ERP components ... 69

3.4.3 The N400 as an electrophysiological marker of semantic processing ... 72

3.4.4 The P600 as an electrophysiological marker of syntactic processing ... 74

3.4.5 Relationship between semantic and syntactic processing ... 75

3.4.6 ERP studies of collocation ... 76

3.4.6.1 The explicit approach ... 76

3.4.6.2 The non-explicit approach ... 82

3.4.7 ERP studies of predictability ... 89

3.4.8 Advantages and disadvantages of the ERP technique ... 95

3.5 Chapter summary and conclusion ... 99

CHAPTER 4: METHODOLOGY ... 101

4.1 Overview of the chapter ... 101

4.2 Selection of stimuli ... 101

4.2.1 Selection of bigrams ... 101

4.2.2 Selection of sentences ... 108

4.3 Presentation of stimuli ... 117

4.4 Electrode placement ... 121

4.5 Experimental setup ... 126

4.6 Data analysis and interpretation ... 136

4.6.1 Overview ... 136

4.6.2 Filtering ... 138

4.6.3 Re-referencing ... 139

4.6.4 Segmentation and baseline correction ... 141

4.6.5 Artifact detection and rejection 142
4.6.6 Averaging 144
4.6.7 Computation and quantification of ERPs 144
4.6.8 Statistical analysis 148
4.6.9 Plotting ERP waveforms 151
4.7 Chapter summary and conclusion 152

CHAPTER 5: EXPERIMENT 1 – A pilot study 154
5.1 Overview of the chapter 154
5.2 Aims and hypotheses: Experiment 1 154
5.3 Methodology: Experiment 1 155
5.3.1 Participants 155
5.3.2 Stimuli 156
5.3.3 Procedure 163
5.3.4 Data analysis 163
5.4 Results: Experiment 1 163
5.4.1 Component-based experimental design 163
5.4.1.1 N400 (300 - 500 ms) 163
5.4.1.2 P600 (500 - 800 ms) 170
5.4.1.3 Unexpected results 174
5.4.1.3.1 Overview 174
5.4.1.3.2 P250 (200 - 300 ms) 176
5.4.1.3.3 P750 (700 - 800 ms) 178
5.4.2 Component-independent experimental design 181
5.5 Discussion: Experiment 1 184
5.5.1 Summary of aims and results 184
5.5.2 N400 and P600 185
5.5.3 Onset latency 189
5.5.4 P250 190
5.5.5 P750 193
5.5.6 Limitations of the pilot study 197
5.6 Chapter summary and conclusion 200

CHAPTER 6: EXPERIMENT 2 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in native speakers of English 202

6.1 Overview of the chapter 202

6.2 Aims and hypotheses: Experiment 2 202

6.3 Methodology: Experiment 2 203

6.3.1 Participants 203

6.3.2 Stimuli 203

6.3.3 Procedure 204

6.3.4 Data processing 208

6.3.5 Data analysis 211

6.4 Results: Experiment 2 213

6.4.1 Component-based experimental design 213

6.4.1.1 N400 (350 - 500 ms) 213

6.4.1.1.1 Initial N400 analysis (anterior central electrode sites) 213

6.4.1.1.2 New N400 analysis (all electrode sites) 216

6.4.1.3 P600 (500 - 650 ms) 218

6.4.2 Component-independent experimental design 220

6.5 Discussion: Experiment 2 220

6.5.1 Summary of aims and results 220

6.5.3 Issues raised by the results 222

6.6 Chapter summary and conclusion 224

CHAPTER 7: EXPERIMENT 3 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in non-native speakers of English 225

7.1 Overview of the chapter 225

7.2 Aims and hypotheses: Experiment 3 225

7.3 Methodology: Experiment 3 227

7.3.1 Participants 227

7.3.2 Stimuli, procedure, and data analysis 229

7.4 Results: Experiment 3 229

7.4.1 Component-based experimental design 229

7.4.1.1 N400 (350 - 500 ms) 229

7.4.1.2 P600 (500 - 650 ms) 232

7.4.2 Component-independent experimental design 234

7.5 Discussion: Experiment 3 235
7.5.1 Summary of aims and results 235
7.5.2 Issues raised by the results 237
7.6 Chapter summary and conclusion 239

CHAPTER 8: EXPERIMENT 4 – An investigation into the psychological validity of different measures of collocation strength 240
8.1 Overview of the chapter 240
8.2 Aims and hypotheses: Experiment 4 240
8.3 Methodology: Experiment 4 241
8.3.1 Participants 241
8.3.2 Stimuli 242
8.3.3 Procedure 247
8.3.4 Data analysis 249
8.4 Results: Experiment 4 252
8.4.1 Part 1: Replication of Experiment 2 252
8.4.1.1 Component-based experimental design 252
8.4.1.1.1 N400 (350 - 500 ms) 252
8.4.1.1.2 P600 (500 - 650 ms) 255
8.4.1.2 Component-independent experimental design 257
8.4.2 Part 2: Correlational analysis 258
8.5 Discussion: Experiment 4 271
8.5.1 Summary of aims and results 271
8.5.2 Issues raised by the results 272
8.6 Chapter summary and conclusion 275

CHAPTER 9: GENERAL DISCUSSION AND CONCLUSION 277
9.1 Overview of the chapter 277
9.2 Summary of findings 277
9.2.1 Experiment 1 277
9.2.2 Experiment 2 278
9.2.3 Experiment 3 280
9.2.4 Experiment 4 281
9.2.5 Overall findings 282
9.3 Implications of the results 284

9.3.1 One or two processing systems?..........284
9.3.2 Network model of language processing..........286
9.3.3 Cognitive neuroscience..........294
9.4 Issues raised by the results..........298
9.4.1 Inconsistencies in scalp distribution findings..........298
9.4.2 Elicitation of a putative P600 in the collocational condition..........300
9.4.3 Conflicting conclusions from N400 and onset latency results..........301
9.5 Limitations of the study and directions for future research..........301
9.6 Conclusions..........305

REFERENCES..........308

APPENDICES..........338
Appendix 1: Counterbalanced stimuli lists used in Experiment 1..........338
Appendix 2: Counterbalanced stimuli lists used in Experiments 2 and 3..........347
Appendix 3: Counterbalanced stimuli lists used in Experiment 4..........353
Appendix 4: Information sheet for native speakers of English..........357
Appendix 5: Information sheet for non-native speakers of English..........359
Appendix 6: Consent form for native and non-native speakers of English..........361

LIST OF FIGURES

CHAPTER 3: A REVIEW OF PRIOR WORK ON THE PSYCHOLOGY OF COLLOCATION

Figure 3.1: The divisions of the human brain (from Pinel 2014:68)…………………………..66
Figure 3.2: The lobes of the cerebral hemispheres (from Gray 1918)……………………….....67
Figure 3.3: Basic structure of a cortical nerve cell (from Nolen-Hoeksema et al. 2009:37)…68
Figure 3.4: An ERP waveform displaying the N400 and P600 (adapted from Swaab et al. 2012:422)………………..…………………..…………………………………..…72

CHAPTER 4: METHODOLOGY

Figure 4.1: Example stimuli list; matched experimental sentences presented at least 5 sentences apart in order to reduce the memory trace and therefore limit repetition priming…....113
Figure 4.2: Ordering of experimental sentences in counterbalanced lists (collocational items come first in List A; non-collocational items come first in List B) ………………………………………………………………………………115
Figure 4.3: Ordering of experimental sentences in counterbalanced lists (half of the collocational items come first in List C; the other half of the collocational items come first in List D).. …………………..……………………………………………………116
Figure 4.4: Screenshot of main instruction screen taken from Experiment 4……………...…120
Figure 4.5: Recently updated version of the International 10-20 (or 10-10) System (from American Clinical Neurophysiology Society 2006:223) ……………………………122
Figure 4.6: BioSemi electrode positions (from the BioSemi website: http://www.biosemi.com)………………………..………………………………124
Figure 4.7: Experimental setup…………………………………………………………...…128
Figure 4.8: Positioning of the electrodes on a test-run participant (Gillian Smith, PhD student in LAEL, Lancaster University) ……………………………………………….…131
Figure 4.9: Problematic waveforms (from Luck 2014a:203) ………………….……………133
Figure 4.10: How a blink appears on the electroencephalogram (from Luck 2014a:195)…...135
Figure 4.11: Data processing operations…………………………………………….………137
Figure 4.12: BioSemi electrode positions grouped for statistical analysis. Purple lines designate left-to-right electrode position while red lines designate anterior-to-posterior electrode position..……………………………………………………….………………149

CHAPTER 5: EXPERIMENT 1 – A pilot study

Figure 5.1: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus. The box on each waveform illustrates the typical latency range of the N400 (300-500 ms)..……………………………………165
Figure 5.2: BioSemi electrode positions grouped into nine zones for statistical analysis. Purple lines designate left-to-right electrode position while red lines designate anterior-to-posterior electrode position. The electrode sites used to represent each zone are circled in red……………...…………………………………………………………166
Figure 5.3: Topographic scalp maps showing mean amplitude between 300 and 500 ms…...168
Figure 5.4: Grand average ERPs across mid anterior and mid central electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus……………………………………………169
Figure 5.5: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus……………………………………………171
Figure 5.6: Topographic scalp maps showing mean amplitude between 500 and 800 ms…...172
Figure 5.7: Grand average ERPs across mid-central and mid-posterior electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus…………………………………….……173
Figure 5.8: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus. The blue box with the solid outline illustrates the latency range of the P250 (200-300 ms); the purple box with the dashed outline illustrates the latency range of the P750 (700-800 ms) ……………………………...175
Figure 5.9: Topographic scalp maps showing mean amplitude between 200 and 300 ms….177
Figure 5.10: Grand average ERPs across mid central and mid posterior electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus…………………………………….…178
Figure 5.11: Topographic scalp maps showing mean amplitude between 700 and 800 ms….180
Figure 5.12: Grand average ERPs across mid central and mid posterior electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus…………………………………….……181

CHAPTER 6: EXPERIMENT 2 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in native speakers of English

Figure 6.1: Positioning of breaks within and between trial blocks…………………………..206
Figure 6.2: EKG activity in channel data from Participant 14………………………………210
Figure 6.3: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus (initial N400 analysis) …………………...214

Figure 6.4: Topographic scalp maps showing mean amplitude between 350 and 500 ms…...215
Figure 6.5: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus (new N400 analysis)……………………..217
Figure 6.6: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus……………………………………………219

CHAPTER 7: EXPERIMENT 3 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in non-native speakers of English

Figure 7.1: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus……………………………………………230
Figure 7.2: Topographic scalp maps showing mean amplitude between 350 and 500 ms…231
Figure 7.3: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus……………………………………………233
Figure 7.4: Topographic scalp maps showing mean amplitude between 500 and 650 ms…234

CHAPTER 8: EXPERIMENT 4 – An investigation into the psychological validity of different measures of collocation strength

Figure 8.1: Positioning of breaks within and between trial blocks…………………………..248
Figure 8.2: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus……………………………………………253
Figure 8.3: Topographic scalp maps showing mean amplitude between 300 and 500 ms…..254
Figure 8.4: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus……………………………………………256
Figure 8.5: Topographic scalp maps showing mean amplitude between 500 and 650 ms…..257
Figure 8.6: Scatterplot showing the relationship between amplitude and transition probability.. ………………………………………………………………………………259
Figure 8.7: Difference waves at Cz for each bigram pair, from -200 ms pre-stimulus to 800 ms post-stimulus. The box on each waveform shows the N400 latency range……..……261
Figure 8.8: Scatterplot showing the relationship between amplitude and Z-score…………...264
Figure 8.9: Scatterplot showing the relationship between amplitude and log-likelihood……265
Figure 8.10: Scatterplot showing the relationship between amplitude and MI3……………..266
Figure 8.11: Scatterplot showing the relationship between amplitude and Dice coefficient...267
Figure 8.12: Scatterplot showing the relationship between amplitude and T-score…………268
Figure 8.13: Scatterplot showing the relationship between amplitude and frequency……….269

Figure 8.14: Scatterplot showing the relationship between amplitude and MI……………..270

CHAPTER 9: GENERAL DISCUSSION AND CONCLUSION

Figure 9.1: Scatterplot showing the relationship between amplitude and backward transition probability…………………………………………………………………..…293

LIST OF TABLES

CHAPTER 3: A REVIEW OF PRIOR WORK ON THE PSYCHOLOGY OF COLLOCATION

Table 3.1: Examples of experimental sentences varying in degree of contextual constraint and cloze probability (table adapted from Kutas and Hillyard 1984:161)…………..…….83
Table 3.2: Stimuli types used by Rhodes and Donaldson (2008:52)…………………………85
Table 3.3: Stimuli types used by Lau et al. (2016:4)………………………………………87
Table 3.4: Summary of previous ERP studies of collocation……………………………..93-94

CHAPTER 5: EXPERIMENT 1 – A pilot study

Table 5.1: Experimental bigrams for Experiments 1, 2, and 3……………………………….156
Table 5.2: Statistical properties of Experiment 1/2/3 bigrams extracted from written BNC1994……………………………………………………………………….157
Table 5.3: Experimental sentences for Experiments 1, 2, and 3…………………………..…159
Table 5.4: Fillers and practice sentences…………………………………………………..…162
Table 5.5: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 300-500 ms latency range…………………………………...167
Table 5.6: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 500-800 ms latency range…………………………………...172
Table 5.7: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 200-300 ms latency range…………………………………….177
Table 5.8: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 700-800 ms latency range…………………………………….179
Table 5.9: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites…………………………………………………......…………..183

CHAPTER 6: EXPERIMENT 2 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in native speakers of English

Table 6.1: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites at the 350-500 ms latency range (initial N400 analysis)……………………..215

Table 6.2: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites in the 350-500 ms latency range (new N400 analysis)…………………………218
Table 6.3: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites in the 500-650 ms latency range……………………………………………….220

CHAPTER 7: EXPERIMENT 3 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in non-native speakers of English

Table 7.1: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites at the 350-500 ms latency range……………………………………..………231

CHAPTER 8: EXPERIMENT 4 – An investigation into the psychological validity of different measures of collocation strength

Table 8.1: Bigrams for Experiment 4………………………………………………………..…244
Table 8.2: Statistical properties of Experiment 4 bigrams extracted from written BNC1994..245
Table 8.3: Experimental sentences for Experiment 4………………………………..……246
Table 8.4: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites in the 350-500 ms latency range…………………………………..…254
Table 8.5: Summary of post hoc ANOVA results carried out at anterior and central electrode sites……………………..……………………………………………………...…258
Table 8.6: Association measures ranked by strength of correlation……………………....…262
Table 8.7: Amplitude and association scores for each collocational bigram……………..…263

CHAPTER 9: GENERAL DISCUSSION AND CONCLUSION

Table 9.1: Association measures ranked by strength of correlation (including backward transition probability) ……………………..……………………………..……294
Table 9.2: Comparison table of results from previous studies of collocation and the N400/P600, and results from the present study…………………………………………………...297

CHAPTER 1: INTRODUCTION

1.1 Overview of the chapter

In this chapter, I introduce the two disciplines on which this thesis is based, namely corpus linguistics (section 1.2) and cognitive neuroscience (section 1.3). In section 1.4, I explain how I combine these two disciplines by using corpus-derived collocations as stimuli for a series of four ERP experiments (section 1.4). I then outline the aims and hypotheses of this thesis in section 1.5, including both the overarching aim and hypothesis, and the aims and hypotheses associated with each individual experiment. Finally, I end the chapter in section 1.6 by providing an overview of the structure of the thesis.

1.2 Introduction to corpus linguistics

Corpus linguistics is the study of large bodies of computerized text, where "the set of texts or *corpus* dealt with is usually of a size which defies analysis by hand and eye alone within any reasonable timeframe" (McEnery & Hardie 2012:1-2). A corpus contains samples of written and/or spoken texts (i.e. transcriptions of authentic speech) (Leech et al. 1995). These texts are often annotated to include linguistic information such as semantic (i.e. meaning-based) category or parts of speech (McEnery & Wilson 2004:32; Hyland 2015:301). A part of speech label (technically known as a *tag*) indicates the grammatical category of a word, such as noun or verb, in the context of the sentence in which it is found (Leech 1987:8; McEnery & Hardie 2012:13). Annotating texts in this way allows researchers to search the corpus for particular semantic features, grammatical categories, or grammatical structures (Leech 2015:148).

The texts are selected strategically, usually with the aim of compiling a balanced collection of texts that can be seen as being representative of a particular language variety

(Leech 1992:116). For example, the British National Corpus 1994 (henceforth BNC1994[1]) contains over 100 million words from over 4000 samples of British English (Aston & Burnard 1998:i). These 4000 samples span a variety of different kinds of texts, both written and spoken, including genres as diverse as fiction books, academic books, newspapers, lectures, radio programmes, classroom interaction, and casual conversation (Aston & Burnard 1998:5, 31-32). The BNC1994 is therefore considered to be fairly representative of British English, at least for the time period from which the texts were selected (Aston & Burnard 1995:30).

Importantly, the texts which make up a corpus are naturalistic and authentic in the sense that they are actual examples of language in use, rather than being linguistic examples invented solely for the purpose of linguistic analysis. This means that corpus-based analyses reflect how language is actually used in the real world (McEnery & Wilson 2004:104). This is in direct contrast to how language is analysed under the Chomskyan tradition, within which linguistic conclusions are based on intuition and invented examples (e.g. Chomsky 1957:15). With the advent of corpus linguistics, it has become apparent that "intuition can be unreliable when it comes to making judgements about language in use, and there are certain aspects of language that are simply not open to intuition, such as word frequency distributions" (Adolphs 2006:7). The use of authentic text is therefore a major advantage of corpus linguistics.

There are four fundamental techniques of corpus analysis, namely keywords, frequency lists, concordances, and collocations. A *keyword* is defined as "a word that is more frequent in a text or corpus under study than it is in some (larger) reference corpus, where the difference in frequency is statistically significant" (McEnery & Hardie 2012:245). Keywords essentially give an overview of what the text or corpus is "about", and what distinguishes one particular

---

[1] This corpus was originally named the BNC, but is retroactively referred to as the BNC1994 following the creation of the BNC2014 (Love 2017; Love et al. 2017; Hawtin forthcoming).

text or corpus from other texts or corpora (Hyland 2015:300). For instance, the keywords in a corpus of texts from the natural sciences include *fragments*, *surface*, and *analysis*, while the keywords in a corpus of texts from the discipline of medicine include *clinical*, *patients*, and *treatment* (Scott & Tribble 2006:83).

A *frequency list* is defined as “a list of all the items of a given type in a corpus (for example, all words; all part-of-speech tags; all four-word sequences) together with a count of how often each one occurs” (McEnery & Hardie 2012:243). While the raw frequencies provided by corpus frequency lists can be useful in some contexts, these frequencies need to be normalized, or relativized (e.g. to frequency per thousand or million words), before they can be compared with the frequencies from another text or corpus (McEnery & Wilson 2004:83; Kirk 2009:33; Gries 2015:52). See Chapter 4, section 4.2.1, for use of frequency lists in this thesis.

Central to corpus analysis is the *concordance*, which is “a display of every instance of a specified word or other search term in a corpus, together with a given amount of preceding and following context” (McEnery & Hardie 2012:241). Concordance analysis has many useful functions including disambiguating word senses (e.g. financial *bank* versus river *bank*), and identifying common patterns of usage and occurrence (Sinclair 1994:4). See Chapter 4, section 4.2.2, for use of concordances in this thesis.

The final major corpus analysis technique, namely collocation analysis, is the main focus of this thesis. There are many different approaches to the study of collocation, with many different conceptualizations and operationalizations of the same concept (Nesselhauf 2004:1) (see Chapter 2, section 2.3). Indeed, Wray (2002:9) lists 57 words that have been used to refer to the phenomenon of collocation. However, in the broadest sense, a collocation can be defined as a “co-occurrence relation between two words” (McEnery & Hardie 2012:240). McEnery and Hardie (2012:240) state that "[w]ords are said to *collocate* with one another if one is more likely to occur in the presence of the other than elsewhere".

While frequency lists and keywords represent types of quantitative corpus analysis (Adolphs 2006; Gries 2015), and concordance analysis is a form of qualitative corpus analysis (McEnery & Hardie 2012:2), collocation analysis can be either qualitative or quantitative. Qualitative collocation analysis is conducted via the concordance-based approach, which involves manually reading concordance lines in order to identify co-occurrence patterns (McEnery & Hardie 2012:125). By contrast, quantitative collocation analysis is conducted via the statistical approach, which involves identifying collocations using association measures. It is this latter approach to collocation analysis that is used in this thesis.

Association measures are statistical scores which allow us to distinguish between words which co-occur due to chance, and words which co-occur due to true statistical association (Evert 2008:32). There are many different association measures, which are introduced in Chapter 2 (section 2.11), but the one used in this thesis is *transition probability*. The transition probability of a pair of adjacent words (i.e. *bigrams*) is calculated by dividing the number of times the bigram X-then-Y occurs in a corpus by the number of times X occurs in the corpus altogether (McEnery & Hardie 2012:195). The justification for using this measure of collocation strength is provided in Chapter 4 (section 4.2.1).

1.3 Introduction to cognitive neuroscience

Having introduced the field of corpus linguistics, I will now introduce the field of cognitive neuroscience. Cognitive neuroscience is an interdisciplinary field which combines the psychological disciplines of neuroscience and cognitive psychology. Cognition is defined as "the collection of mental processes and activities used in perceiving, remembering, thinking, and understanding, as well as the act of using those processes" (Ashcraft & Radvansky 2010:9). Cognitive psychology is the scientific study of those mental activities and processes or, more generally, “the scientific study of the mind” (Ashcraft & Radvansky 2010:2, 27). Cognitive psychologists study the mind independently of the structural brain (Johnson 1997:xv; Churchland & Sejnowski 2000:14). By contrast, neuroscientists study the structural brain independently of the mind; neuroscience is the “branch of biology dealing with the brain and central nervous system” (Atkinson et al. 1996:24).

Churchland and Sejnowski (2000:14) state that:

> [i]n the past, discoveries at the neuronal level and explanations at the cognitive level were so distant that each often seemed of merely academic significance to the other […] It would be convenient if we could understand the nature of cognition without understanding the nature of the brain itself. Unfortunately, it is difficult if not impossible to theorize effectively on these matters in the absence of neurobiological constraints.

Similarly, Johnson (1997:xv) argues that:

> information from brain development is more than just a useful additional source of evidence for supporting particular cognitive theories. Rather, information about brain development is viewed as both changing and originating theories at the cognitive level.

This integration of cognitive and biological information is the domain of cognitive neuroscience.

Cognitive neuroscience differs from the related discipline of cognitive neuropsychology in that, whereas cognitive neuroscience “focuses on normal cognitive functioning” (Johnson (1997:xv), cognitive neuropsychology typically focuses on what can be learnt from studying brain-damaged patients (Ellis & Young 2004:4). For example, the discovery of Broca’s area and Wernicke’s area, which are thought to be functionally related to language production and comprehension, respectively, is an achievement of cognitive neuropsychology rather than cognitive neuroscience (Beaumont 2008:137-138).

Neuropsychology is also referred to as behavioural neurology, behavioural neurophysiology, or biological psychology (Rosenzweig et al. 1996:5).

While some researchers, such as Elias and Saucier (2006:3) consider cognitive neuroscience and cognitive neuropsychology to be interchangeable labels for the same discipline, one of the founders of cognitive neuroscience, Gazzinga (2000:4), explicitly states that "neuropsychology was not what we had in mind. Tying specific functions to lesioned brain areas was not going to be our expertise". Instead, cognitive neuroscientists make extensive use of neuroimaging, or brain imaging, techniques (Gazzaniga et al. 2014:14).

Modern neuroimaging techniques include *positron emission tomography* (PET), *functional magnetic resonance imaging* (fMRI), *electroencephalography* (EEG), and *magnetoencephalography* (MEG) (Raichle 2001:3). PET and fMRI are haemodynamic techniques, meaning that they measure the blood flow in the brain (Chen et al. 2008:1). An increase in blood flow to a particular brain region is linked to an increase in neural activity in that region (Aguirre 2014:10). Haemodynamic techniques allow researchers to investigate *where* in the brain the neural activity is taking place in response to a particular stimulus (Aguirre 2014:11). However, since changes in blood flow lag seconds behind changes in neural activity, hemodynamic techniques do not allow researchers to investigate precisely *when* the neural activity is taking place (Aguirre 2014:11). In other words, haemodynamic techniques have a high spatial resolution but a low temporal resolution (Hillyard 2000:25). By contrast, EEG and MEG are electromagnetic techniques, which have a high temporal resolution but a low spatial resolution (Hillyard 2000:25). I will now provide a very brief explanation of how each of these neuroimaging techniques work.

PET is an invasive haemodynamic technique which involves injecting a radioisotope with a short half-life into the bloodstream of a participant (Elias & Saucier 2006:95). The radioisotope is then carried to the brain, where it is taken up by nerve cells (Aguirre 2014:10).

Nerve cells are also known as *neurons* (Pinel 2014:81). As the radioisotope decays in the body, electromagnetic radiation is emitted, and it is this radiation which is detected outside the body using a PET scanner which surrounds the participant's head (Aguirre 2014:10).

fMRI is a less invasive haemodynamic technique because it does not involve the injection of a radioisotope. Instead, it involves the participant lying in an fMRI scanner which detects the oxygen level of the blood (Aguirre 2014:10). More specifically, the fMRI scanner detects the magnetic properties of the iron atoms inside *haemoglobin* – "the primary oxygen-carrying molecule in the blood" (Aguirre 2014:10). Since an increase in blood flow to a particular brain region is linked to an increase in neural activity in that region, blood oxygen levels serve as a proxy for neural activity (Aguirre 2014:10).

Whereas PET and fMRI provide an indirect measure of brain activity, using blood flow as a proxy for neural activation, the electromagnetic techniques of MEG and EEG are the only neuroimaging techniques which provide a direct measure of brain activity (Papadelis et al. 2015:6). In MEG experiments, external sensors detect the *magnetic* signals that are produced by neural activity (Beaumont 2008:283); in EEG experiments, electrodes placed across the scalp detect the *electrical* signals that are produced by neural activity (Harley 2008:17, 494). As mentioned earlier, both MEG and EEG have a high temporal resolution but a low spatial resolution (Hillyard 2000:25). Even so, the spatial resolution of MEG is somewhat higher than that of EEG (Luck 2012:44). In EEG experiments, effects cannot be accurately localized because "the high resistance of the skull relative to the low resistance of the underlying brain and overlying scalp causes the voltage to spread laterally as it travels" (Kappenman & Luck 2011:7). By contrast, magnetism is not so strongly affected by the presence of the skull (Luck 2012:44, 351). However, MEG experiments are much less common than EEG experiments as they are significantly more expensive (Luck 2012:44).

The electrical activity detected by EEG experiments can be analysed in either the time domain or the frequency domain (see Chapter 3, section 3.4.1). Language-related EEG studies almost always focus on the time domain (Kutas and Van Petten 1994:88), which involves the analysis of *event-related potentials* (ERPs), i.e. “the momentary changes in electrical activity of the brain when a particular stimulus is presented to a person” (Ashcraft & Radvansky 2010:61). In turn, ERP experiments typically involve the analysis of *ERP components*. An ERP component can provisionally be defined as “a scalp-recorded voltage change that reflects a specific neural or psychological process” (Kappenman & Luck 2012:4).

The two most widely-studied language-related ERP components are known as the *N400* and the *P600*. ERP components are labelled according to their polarity (positive or negative voltage) and the time at which they occur in relation to the onset of the stimulus (Kutas & Van Petten 1994:86; Kolb & Whishaw 1996:142; Elias & Saucer 2006:91). The N400 is a negative shift in the ERP waveform that occurs roughly 400 ms after stimulus onset, and it is elicited in response to reading a semantic error (Kutas & Hillyard 1980); the P600 is a positive shift that occurs roughly 600 ms after stimulus onset, and it is elicited in response to reading a syntactic error (Osterhout & Holcomb 1992). In this thesis, I conduct a series of four ERP experiments focusing on the N400 and P600, as well as a component-independent measure known as *onset latency*. Onset latency is the point in time at which the waveforms in two conditions begin to diverge (Kappenman & Luck 2011:23).

1.4 Combining corpus linguistics and cognitive neuroscience

Arppe et al. (2010:6) point out that “linguists have made relatively few efforts up until now to test the cognitive reality of corpora”. Likewise, Gries (2014a:12) argues that “there will be, and should be, an increase of corpus-based studies that involve at least some validation against experimental data”. Some previous psycholinguistic studies have attempted to ascertain the psychological validity of corpus-derived collocations using techniques such as eye-tracking or self-paced reading.

In eye-tracking experiments, participants are required to read text on a computer screen whilst an eye-tracking camera tracks their pupil and corneal reflection (Holmqvist et al. 2011:181). The cognitive load required to process each word is inferred from measures such as the number and duration of fixations, i.e. the points at which the eyes are almost still (Rayner 1998:373). Similarly, in self-paced reading experiments, participants are required to silently read either individual words or short phrases on a computer screen, and then press a specified key on the keyboard to reveal the next word or phrase. The amount of time taken to read each word or phrase is assumed to reflect the cognitive load associated with processing that element (McDonough & Trofimovich 2012:119).

Results from eye-tracking and self-paced reading experiments reveal that sequences of words which form collocations are read more quickly and receive fewer fixations than sequences of words which do not form collocations (e.g. Conklin & Schmitt 2008; McDonald & Shillcock 2003a; 2003b; Underwood et al. 2004; Huang et al. 2012; see Chapter 3, sections 3.2 and 3.3). However, behavioural data and eye-tracking data can only ever provide an indirect measure of what is going on in the brain during language processing, as they measure reading time and eye movements as opposed to neural activity. Millar (2010:277) therefore suggests using EEG, as this method provides a direct measure of neural activity by recording the brain’s electrophysiological response to stimuli (albeit from the surface of the scalp rather than invasively).

Very few EEG/ERP studies focus on collocation, and those that do focus on collocation conceptualize and operationalize the notion differently from how it is conceptualized and operationalized in this thesis, or indeed in most corpus linguistics work (see Chapter 3, section 3.4.6). For example, although Molinaro and Carreiras (2010) extract collocations from a corpus for an ERP study, they explicitly state that they only extract collocations which are "idioms or clichés". Idioms are a special case of collocation as they are non-compositional (Taylor 2002:100), i.e. they are a "phrase or other multi-word unit whose meaning cannot be deduced by combining the meanings of the words within it" (McEnery & Hardie 2012:244; see Chapter 2, section 2.2). By contrast, in this thesis, collocation is conceptualized as a more fluid phenomenon. Collocations are not necessarily fixed non-compositional phrases. Rather, they are compositional or non-compositional word pairs where the words have a high probability of occurring together.

1.5 Aims of the thesis

In this thesis, I combine methods from corpus linguistics and cognitive neuroscience by using collocations (specifically, bigrams) extracted from the BNC1994 as experimental stimuli in a series of four ERP experiments. The overarching aim of this thesis is to find out whether or not there is a neurophysiological difference in the way that the brain processes collocational adjective-noun bigrams compared to non-collocational adjective-noun bigrams, and thus contribute to our understanding of whether or not the phenomenon of collocation can be seen as having psychological validity. This thesis extends the work of Millar (2010) and Hughes and Hardie (forthcoming), who use other psycholinguistic methods to explore the psychological validity of collocation. Specifically, through a combination of self-paced reading and eye-tracking techniques, Millar (2010) investigates the processing burden that results from reading *learner collocations*, i.e. collocations which occur in a learner corpus, and which are intuitively unacceptable from a native speaker's perspective. Similarly, Hughes and Hardie (forthcoming) use the self-paced reading technique to investigate whether or not adjective-noun bigrams with a high transition probability are processed more quickly than adjective-noun bigrams with a low transition probability by native and non-native speakers of English.

Millar (2010) finds that collocations that occur in the BNC1994 (e.g. *ideal partner*) are read significantly more quickly than equivalent learner collocations (*best partner*) by native speakers of English. Similarly, Hughes and Hardie (forthcoming) find that bigrams with a high transition probability (e.g. *complex process*) are processed significantly more quickly compared to bigrams with a low transition probability (e.g. *complex question*), by both native and non-native speakers of English. Therefore, based on the results of these psycholinguistic studies, I hypothesize that I *will* find a neurophysiological difference in the way that the brain processes collocational adjective-noun bigrams compared to non-collocational adjective-noun bigrams, providing further evidence that the phenomenon of collocation *can* be said to have psychological validity.

Experiment 1 is a pilot study which compares the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in native speakers of English. The aims of Experiment 1 are as follows:

Aim 1: To pilot a procedure for determining whether or not there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams.

Aim 2: To refine the hypotheses and methods in preparation for Experiments 2 and 3.

Two hypotheses are associated with these aims:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will elicit a P600.

These hypotheses are based on the results of previous ERP studies which find an N400 and a P600 in response to violations in expectancy and predictability (Laurent et al. 2006; Rhodes & Donaldson 2008; Davenport & Coulson 2010; Lau et al. 2016; Siyanova-Chanturia et al. 2017).

See Chapter 3 (sections 3.4.6 and 3.4.7) for an explanation of how these studies relate to the study of collocation.

The purpose of Experiment 2 is to replicate the results from the pilot study in order to provide stronger evidence in support of the idea that there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams. This can be expressed as three distinct aims:

Aim 1: To replicate the finding that a larger N400 is elicited in response to reading non-collocational bigrams compared to collocational bigrams.

Aim 2: To replicate the finding that there is a difference in onset latency between collocational and non-collocational bigrams.

Aim 3: To find out whether or not a P600 is elicited in response to reading non-collocational bigrams (the results regarding this component were inconclusive in the pilot study).

Based on the results of Experiment 1, and on the results from the ERP literature, I make the following hypotheses for Experiment 2:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will elicit a P600.

Hypothesis 3: The onset latency will be greater for the collocational condition compared to the non-collocational condition.

Experiment 3 investigates the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in *non-native* speakers of English (specifically, native speakers of Mandarin Chinese). The aims of Experiment 3 are as follows:

Aim 1: To find out whether or not there is a neurophysiological difference in the way that collocational adjective-noun bigrams and non-collocational adjective-noun bigrams are processed by non-native speakers of English.

Aim 2: To see how the ERP results for the non-native speaker group compare to those of the native speaker group in Experiment 2.

In the self-paced reading study conducted by Hughes and Hardie (forthcoming), the results show that the non-native speakers actually seem to be *more* sensitive to transition probabilities between words than the native speakers. The proposed explanation for this is that the non-native speakers have probably not encountered the bigrams with low transition probabilities before or, if they have, they are unlikely to have encountered them frequently enough for them to become entrenched. By contrast, the native speakers are likely to have encountered the low transition probability bigrams at some point, making them somewhat entrenched, or they will at least have greater flexibility in their use of (non-)collocational patterns than non-native speakers. Based on this result, I predict that the non-native speaker group *will* exhibit a neurophysiological difference in the processing of collocational and non-collocational bigrams, and that this difference will be bigger than that demonstrated by the native English speaker group. In line with the results of Experiment 2, the hypotheses for Experiment 3 are as follows:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will *not* elicit a P600.

Hypothesis 3: The onset latency will be greater for the collocational condition compared to the non-collocational condition.

Hypothesis 4: The ERP responses will be larger than those demonstrated by the native English speakers in Experiment 2.

According to Millar (2010:275), “there is a need for more studies to validate the numerous statistical measures available to corpus linguists”. Likewise, Wiechmann (2008:283) argues that “there is still a strong need for empirical evaluations of competing measures of collocativity”. The final experiment of this thesis, namely Experiment 4, therefore investigates the psychological validity of different measures of collocation strength (i.e. association measures), including those introduced in section 1.3. The aims of Experiment 4 are as follows:

Aim 1: To replicate Experiment 2 in order to strengthen the confidence of its conclusions.

Aim 2: To investigate the strength of the correlation between the transition probability of a bigram and the amplitude of the ERP response.

Aim 3: To find out which measure of collocation strength most closely correlates with the amplitude of the ERP response, and thus may be seen as having the most psychological validity.

Based on the results of Experiment 2, I make the following hypotheses for Experiment 4:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will *not* elicit a P600.

Hypothesis 3: The onset latency will be greater for the collocational condition compared to the non-collocational condition.

Hypothesis 4: There *is* a correlation between the transition probability of a bigram and the amplitude of the ERP response.

The aims outlined in this section will be discussed at length in the relevant chapters.

A final point to note in relation to the aims of this thesis is that I wanted the stimuli to represent authentic language data that the participants are likely to have encountered in real life. It was important for the collocations to be corpus-derived because, if I were to invent linguistic examples based on intuition, the data would be less likely to match up with the participants' language experience. As a result, the participants might not process the experimental stimuli in the way that they process language in the real world. Therefore, using authentic corpus-derived stimuli allowed me to investigate how real-life language data is processed in the brain (insofar as this is possible in the context of an ERP experiment[2]).

1.6 Outline of the thesis

In Chapter 2 of this thesis, I provide a review of the theoretical and practical issues in the study of collocation. In Chapter 3 I provide a review of prior work on the processing of collocations, first from self-paced and eye-tracking experiments (sections 3.2 and 3.3), and then from electrophysiological studies (section 3.4). Then, in Chapter 4, I describe and justify the methodological steps and decisions that are central to all of the ERP studies presented in this thesis. In Chapter 5 through to 8, I dedicate one chapter to each of the ERP experiments that I conducted. Finally, in Chapter 9, I conclude the thesis with a summary and discussion of the results from all four experiments, along with suggestions for future research.

---

[2] The reading that takes place during an ERP experiment cannot be fully naturalistic because of the need to present sentences word-by-word (see Chapter 4, section 4.3).

CHAPTER 2: A REVIEW OF THEORETICAL AND PRACTICAL ISSUES IN THE STUDY OF COLLOCATION

2.1 Overview of the chapter

The aim of this chapter is to review the different ways in which researchers discuss the notion of collocation, and to review the theoretical and practical issues associated with each approach. In section 2.2 I define *collocation* and related terminology. In section 2.3 I provide an overview of the different ways of conceptualising collocation, and in subsequent sections I go into more detail about the different approaches. Specifically, in section 2.4 I focus on the neo-Firthian approach to collocation, looking at Sinclair's approach (section 2.4.1), Hoey's (2005) theory of Lexical Priming (section 2.4.2), and Hunston and Francis' (2000) theory of Pattern Grammar (section 2.4.3). In section 2.5 I discuss the syntax-based approach to collocation, in section 2.6 I discuss Wray's notion of *formulaicity*, and in section 2.7 I discuss the usage-based approach. In section 2.8 I provide an overview of how the different approaches represent collocation at a psychological level, and in section 2.9 I outline the main parameters of difference between the different approaches. Finally, in section 2.10 I provide a conclusion to the chapter, including a summary of the main points.

2.2 Definitions

Different approaches to collocation conceptualise and operationalize the phenomenon in different ways. These different approaches are addressed in subsequent sections. However, it is necessary for the purpose of this discussion to set out some simplified definitions. In the broadest sense, a collocation can be defined as a "co-occurrence relation between two words" (McEnery & Hardie 2012:240). McEnery and Hardie (2012:240) state that "[w]ords are said to *collocate* with one another if one is more likely to occur in the presence of the other than elsewhere".

In corpus linguistics, potential collocates are typically identified by conducting statistical significance tests on words which appear within a *span* of 4 words to the left or 4 words to the right of the *node*, where the node is the word being studied (Sinclair et al. 2004:13, 42). In this approach, pairs of regularly co-occurring words are considered to form collocations, even if they are not adjacent and even if they do not always occur in the same order (McEnery & Hardie 2012:123). A word that significantly collocates with the node at just one position within the specified span is known as a *position-dependent* collocate (Sinclair et al. 2004:35). These are typically grammatical words (Sinclair et al. 2004:83). By contrast, a word that significantly collocates with the node within a specified span but not at one particular point of the span is known as a *position-free* collocate. These are typically lexical words (Sinclair et al. 2004:83).

Sinclair (1991:115-116) distinguishes between two more types of collocation, namely *upward collocation* and *downward collocation*. An upward collocate has a higher overall text frequency than the node; by contrast, a downward collocate has a lower text frequency than the node. The upward/downward collocation distinction is not often discussed, but it is important in demonstrating the strength of collocation, as it could be the case that a word has a very low frequency in general but, when it does occur, it frequently occurs with the node. When discussing upward/downward collocations, it is important to note that, since the word that is considered to be the node changes depending on which word is being studied, the status of a collocation as either upward or downward can change depending on the direction of the analysis.

Other concepts related to the notion of collocation include *colligation*, *semantic preference*, and *semantic prosody*. Colligation can be defined as the co-occurrence of grammatical categories (Firth 1957:13) or as a “co-occurrence relationship between a word and a grammatical category or context” (McEnery & Hardie 2012:240). For example, there is a strong tendency for a possessive adjective to immediately precede the expression *true feelings* (Sinclair 2004:35).

Similarly, semantic preference is defined as a “co-occurrence pattern between a word and a semantic category of words”, while semantic prosody refers to “[t]he tendency exhibited by some words or idioms to occur consistently with either positive or negative meanings” (McEnery & Hardie 2012:250). In the paper that introduced the concept, Louw (1993:171) also defines semantic prosody in terms of positive or negative evaluations. However, Sinclair (2004:34) and Stubbs (2002:66) offer an alternative conceptualisation of semantic prosody (or *discourse prosody* in Stubbs’ terminology) by defining it as a pragmatic concept that expresses the attitude of the speaker/writer. For example, Sinclair (2004:35) states that *true feelings* has a semantic prosody of “reluctance” and a semantic preference for “expression”. These concepts of semantic preference and semantic prosody have been criticized for being difficult to distinguish from one another (McEnery & Hardie 2012:137). Indeed, even Sinclair (2004:35) notes that “in a number of cases … the semantic preference and the semantic prosody are fused”.

Collocations are usually distinguished from the related concept of an *idiom* by their relative *transparency* or *compositionality*. McEnery & Hardie (2012:244) define an idiom as a “phrase or other multi-word unit whose meaning cannot be deduced by combining the meanings of the words within it”. Since idioms have a meaning that is not reducible to the meaning of the component words, they are said to be *opaque* or *non-compositional* (Taylor 2002:100). By contrast, since the meaning of collocations can in many cases be derived from the meaning of the component words, collocations can be said to be *transparent* and *compositional* (Laufer & Waldman 2011:649). This idea that collocations are compositional is problematized in section 2.5.

2.3 Overview of different approaches to collocation

Although the definition of *collocation* provided in the previous section is seemingly straightforward, the term has actually been used and conceptualised in a variety of different ways (Nesselhauf 2004:1). Wray (2002:9) lists 57 words that have been used to refer to collocation. She notes that, although some of the words are used to refer to the same concept, they typically have different connotations and refer to different aspects of the same phenomenon (Wray 2002:8; Wray & Perkins 2000:3).

Nesselhauf (2004:1) points out that "[t]he only common point in a wide spectrum of definitions offered is that collocations are always considered some kind of syntagmatic relation of words", where the term *syntagmatic* refers to their linear order. She attributes the variation in use of the notion of collocation to the fact that it is used for different purposes by researchers in different fields. As a result, two dominant approaches to the study of collocation have emerged: namely the *proximity-based approach* and the *syntax-based approach*.

The proximity-based approach, also known as the *frequency-based approach*, the *distributional approach* (Evert 2005:15), or the *contextualist approach* (Seretan 2011:9), is typically used by corpus linguists (Millar 2010:16). This approach involves identifying collocations based on their frequency or probability. The words which form the collocation must appear in close proximity (Sinclair 1991:170), and words are considered to be collocates even if they are not adjacent and/or do not always occur in the same order (McEnery & Hardie 2012:123). My discussion in the previous section was based largely within this perspective.

By contrast, the syntax-based approach (Grefenstette 1992; Seretan 2011), also known as the *phraseological approach*, the *lexicographic approach*, or the *intensional definition* (Evert 2005), is used by researchers in the fields of lexicography, natural language processing (NLP), and language pedagogy (Evert & Kermes 2003; Nesselhauf 2005:). In this approach, collocation is defined as the "co-occurrence of two or more lexical items as realizations of structural elements within a given syntactic pattern" (Cowie 1978:132). For example, a verb and a noun that tends to occur as that verb's object would be considered to be a collocation.

Another way in which approaches to collocation vary is in terms of whether they pose one or two linguistic processing systems. Approaches that pose two processing systems posit separate lexical and grammatical systems, with collocation being a purely lexical phenomenon. By contrast, approaches that pose a single processing system explain collocation within a theory that merges grammar and lexis. An example of an approach that argues for the existence of a separate grammatical system is Wray's (2002) notion of *formulaicity*. In this approach, which is widely adopted in psycholinguistics, a *formulaic sequence* is defined in the following way:

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (Wray 2002:9).

In positing the existence of a separate grammatical system, Wray (2002:15) goes as far as to say that "a single-system model is implausible". Despite this, Wray (2002:10) argues that the *analytic system* (i.e. the grammatical system) is secondary to the *holistic system* which processes sequences as memorized chunks. Wray's notion of formulaicity is discussed further in section 2.6.

Another approach to collocation that posits the existence of a separate grammatical system is the approach taken in Sinclairian theory. Sinclair (1991:71) proposes the *Idiom Principle*, which is the idea that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments". This notion of "semi-preconstructed phrases" is directly equivalent to Wray's (2002:9) point that formulaic sequences are "prefabricated". Furthermore, like Wray, Sinclair (1991:114) suggests that this holistic system is the primary mode of interpreting text. Yet he also posits what he refers to as the *Open-Choice Principle*: an analytic system that is activated whenever an individual encounters a word that is unexpected in its environment (Sinclair 1991:109-110).

Sinclair works in a framework known as the *neo-Firthian* approach. This approach is based on the ideas of Firth (1957:12), who defines collocations as "statements of the habitual or customary places of that word in collocational order but not in any other contextual order and emphatically not in any grammatical order". Firthian linguistics focuses heavily on meaning in context (McEnery & Hardie 2012:131-132), and neo-Firthian linguistics applies this approach to corpora. Across both Firthian and neo-Firthian linguistics, there is a general scepticism of structural analysis. For Firth, this means a scepticism of Structuralism, which posits that language can be analysed at different linguistic levels (e.g. phonology, morphology, lexis, syntax); for Sinclair, this means a scepticism of Chomskyan formalism (see section 2.4.1.1).

Hoey (2005) and Hunston and Francis (2000) also work within the neo-Firthian framework. Yet, instead of positing the existence of a separate grammatical system, they attempt to account for the whole of language using collocational mechanisms (McEnery & Hardie 2012:144-145). These neo-Firthian theories are discussed in section 2.4.

Another approach that argues for the inseparability of syntax and lexis, independently of the neo-Firthian approach, is the *usage-based* approach. Researchers working within a usage-based framework typically use the work of Halliday (1961) as a starting point. In his seminal article, Halliday (1961:247, 275) states that any linguistic theory must provide a way of relating grammar and lexis. The interrelation of grammar and lexis (forming lexicogrammar) is central to the collection of usage-based theories known as Construction Grammar. Construction Grammar posits that all units of language are *constructions*, which are defined as syntactic frames that "may specify, not only syntactic, but also lexical, semantic, and pragmatic information" (Fillmore et al. 2003:243). Construction Grammar is discussed further in section 2.7.

2.4 The neo-Firthian approach

2.4.1 Sinclairian theory

2.4.1.1 The Idiom Principle

As mentioned previously, the neo-Firthian approach is based on the ideas of Firth (1957:14), who states that "[c]ollocations are actual words in habitual company". Neo-Firthian's attempt to recast Firth's ideas in the context of corpus linguistics (McEnery & Hardie 2012:122, 247). Sinclair (1991:3, 104), who is the most prominent researcher within the neo-Firthian framework, questions the traditional distinction between grammar and lexis, arguing that this distinction obscures the way in which sequences of words can have meanings that are independent of the individual words which make up the sequence. Traditional, formal grammatical theories account for this by postulating that idioms are a minor part of language that act as an exception to the usual grammatical rules (e.g. Chomsky 1965:186). However, Sinclair (1991:104, 108) argues that idioms are not just a peripheral part of language. Instead, he argues that "a substantial portion of the language" is idiomatic in the sense that "[m]ost everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up text".

With this in mind, Sinclair (1991:110) proposes the Idiom Principle, i.e. the idea that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments". Sinclair (1991:8, 1994:24) variably refers to these semi-preconstructed phrases, which are essentially fixed or semi-fixed collocations that are stored holistically, as *lexical units, lexical items,* and *extended units of meaning*. Sinclair (1991:114) states that the Idiom Principle is the default mode for interpreting text. However, when an individual encounters a word that is unexpected in its environment, the mode of interpretation can switch to what he calls the Open-Choice Principle. This is essentially the traditional way of describing language whereby any word can be chosen to occupy a slot, providing that the result is grammatical (Sinclair 1991:109-110). An example of a theory of grammar that works in this way is Transformational Generative Grammar (Chomsky 1965). In this theory, Chomsky posits formal phrase-structure rules such as S → NP VP (i.e. *a sentence consists of a noun phrase followed by a verb phrase*). Any words can be inserted into these abstract slots so that, although the sentences generated by the phrase-structure rules are grammatical, they do not necessarily have to make sense semantically (Chomsky 1957:15). Hence, the existence of Chomsky's oft-cited example *colourless green ideas sleep furiously* (1957:15).

Although Sinclair maintains a distinction between grammar and lexis/semantics by postulating two separate systems, his theory of language assigns a more central role to collocation by arguing that "[w]e would not produce normal text simply by operating the open-choice principle" (Sinclair 1991:110). It is worth nothing, however, that neo-Firthian linguists such as Hoey (2005) and Hunston and Francis (2000) try to account for language generated by the Open-Choice Principle using collocational mechanisms. In this way, they provide a unified account of language based purely on Sinclairian theory. This is discussed further in sections 2.4.2 (Hoey) and 2.4.3 (Hunston and Francis).

2.4.1.2 Problems with Sinclairian theory

The Idiom Principle has been criticized for putting too much emphasis on collocation while ignoring other prominent features of lexical semantics (McEnery & Hardie 2012:162). In particular, McEnery and Hardie (2012:163) argue that Sinclair goes too far in stating that "the idea of a word carrying meaning on its own [can] be relegated to the margins of linguistic interest" (Sinclair 1996:82). They point out that words such as *house* and *build* collocate with each other because of the real-world properties of their referents. Sinclair (2004:29) himself notes that, in some cases, “language does little more than correlate with the world”. Yet he also states that conventional units of meaning are pervasive even when accounting for factors such as lexical semantics and register (Sinclair 1991:109). When analysing language from the perspective of the Idiom Principle, care must therefore be taken to distinguish true collocations from words that co-occur due to their real-world associations.

Another problem with Sinclairian theory is the way in which Sinclair defines the notion of span. As mentioned in section 2.2, Sinclair identifies collocates using a span of 4 words either side of the node (Jones & Sinclair 1974:21). However, it is problematic to identify collocates based on individual words around the node when Sinclair has already posited that it is “multi-word patterns” (1991:104), as opposed to individual words, that carry meaning (Stubbs 2002:29). As Stubbs (2002:29) points out, this is a problem that has not yet been solved in neo-Firthian linguistics.

A related problem is the assumption that all words have the same span size regardless of factors such as word class and semantic content. Mason (2000) proposes the notion of *lexical gravity* to account for the way in which different words have varying levels of influence over the surrounding context. The lexical gravity of a node is calculated by measuring the variability at each span position (up to a maximum of about 10 words) using a metric such as entropy or type-token ratio (Mason 2000:270-272). The assumption is that, if the surrounding words have a low level of variability, meaning that not many different words or classes of words can occur in that position, then that node word must have a high level of influence, or lexical gravity, over its lexical environment (Mason 2000:270).

However, I would argue that, since the lexical gravity of a node is calculated only up to a pre-determined span of 10 words (Mason 2000:270), this is not that different from the traditional eight-token span proposed by Sinclair et al. (2004:13, 42). Jones and Sinclair (1974:21) claim that the eight-token span accounts for 95% of the collocational influence of the node. When lexical gravity measures predict a window span of 8, this would give the same accuracy percentage. Nevertheless, it could be the case that the outcome is more accurate for node words which have a small span according to lexical gravity measures. It could be argued, therefore, that the eight-token span size proposed by Sinclair might not be the most appropriate span size for all words, particularly those with a very low level of influence over the surrounding context.

Another problem with Sinclairian theory is the focus on linearity. Sinclair and other neo-Firthians believe that it is essential to take a linear approach to the study of language. Indeed, Sinclair and Mauranen (2006:xxviii) even propose a grammar entitled Linear Unit Grammar, which is characterised by the "maintenance of linearity in the description wherever possible". It is intuitively plausible to take a linear approach to the study of language because, as Brazil (1995:4) points out, "[s]peech is an activity that takes place in time: speakers necessarily say one word, follow it with another and then with another, and so on". Brazil (1995:229) goes on to state that "no mechanisms which require non-linear explanations are necessary". However, this linear approach deserves critique as, while the sound stream is linear, language production cannot be wholly linear, due to the way in which a word can precede the word which it is modifying. For instance, adjectives modify nouns, and adverbs modify verbs, but the noun or verb must be selected before the adjective or adverb in phrases such as *the tall woman* or *to slowly walk*. This suggests that language production involves creating some sort of mental hierarchy.

Despite this apparent focus on linearity, it can be argued that the neo-Firthian approach is not as linear as it is often presented. This is because, since words are considered to form collocations even if they are not adjacent, the intervening words between the node and collocate will have their own collocates embedded within, and overlapping with, the node and collocate under investigation. This non-linear feature is also apparent in other neo-Firthian theories such as Lexical Priming, where Hoey (2005:8) discusses the concept of "nesting", and Pattern Grammar, where patterns can flow into each other and have other patterns embedded within them (Hunston and Francis 2000:230). This will be discussed in the following sub-sections.

2.4.2 Lexical Priming

In cognitive psychology and psycholinguistics, the word *priming* is used to refer to the facilitated mental activation of a concept following exposure to a stimulus that is somehow related. For example, in Meyer and Schvaneveldt's (1971) seminal article on what later came to be known as *semantic priming*, participants are presented with two strings of letters and are asked to decide whether or not both letter strings are real English words. In the condition where the letter strings were both real English words, half are semantically related (e.g. *nurse-doctor*) and half are not semantically related (e.g. *nurse-butter*). Meyer and Schvaneveldt (1971) find that participants respond significantly more quickly when the words are semantically related than when they are not semantically related. This shows that a word can *prime* (i.e. mentally activate) a word from the same semantic category.

The psycholinguistic notion of priming is *paradigmatic* (i.e. non-linear), as the *prime* (i.e. the activating stimulus) and the *target* (i.e. the word being mentally activated by the prime) exist in a non-linear relationship whereby one can be substituted for the other. However, in Hoey's (2005) theory of Lexical Priming, the term *priming* is used in an entirely syntagmatic sense, where the prime and the target exist in a linear relationship. This distinction between syntagmatic and paradigmatic relations was introduced by de Saussure (1916:123), though de Saussure used the term *associative* instead of paradigmatic. The syntagmatic relationship between the prime and the target is evident in Lexical Priming, as Hoey (2005:8) posits that any word that frequently follows another word is mentally activated whenever the first word is spoken or heard. However, Hoey (2005) does not explicitly state that he shifts the notion of priming from the paradigmatic to the syntagmatic plane, and the title of the theory – Lexical Priming – does not provide any indication of this shift.

Hoey's use of the term *priming* is also different from the way in which it is used in psychology. In psychology, the term is used to refer to the mental activation itself. By contrast, in Hoey's theory of Lexical Priming, the term is used to refer to the links between words which allow the mental activation to occur. Furthermore, in psychology, lexical priming already exists as a term but it does not refer to the concept that is proposed in Hoey's theory. Instead, psychologists seem to use the term lexical priming synonymously with the notion of semantic priming mentioned earlier (e.g. Spence & Owens 1990:318), or as a hypernym for different types of word-related priming including semantic and associative priming (Jones & Estes 2012:45). These discrepancies between the psychological notion of priming and Hoey's theory of Lexical Priming are not explicitly addressed in Hoey's work.

Furthermore, before Hoey introduced his notion of Lexical Priming, psychology already used the term *associative priming* for words that frequently co-occur (e.g. Moss et al. 1994:413) (also see section 3.4.6). Nevertheless, the introduction of a new term for this concept was arguably necessary for a number of reasons. First, across many psychology sources, "there seems to be a confound between semantic and associative relatedness in the materials used" (Alario et al. 2000:742). This is unsurprising because, even in the seminal article on semantic priming mentioned earlier (Meyer & Schvaneveldt 1971), the discussion is framed in terms of associations even though this article is widely recognized as being the first to demonstrate semantic priming. Secondly, some psychologists do not use frequent co-occurrence of words as the only or even the main criteria in defining associative priming. Instead, frequent co-occurrence of words is often seen as "an important cue for detecting word associations" (Chaudhari et al. 2010:1058) that is secondary to other cues such as phonological relatedness, e.g. rhyming word pairs (Alario et al. 2000:749). Thirdly, associative priming in psychology only focuses on nouns as the prime and the target whereas, in Hoey's theory of Lexical Priming, all words including function words are involved in priming relationships. Hoey's introduction of a new term for this phenomenon is therefore very useful, even if a term such as *syntagmatic priming* might be more descriptive and transparent for the theory that he proposes.

A key feature of the theory of Lexical Priming is that the primings are not the same for each individual; everyone acquires their own unique set of primings based on the language that they have produced, heard, and read throughout their lifetime (Hoey 2005:14, 178). Whenever words co-occur in language produced or perceived by an individual, the priming relationship between those words is reinforced for that individual. By contrast, whenever a word is combined with an unexpected word, its existing priming relationships with other words are weakened. Through this process, the primings of a word can shift throughout an individual's life, and this can lead to a shift in the meaning or function of the word for that individual. Hoey (2005:9) refers to this shift as "a drift in the priming". Since primings are unique to each individual, a "corpus cannot tell us what primings are present for any language user". This makes conducting research on Lexical Priming problematic. Nevertheless, in spite of this, Hoey states that a corpus "can serve as a kind of laboratory in which we can test for the validity of claims made about priming" and that a corpus can "indicate that certain primings are likely to be shared by a large number of speakers" (Hoey 2005:14, 15).

Another key feature of Lexical Priming is that, as well as words being syntagmatically primed by other words, words are primed by the contexts in which they occur (Hoey 2005:8). This form of priming is not syntagmatic or paradigmatic. For example, *recent research* is a frequent collocation in academic texts (Hoey 2005:9). Therefore, the word *research* is primed to occur with *recent* when academics are producing or comprehending academic language but not when they are producing or comprehending other forms of language. The idea that collocations are context-dependent was first suggested by Firth (1957:12-13), who notes that different collocations are found in different types of text.

Hoey (2005:11, 180) adds that individuals can experience a “crack” in their priming. This means that their implicit collocational knowledge is undermined by the knowledge imposed by education (e.g. teachers, dictionaries). However, this crack can be “healed” by assigning different primings to different contexts. For instance, if a teacher tells a student that their use of *you was* is ungrammatical, that student might continue to use *you was* in the home environment but use the standard *you were* in the school environment (Hoey 2005:11). Hoey (2005:8) therefore argues that a word “becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context”.

As well as applying to individual words, Lexical Priming also applies to collocations. Hoey (2005:8) explains that words which are primed to co-occur with each other form sequences that can have their own primings which do not apply to the words making up the sequence. Hoey (2005:8) refers to this as “nesting”, and argues that the concepts of priming and nesting enable us to account for the whole of the language rather than just a subset of the language, as is the case with the Idiom Principle. This is because, whereas the lexical items in the Idiom Principle have clear start and end points, the notions of priming and nesting can account for the sequencing of syllables, words, collocations, and the wider discourse structure, without the need to impose artificial boundaries on what is and is not a collocation (Hoey 2005:158).

One problem with Lexical Priming pointed out by Hoey is that the theory uses the word as the starting point of analysis despite working within the framework of Sinclair’s ideas, one of which is that the word is not necessarily the main unit of meaning. Indeed, Hoey (2005:158) admits that he focuses “on the word as a convenient starting point for the description of priming, rather than for theoretically grounded reasons". Furthermore, as well as acknowledging that units longer than the word might be a more theoretically appropriate starting point for his theory, Hoey (2005:158-159) points out that sub-word units such as syllables and phones are also primed in the same way as words. However, the theory of Lexical Priming is part of the neo-Firthian approach, which is grounded in lexicography. As McEnery and Hardie (2012:142) point out, "it is perhaps unsurprising that a school of thought with roots in lexicography should emphasise the absolute centrality of the word – as opposed to some other unit of linguistic description, such as the phoneme, morpheme, phrase or clause – to the understanding of language".

2.4.3 Pattern Grammar

Another neo-Firthian theory which attempts to account for the whole of language rather than just a subset of language is Pattern Grammar (Hunston & Francis 2000:14). In this theory, it is argued that syntax and lexis are inseparable in the sense that "[p]articular syntactic structures tend to co-occur with particular lexical items, and – the other side of the coin – lexical items seem to occur in a limited range of structures" (Francis 1993:143). Syntax and lexis cannot be discussed independently of each other. Therefore, Hunston and Francis (2000) propose the notion of a *pattern* which encapsulates both syntax and lexis. Specifically, the patterns of a word are defined as "all the words and structures which are regularly associated with the word and which contribute to its meaning" (Hunston & Francis 2000:37).

Francis (1993:141) describes this notion of patterning as a "blend of colligation and collocation", as it involves identifying both the words and the grammatical structures associated with a particular word. Grammatical categories such as noun and verb are used in the definition and the notation of patterns. For instance, the notation **V n** is used to refer to the pattern of a verb or "verb group" (i.e. a main verb and its auxiliary verbs) followed by a noun or "noun group" (Hunston & Francis 2000:45). The word class that is being focused on by the researcher is written in an upper-case letter, and any lexical items that are consistently part of the pattern are written in italics. For example, the notation ***to* N** is used to refer a pattern where a noun is preceded by the preposition *to* (as in the sentence *They went* <u>to school</u> *together every day*) and the noun is the element being focused on by the researcher (Hunston & Francis 2000:57).

A key feature of Pattern Grammar is the notion of *prospection*. Hunston and Francis (2000:241) state that "[t]he term 'prospection' is used to indicate that something that occurs in a discourse leads the reader or hearer to expect that some other thing will occur". This notion of prospection is similar to the linear approach to the study of language mentioned in section 2.4.1.2, but it originates from Sinclair's (1992) theory of discourse structure. Sinclair (2004:88) states that "[p]rospection occurs where the phrasing of a sentence leads the addressee to expect something specific in the next sentence". Prospection contributes to the coherence of discourse if the prospections are consistently fulfilled (Mauranen 1993:110; Sinclair 2004:97). However, in Hunston and Francis' (2000:241) reformulation of prospection, the term is used to refer to the collocational/colligational patterns that are central to Pattern Grammar rather than the wider discourse patterns discussed by Sinclair.

Hunston and Francis (2000:208) state that a particular word 'prospects' a particular pattern, and any of the words within that pattern can 'prospect' another pattern. In this way, patterns overlap and flow into each other. Hunston and Francis (2000:211-212) refer to this phenomenon as *pattern flow*. In addition to pattern flow, three other pattern configurations have been proposed by Hunston and Francis (2000:215, 224, 229, 230), namely "pattern string", "pattern loop", and "pattern accumulation". *Pattern string* refers to adjacent patterns that do not overlap and flow into each other, *pattern loop* refers to patterns which are complete but have another pattern embedded within them, and *pattern accumulation* refers to the situation where a pattern is prospected by more than one word.

These different pattern configurations, particularly pattern flow, demonstrate how discourse might be constructed without assuming that there always has to be a clear beginning and end point to a collocation. However, one problem with Pattern Grammar that is pointed out by Hunston and Francis is that it is difficult to state with any certainty that some sequences are analysable in terms of patterns while other sequences are not. Hunston and Francis (2000:49) state that "the question of what is and is not a pattern is one that is not always easy to answer", and that "[c]ertain elements are excluded on the grounds that they can occur with almost any word of the same class". However, if there are word sequences that cannot be analysed in terms of patterns, this implies that Pattern Grammar is not and cannot be a complete explanation of the grammar of a language.

Furthermore, Hunston and Francis (2000:14) claim that Pattern Grammar attempts "to describe the whole of the language (or rather, all the frequently-occurring items in the language)". Yet the most frequently-occurring items are precisely the words that are likely to be able to combine with "almost any word of the same class". For instance, the word *the* is the most frequent word in any corpus of written English, yet this word can be combined with almost any noun. This implies that Pattern Grammar cannot actually account for the most frequent items in a language.

Another problem with Pattern Grammar that is noted by Hunston and Francis (2000:107) is that, while the "core" words which occur with a particular pattern are usually clearly identifiable, it is "extremely difficult" to list all of the words which occur with the pattern but do so with lower frequency. We may suspect that it is only the lexical focus of Sinclairian theory that makes Hunston and Francis believe that it ought to be possible to itemise all of the words that can occur with each pattern, in spite of the extreme difficulty they themselves identify in doing this. Indeed, it could actually be the case that it is not merely extremely difficult, but actually impossible to compile a full list of words which occur with a particular pattern. After all, outside of neo-Firthian theory (and to some extent within, in the case of Sinclair's (1991:109-110) Open-Choice Principle), there is a general belief originating from Chomsky's (1965) Transformational Generative Grammar that any word can occupy a slot in a grammatical structure as long as the resulting sentence is grammatical. Hunston and Francis (2000:3) do not hold this assumption, arguing instead that patterns fundamentally belong to words and that the patterns do not exist as independent abstract structures. However, if a pattern exists which does not have a closed set of words that it occurs with, then clearly that pattern *does* have an independent existence. Therefore, Hunston and Francis' (2000:107) point that enumerating any pattern's complete list of possible words is extremely difficult suggests that they might actually be wrong in claiming that patterns do not exist as independent structures, and they are thereby inadvertently supporting the opposite assumption. This is comparable with Lexical Priming because, by proposing the concept of "nesting" mentioned in the previous section, Hoey (2005:8) accepts that larger units do in fact have an independent existence in the system.

Hunston and Francis (2002:100) go further in inadvertently supporting the non-neo-Firthian assumption by positing that, occasionally, it is the pattern itself that has the meaning. For example, when the pattern **V *way* prep/adv** is used with a verb that denotes talking (e.g. *talked his way into*), Hunston and Francis (2000:100) state that "the meaning of the whole phrase is that someone uses clever, devious, or forceful language to achieve a goal, usually extricating themselves from a difficult situation, or getting into a desirable situation". This meaning is not related to any of the individual words that make up the pattern, suggesting that patterns themselves can carry meaning. In this way, patterns become comparable to constructions in Construction Grammar. Construction Grammar will be discussed in section 2.7.

2.5 The syntax-based approach

Early work in the syntax-based approach was carried out by Cowie (1978:132), who defines collocation as the "co-occurrence of two or more lexical items as realizations of structural elements within a given syntactic pattern". Cowie divides phraseological units into two different types, namely *composites* and *formulae*. In this context, *formulae* refers to phraseological units with a pragmatic function such as *how are you?* (Nesselhauf 2004:10), whereas *composites* are a class of four different types of phraseological unit, namely pure idioms, figurative idioms, restricted collocations, and open collocations. These different types of phraseological unit are conceptualised along a scale of compositionality. Pure idioms such as *kick the bucket* lie at one extreme of this scale, as they have a fixed structure and a meaning that is not reducible to the meaning of the component words. By contrast, open collocations such as *drink tea* fall at the other extreme of the scale, as their structure is not fixed and the meaning can in many cases be derived from knowing the meaning of the component words. Figurative idioms such as *to catch fire* and restricted collocations such as *make a comment* lie in between these two extremes.

The idea that collocations can be compositional is problematized in the Sinclairian approach, which posits that collocations are, effectively, non-compositional word sequences or word combinations (in the case of co-occurrences in a non-fixed order). For instance, the phrase *traffic light* would not typically be considered to be idiomatic because the meaning is fairly transparent. However, the meaning is not entirely transparent because *traffic light* designates a specific type of light that is associated with traffic (Baldwin & Kim 2010:273). Specifically, a *traffic light* is a light that informs drivers when they need to stop or start driving in order to control the flow of traffic. Yet there are other lights that are associated with traffic such as the lights at the side of the road or the lights that are attached to vehicles. In Sinclair's view, "[m]ost everyday words do not have an independent meaning (1991:108). Therefore, if a word supposedly does not have much intrinsic meaning, collocations cannot be seen as being compositional, because their meaning cannot be reduced to the meanings of the component words. However, as mentioned in section 2.4.1.2, this perspective is problematic because there are clearly aspects of lexical semantics other than collocation which are important to the meaning of a word (McEnery & Hardie 2012:162).

While the frequency-based approach identifies collocations "more or less independently of grammatical pattern or positional relationship" (Sinclair 1991:17), the syntax-based approach considers grammatical patterns to be a "central defining feature" in the identification of collocations (Seretan 2011:12). As Nesselhauf (2005:25) points out, in the syntax-based approach, "collocations are considered a type of word combination in a certain grammatical pattern". For example, a verb and a noun that tends to occur as that verb's object would be considered to be a collocation. In more complex structures such as the ditransitive, every verb relation within the structure (i.e. the verb and its subject, the verb and its direct object, and the verb and its indirect object) is analysed separately for the purposes of syntactic collocation. According to Grefenstette (1992:90), the syntax-based approach therefore "opens up a much wider range of contexts" than the frequency-based approach.

However, the syntax-based approach to collocation is not actually as different from the proximity-based approach as it initially seems to be. This is because the eight-token span used to identify collocates in the proximity-based approach is likely to capture the grammatical patterns that the node is a part of. For example, the eight-token span is likely to capture the ditransitivity of the verb *give* because this span is long enough to contain a subject, a direct object, and an indirect object. Similarly, for words to be syntactically related, they must occur within the same phrase, clause, or sentence, and therefore probably in close proximity to each other. Proximity limitations therefore implicitly exist within the syntax-based approach (Seretan 2011:13). Thus, in this way, proximity can be considered to be a proxy for syntax.

However, in the proximity-based approach, the syntactic relations are not analysed separately in the way that they are in the syntax-based approach.

Some collocation researchers combine aspects of the syntax-based and proximity-based approaches. For instance, Greenbaum (1974:80, 82) defines a collocation as “a frequent co-occurrence of two lexical items” and refers to the neo-Firthian notion of collocational span. Yet Greenbaum also acknowledges that the identification of collocations “require[s] syntactic information in at least some instances”. Furthermore, Greenbaum (1970:11) states that it is not possible to confirm that two words constitute a collocation without considering their grammatical relationship. Mitchell (1971:54) uses a similar combined approach to the study of collocation, despite working independently from Greenbaum. This is because, like Firth, he emphasizes the “habitualness” of collocations yet, like Cowie, he distinguishes between collocations and other phraseological units such as idioms. Even Nesselhauf (2004:34), who is a central figure in the syntax-based approach, uses frequency as a criterion, as evidenced in the following quotation:

> For a combination to be considered to exist, it is not sufficient that it has been used by a (native) user of the language at some point, as this would mean that practically every conceivable combination could be considered to exist. Instead, it will be taken to mean that a combination is either used with a certain frequency and/or is usually considered an acceptable combination in English by adult native users of a standard variety of British or American English (Nesselhauf 2004:34-35).

Indeed, Nesselhauf (2004:18) argues that the proximity-based approach and the syntax-based approach to collocations are equally important “as they focus on different important aspects of syntagmatic relations”.

2.6 Formulaicity

Wray (2002:3-4) points out that speakers often produce sequences of words without analysing them according to their component parts. Idioms, by definition, cannot be understood by combining the meanings of their component words (McEnery & Hardie 2012:244). Theories of language such as Chomskyan formalism assume that idioms, in this strict sense, are stored as holistic chunks in the lexicon (e.g. Chomsky 1965:186). However, Wray (2002:4) argues that sequences which are not typically considered to be idioms are also analysed as non-compositional sequences that are stored holistically in the lexicon. Wray (2002:4) considers this formulaicity to be pervasive throughout language, and proposes the term *formulaic sequence*. She defines a formulaic sequence as:

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (Wray 2002:9).

Wray (2002:8) proposes this term out of dissatisfaction with other terms that have previously been used to describe aspects of formulaicity, “all of which have something useful to say, but none of which seems fully to capture the essence of the wider whole”. However, Wray’s notion of formulaicity is similarly restricted in the sense that it focuses on sequences of words that are completely fixed.

Wray (2002:14) posits two separate processing systems, namely the analytic system and the holistic system. Wray (2002:14) states that, while “[a]nalytic processing entails the interaction of words and morphemes with grammatical rules”, “[h]olistic processing relies on prefabricated strings stored in memory”. Drawing upon Pawley and Syder (1983:218), Wray (2002:18) argues that the advantage of the holistic system is that it minimizes processing effort. By contrast, drawing upon Bloom (1973:17), Wray argues that the advantage of the analytic system is that it is can process and produce novel utterances that have never been heard before.

Wray (2002:263) proposes a model entitled the *Heteromorphic Distributed Lexicon*. In this model, holistic units are classified into one of five different lexicons depending on their function. There is a "grammatical" lexicon which stores grammatical units such as *in order to*, a "referential" lexicon storing units such as *dog*, and an "interactional" lexicon storing units such as *great to see you*. There is also a separate lexicon for "memorized" material such as songs and phone numbers, and a separate lexicon for "reflexive" units such as swear words. In each lexicon, there are three types of unit that can be stored holistically, namely the "morpheme", the "formulaic word", and the "formulaic word string" (Wray 2002:262). For example, in the "grammatical" lexicon, there are morphemes such as *–ly* and *–able*, formulaic words such as *because*, and formulaic word strings such as *in order to*.

One problem with the Heteromorphic Distributed Lexicon model is that it is not entirely clear how Wray decides which of the five lexicons a particular formulaic unit should be stored in. Some of the examples given by Wray are straightforward. For example, formulaic units such as *Hey!* and *great to see you* clearly fit in the "interactional" lexicon. However, it is not entirely clear why she places the sequence *the most important thing is* in the "interactional" lexicon. Furthermore, the fact that Wray conceptualises formulaic sequences as having clear beginning and end points is problematic. For example, Wray gives no justification for why she posits *the most important thing is* as a formulaic sequence instead of shorter versions such as *the most important thing*, longer versions containing a noun after *is*, or *the most important thing is* without a modifier. In addition, it is also questionable whether positing five separate lexicons is a psychologically plausible model of holistic storage.

Wray (2002:12) states that the "capacity for handling novelty, both ideational and grammatical, is sufficient to rule out the possibility that language knowledge consists only of a set of prefabricated phrases and sentences memorized from previous encounters with them". However, in Wray's later work, she becomes more open to the possibility that positing a separate analytic system might be unnecessary. For instance, instead of drawing a distinction between formulaic and non-formulaic sequences, Wray (2012:245) suggests that "perhaps everything we say is formulaic at one level or another". This brings her position closer to the stance taken in theories such as Lexical Priming (section 2.4.6) and the usage-based approach (section 2.7), and indeed the approach taken in this thesis. Despite this, Wray (2012:245) believes that accounting for the whole of language in terms of formulaicity results in "a flattening effect". For instance, she states that "[w]hat attracted our attention to formulaic language was that certain wordstrings stood out as more formulaic than others, even within the subset of those that are idiomatic" (Wray 2012:245).

2.7 The usage-based approach

The basic idea behind the usage-based approach is that "language structure emerges from language use" (Tomasello 2003:5, 105). This means that, over historical time, words that are consistently juxtaposed in meaningful patterns become fixed grammatical structures through the cultural process of *grammaticalization* (Tomasello & Bates 2001:8; Tomasello 2003:1). This is recapitulated in child language acquisition, as children begin to juxtapose concrete words in meaningful combinations before they are able to generalise and abstract across the resulting linguistic patterns (Tomasello 2001:169; Tomasello 2003:5). Importantly, whereas Wray (2002:113) argues that children make use of both analytic and holistic processing, Tomasello and Bates (2001:8) argue that children initially process *all* sequences of words holistically and only later break up the units into their component parts. Meaning is therefore central to the usage-based approach, with grammar being derived from meaning.

One collection of theories that take a usage-based approach is Construction Grammar. There are many different versions of Construction Grammar, each with a different focus. For example, Radical Construction Grammar takes a typological perspective (Croft 2001). However, all versions of Construction Grammar share the view that "there is a uniform representation of all grammatical knowledge in the speaker's mind" (Croft & Cruse 2004:255). Specifically, all units of language are considered to be *constructions*. Constructions are defined as syntactic frames that "may specify, not only syntactic, but also lexical, semantic, and pragmatic information" (Fillmore et al. 2003:243). In this way, constructions symbolically link form with meaning (Croft 2007:490). This is what distinguishes Construction Grammar from formal theories of grammar, where grammatical rules are meaningless and only words have meaning.

While some constructions are syntactically complex, such as the passive construction, simpler units of language such as morphemes and individual words can also constitute constructions (Goldberg 1995:4; Fillmore et al. 2003:243; Tomasello 2003:6). This is because, like the more complex constructions, individual words and morphemes link form with meaning. It was de Saussure (1959:67) who first suggested that individual words are signs that constitute a combination of form (or, in de Saussure's terminology, the "signifier") and meaning (the "signified"). Therefore, since Construction Grammar posits a symbolic link between form and meaning in both individual words and more syntactically complex units, Construction Grammar can be considered an extension of de Saussure's idea.

As well as varying along the dimension of complexity, constructions can also vary along the dimension of abstractness. For instance, some constructions are totally fixed idioms that have a meaning that is not reducible to their component parts, e.g. *kick the bucket* (Tomasello 2003:101; Fillmore et al. 2003:243). By contrast, some constructions are highly schematic and allow a variety of different words to occupy the empty slots (Croft & Cruse 2004:256). For example, the regular English plural construction and the English passive construction are both highly abstract (Tomasello 2003:101). In this way, Construction Grammar posits a *syntax-lexicon continuum* along the dimensions of complexity and abstractness, with collocations lying at “[a]n intermediate point on this continuum” (Croft & Cruse 2004:256).

In Construction Grammar, and in cognitive linguistics more broadly, “collocations can be viewed as particular instances of constructions” (Torres Cacoullous & Walker 2011:236). For example, the collocation *the bigger, the better* instantiates the construction *the X-er, the Y-er* (Croft & Cruse 2004:234). This link between constructions and collocations was pointed out by Bybee (2010:28), who states that collocations are “conventionalized instances or exemplars of constructions that are not unpredictable in meaning or form ... but are known to speakers as expressions they have experienced before”. For instance, Bybee (2010:28) points out that the idiom *pull strings* instantiates the verb-object construction and the collocation *dark night* instantiates the adjective-noun construction. Both the constructions and the instantiations of constructions leave memory traces in the mind (Bybee 2010:28).

Stefanowitsch and Gries (2003) combine the notion of constructions with the notion of collocations by proposing a new concept termed *collostruction*. A collostruction can be defined as a co-occurrence relationship between a grammatical construction and a word which occurs more frequently than expected in a particular slot of the construction (Stefanowitsch & Gries 2003:214-215). Collostructional analysis involves first choosing a construction to focus on, and searching for this construction in a corpus in order to identify the words which are strongly attracted to (or repelled by) the open slots in the construction being investigated (Stefanowitsch & Gries 2003:214). However, this is methodologically problematic because it is difficult to search for constructions in a corpus even if the construction is partly fixed (McEnery & Hardie 2012:181). The identification of collostructions in a corpus therefore involves at least some manual analysis, or the use of a manually-parsed corpus or one whose parsing has been manually corrected. By contrast, the identification of proximity-based collocations (see section 2.3) in a corpus can be fully automated.

Collostructional analysis is similar to the syntax-based approach to collocation (section 2.5) in that both of these approaches involve identifying grammatical structures. However, while collostructional analysis is strongly grounded in the theory of Construction Grammar (Stefanowitsch & Gries 2003:209), syntax-based collocation does not necessarily have this theoretical association. Similarly, collostructional analysis is very similar to Pattern Grammar (section 2.4) as they have converging concepts and results (McEnery & Hardie 2012:212). However, while collostructional analysis begins with the construction and then identifies co-occurring words, Pattern Grammar begins with the individual word and then identifies the grammatical pattern that the word is found in.

Some traditional neo-Firthian linguists such as Louw (2010) and Teubert (2005, 2010) argue against the use of corpus linguistics as a tool for gathering linguistic examples in the study of Construction Grammar, or cognitive linguistics more generally. This is for two distinct but interrelated reasons: (1) linguists who take this perspective see the corpus as a theory, rather than a method (or, more specifically, corpus linguistics is seen "as a theoretical approach to the study of language" characterised by "an insistence on working only with real language data taken from the discourse in a principled way and compiled into a corpus" (Teubert 2005:2, 4)), and (2) linguists who take this perspective argue that the corpus should be studied independently of any other predefined theoretical positions.

For Teubert (2005, 2010), corpus linguistics is not just a method which can be utilised by a linguist working with any linguistic or psychological theory; rather, the corpus *is* the theory in the sense that any conclusions about language that are reached using corpus data must come from the data itself, as opposed to fitting the corpus data into a pre-established theory. In other words, to use the terminology adopted by Tognini-Bonelli (2001:65), Teubert (2005, 2010) advocates for a corpus-*driven* approach rather than a corpus-*based* approach to the study of language.

One consequence of the corpus-driven approach is that it devalues corpus annotation[3], as corpus annotation "presuppose[s] categories not validated by corpus evidence" – even seemingly uncontroversial ideas about language such as what constitutes a noun or a verb (Teubert 2010:355). Another consequence of the corpus-driven approach is the supposition that corpora cannot be used to tell us anything about how language is represented in the mind, as this would involve studying corpus data in relation to pre-established psychological theories. Indeed, Teubert (2010:357) states that meaning "is irreducible to hypothesised mental representations and equally to neurons firing in our brains", arguing instead that meaning is socially constructed through discourse, and that meaning exists within the discourse (or within a sample of discourse, as captured in a corpus) rather than in the mind.

Construction Grammar falls within a cognitive framework that attempts to hypothesize about the mental representation of linguistic forms. Teubert (2005:8) asserts that "corpus linguistics and cognitive linguistics are two complementary, but ultimately irreconcilable paradigms". Similarly, Louw (2010:346) argues that any claims that are made about the mind on the basis of corpus evidence "can never be proved". Thus, since some neo-Firthians consider corpus linguistics to be valid but cognitive theories to be invalid, applying a valid method to an invalid theory will never produce valid results for those who take that perspective.

Looking beyond Construction Grammar, some traditional neo-Firthian linguists argue against the use of statistics in the identification of meaning. For instance, Teubert (2010:357) claims that "collocation, and certainly not statistics, is at the very heart of meaning"; Louw (2010:347) even goes as far as to state that linguists who use statistics "may be motivated by a desire to look modern and scientific", rather than a desire to find "the truth". This is now

---

[3] Corpus annotation refers to the linguistic information that has been added to a corpus, or the process of adding this linguistic information (McEnery & Wilson 2004:32; Hyland 2015:301) – see Chapter 1 (section 1.2).

somewhat ironic, given the prevalence of the quantitative (i.e. collocation via statistics) as opposed to the qualitative approach (i.e. collocation via concordance) to the identification of collocations (see Chapter 1, section 1.2). Since this thesis adopts the quantitative approach to the identification of collocations, and focuses on the psychological validity of collocation, this thesis would be of no interest to traditional neo-Firthian corpus linguists such as Louw and Teubert. Indeed, when commenting on the psychological validity of corpus-based observations, Teubert (2010:357) claims that corpus linguists “neither know nor care about these realities”.

2.8 How the different approaches represent collocation at a psychological level: Introducing the network model of language processing

Across the different approaches to collocation, there is either an explicit or an implicit assumption that collocations are represented in the brain as transitions across a network. The network consists of nodes which, depending on the theory, constitute individual words, collocations, or constructions. These nodes are connected to other nodes via weighted connections. When an individual produces or hears word X immediately followed by word Y, a connection is formed between those two words (or nodes). Then, on subsequent occasions when that individual produces or hears word X, word Y will be mentally activated, along with any other words that have previously followed word X for that individual. Mental activation is defined as a “state of memory traces that determines both the speed and the probability of access to a memory trace” (Anderson 2005:455). Through repeated exposure to the same sequence X-then-Y, the connection between word X and word Y is strengthened so that there is an increased probability that word Y will occur after word X. In this way, the weightings between nodes are essentially a type of transitional probability whereby the probability of a particular word occurring after the preceding word depends upon the strength of the connection between those words. This connection strength is determined by the level of prior exposure to the collocation.

This network model of language processing is in line with what is known about how *neurons* work in the brain. Neurons are the basic signalling units in the brain that can be understood in terms of their membrane potential, i.e. “the difference in electrical charge between the inside and outside of a cell” (Pinel 2014:81). At rest, neurons have a negative membrane potential of roughly -70 millivolts (mV). In other words, there is a lower voltage inside the neuron than outside the neuron. When a neuron receives neurotransmitters (i.e. electrochemical signals) from other neurons, its resting potential of -70 mV can increase to -72 mV (Pinel 2014:83). This is known as *hyperpolarization*, and it decreases the probability that the neuron will fire, i.e. transmit a signal to other neurons. Alternatively, the resting potential of the neuron can decrease to -67 mV. This is known as *depolarization*, and it increases the probability that the neuron will fire. If the resting potential of the neuron decreases to roughly -65 mV (i.e. its *threshold of excitation*), an *action potential* is generated, i.e. a momentary change in a neuron’s membrane potential to roughly +50 mV (Pinel 2014:557). This action potential causes neurotransmitters to be released to other neurons.

Through the transmission of electrochemical signals, neurons interact and form networks that are capable of representing knowledge (Rosenweig et al. 1996:35; Sternberg 2009:35). This network idea was first proposed by Hebb (1949), who posited that neurons are organized into networks by sensory input and that these networks are the basis of complex cognitive behaviour. These networks of neurons must constitute the ultimate basis of language knowledge. Therefore, it is intuitively plausible to propose a model of language that uses a representation similar to what we know about the ultimate neural substrate. Note that, if we assume that each node in the language network consists of a word (or a construction), each node contains much more information than could be represented in a single neuron. Therefore, neurologically, the nodes themselves would be composed of some sort of network representation of the word or construction that it represents.

The idea that collocations are represented in the brain as transitions across a network is most explicit in Hoey's (2005) theory of Lexical Priming. Hoey (2005:7, 8) explicitly links the notion of collocation to psychology, arguing that "collocation is fundamentally a psychological concept", and argues that any word that frequently follows another word is mentally activated whenever the first word is spoken or heard. Furthermore, Hoey (2005:8) states that "[a]s a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered". In other words, through repeated exposure to the same words in collocational patterns, the connection strength between the nodes representing those words is strengthened.

The network model of language processing is also evident in Halliday's (1961) paper. Halliday (1961:276) explicitly discusses collocation in terms of probabilities, stating that:

> Collocation is the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur, at n removes (a distance of n lexical items) from an item x, the items a, b, c .... Any given item thus enters into a range of collocations, the items with which it is collocated being ranged from more to less probable.

Similarly, the network model is very explicit in Construction Grammar. Constructions exist in a network in which, each time a construction is heard or produced, a pattern of nodes is activated. If the nodes are activated frequently enough, the construction becomes entrenched in the mind as a holistically stored processing unit (Langacker 1987:59-60). A network relationship also exists between an abstract construction and an instantiation of that construction, i.e. a collocation (Torres Cacoullous & Walker 2011:236). Furthermore, the network is described as a "structured inventory of constructions" because it contains core constructions which connect to many other constructions as well as peripheral constructions with fewer connections

(Tomasello 2003:6). In this way, "[t]he collection of constructions ... constitute a highly structured lattice of inter-related information" (Goldberg 1995) – similar to the pattern of interconnected neurons in the brain.

The network idea is not explicitly discussed by Wray (2002). This is surprising, given that Wray is a psychologist. Wray (2002:270) actually states that she is not concerned with the mental representation of formulaic sequences because "it is rarely addressed in other models". However, to some extent, the network idea is there implicitly because she argues that frequently encountered sequences are accessed more quickly than sequences that are encountered less frequently (Wray 2002:268). This is relevant to the network idea because, as mentioned earlier in this section, mental activation is defined as a "state of memory traces that determines both the *speed* and the probability of access to a memory trace [emphasis added]" (Anderson 2005:455).

The network idea is also not explicitly discussed in Sinclairian theory or in the basic syntax-based approach. Indeed, Stubbs (1993:21) states that "Sinclair's work is strangely detached from any psychological or social theory". This is unsurprising, given that Teubert (2005:2-3) argues that "[c]orpus linguistics ... is not concerned with the psychological aspects of language". However, although Sinclair does not explicitly discuss the psychological basis of his theory, he does argue that a "word becomes associated with a meaning through its repeated occurrence in similar contexts". In doing so, he seems to be suggesting that the brain is sensitive to frequency information in the input, and thereby implicitly supports a key tenet of the network model of language processing.

Similarly, although the syntax-based approach is seen as lying in opposition to the frequency-based approach, Nesselhauf (2004:34), who is a central figure in the syntax-based approach, acknowledges the importance of frequency as a criterion in identifying collocations (see section 2.5). This again suggests that the mind is sensitive to frequency information in the input, and that the connections between nodes are strengthened through repeated exposure. Thus, although the Sinclairian approach and the syntax-based approach do not explicitly discuss collocations as forming a psychological network, the assumption is there implicitly. Indeed, Stubbs (1993:21) argues that Sinclair's "detailed observations about the co-selection of lexical items are ripe for input into psycholinguistic theory on language learning".

Hoey (2005:158) states that his theory of Lexical Priming "contextualises theoretically and psychologically Sinclair's insights about the lexicon". Indeed, Hoey goes much further than Sinclair in posing a psychologically plausible model of collocation as he uses the psycholinguistic notion of priming to explain how words become mentally activated. However, as mentioned in section 2.4.2, Hoey's use of the term 'priming' is different to the way in which it is traditionally used in psycholinguistics. Specifically, while psycholinguists consider priming to be a paradigmatic phenomenon (i.e. an item primes a related item in a non-linear relationship), Hoey considers priming to be a syntagmatic phenomenon (i.e. an item primes the next item in a linear sequence). Since there is already psycholinguistic evidence in support of paradigmatic priming (see section 2.4.2), a psychologically plausible network-based model of language knowledge needs to account for both paradigmatic and syntagmatic relations. In this way, Hoey's theory of Lexical Priming does not fully account for how language works in the brain.

A strength of Hoey's theory is that he attempts to link language processing to wider cognition. Hoey (2005:163) posits a two-way priming process whereby words prime thoughts and feelings but thoughts and feelings also prime words. For example, he states that "[i]f *sorry* is primed in different combinations to occur as part of an expression of sympathy or as an apology, so also feelings of sympathy and apology must be primed to elicit *sorry*". This is actually similar to Sinclair's (2004) formulation of semantic prosody. As mentioned in section 2.2, Sinclair (2004:34) defines semantic prosody as a pragmatic concept that expresses the attitude of the speaker/writer. This suggests that encountering a word automatically invokes the attitudes that are associated with that word. Thus, by framing Sinclair's formulation of semantic prosody in terms of priming relationships, Hoey (2005:158) is further contextualising "theoretically and psychologically Sinclair's insights about the lexicon". In addition, by linking words to the thoughts and feelings associated with those words, Hoey is also contextualising psychologically Saussure's (1959:67) idea that signs constitute a combination of form (the "signifier") and meaning (the "signified"). Thus, in this way, Hoey positions words as being part of a network that consists not only of linguistic information but also information from other cognitive domains.

I would argue that Hoey would be justified in going even further in linking language to other cognitive processes. This is because, although language processing is traditionally thought to be largely restricted to two functionally specific brain regions in the left hemisphere, namely Broca's area and Wernicke's area, there is now growing recognition in psycholinguistics that language processing engages domain-general brain regions and mechanisms (Kutas & Dale 1997:228; Thompson-Schill et al. 2005:220). For instance, Fedorenko and Thompson-Schill (2014:121, 124) state that "domain-general brain regions – like those that support, for example, cognitive control and working memory – are likely to participate in all mental processes, including language comprehension and production". In addition, they question the existence of language-specific processes and argue that "the whole brain is probably engaged – in some way – during language processing".

There is evidence from cognitive neuroscience to suggest that language processing is linked to wider cognition. For example, there is evidence to show that the motor cortex is activated by reading verbs such as *kick* (Hauk et al. 2004), and the colour perception cortex is activated when processing words such as *purple* and *yellow* (Simmons et al. 2007). In addition, the auditory cortex is activated when processing words such as *telephone* (Kiefer et al. 2008).

Since there is neuroscientific evidence to suggest that processing words activates brain regions that are not functionally specific to language, this suggests that non-linguistic concepts should be represented in any model of language processing. However, this would be very difficult to achieve in practice.

Overall, given the prevalence of the network assumption across the different approaches to collocation, there is strong justification for building my own approach upon this network assumption in subsequent chapters. Additional justification comes from what is known about how neurons work in the brain. Since neurons interact and form networks that are capable of representing knowledge (Rosenweig et al. 1996:35; Sternberg 2009:35), language knowledge must also be represented in the brain in this way. Thus, it is intuitively plausible to propose a model of language that uses a representation similar to what we know about the ultimate neural substrate.

2.9 Main parameters of difference between the different approaches

In the previous section I demonstrated that some approaches to collocation explicitly discuss the idea that collocations are represented in the brain as transitions across a network, while other approaches only implicitly make this assumption. For instance, the network assumption is explicit in Construction Grammar while it is implicit in Wray’s notion of formulaicity. There is even variation across theories within the same approach. For example, although Sinclairian theory and the theory of Lexical Priming both fall within the neo-Firthian approach, the network assumption is explicit in Lexical Priming but implicit in Sinclairian theory.

In addition to differences in how the network assumption is addressed, there are many other parameters of difference between the different approaches. A clear parameter of difference is whether an approach posits two processing systems, with separate systems for the processing of grammar and lexis, or whether an approach posits just one processing system that accounts for the whole of language using collocational mechanisms. Sinclair posits two processing systems, namely the Idiom Principle and the Open-Choice Principle (1991). Wray also posits two processing systems, which she terms the analytic system and the holistic system. Yet both Sinclair and Wray believe that the grammatical system (namely the Open-Choice Principle and the analytic system, respectively) is secondary to the system which processes language using memorized preconstructed or prefabricated chunks (Sinclair 1991:114; Wray 2002:10). By contrast, the theories of Lexical Priming, Pattern Grammar, and Construction Grammar all attempt to account for the whole of language without positing the existence of a separate grammatical system.

Another parameter that distinguishes between the different approaches is whether the approach assumes that collocations have clear beginning and end points, or whether there are no clear boundaries to when a collocation starts and ends. Formulaic sequences are conceptualised as having clear beginning and end points. The lexical units posited in Sinclairian theory also seem to be conceptualised in this way, as Sinclair (1991:71) states that they “constitute single choices”. Similarly, the syntax-based approach and Construction Grammar assume clear beginning and end points in the sense that the collocations/constructions under analysis constitute a particular syntactic frame. By contrast, the different pattern configurations that are proposed in Pattern Grammar, such as pattern flow, pattern loop, and pattern accumulation, demonstrate how discourse might be constructed without assuming that there always has to be a clear beginning and end point to a collocation. Similarly, in Lexical Priming, the notions of priming and nesting can account for the sequencing of syllables, words, collocations, and the wider discourse structure, without the need to impose artificial boundaries on when a collocation starts and ends (Hoey 2005:158).

2.10 Defining and measuring psychological validity

In this thesis, I aim to contribute to our understanding of whether or not the phenomenon of collocation can be seen as having psychological validity. There are different types of psychological validity. Gries argues for a distinction between *psychological/cognitive reality* and *psychological entrenchedness*, with the former being determined experimentally (Gries & Mukherjee 2010:521), and the latter being determined by the frequency with which a collocation is encountered (Gries 2008b:409-410, 413; 2014:11; 2017:11). Importantly, any measure of psychological validity is only a proxy for what is happening at the neuronal level.

According to Gries and Wulff (2005:183), if a collocation (or construction) has psychological reality, this means that it has some sort of stored "mental representation" as an independent linguistic unit. The size of this mental representation is determined by the ease of processing of the collocation, as quantified by psycholinguistic measures such as number of eye fixations, reading speed, repetition speed and accuracy, and length of time taken to produce (Gries & Mukherjee 2010:521). However, "no position is taken as to the exact mental representation or processing" of collocations (Divjak & Gries 2008:198).

The aforementioned psycholinguistic measures are often correlated with frequency of occurrence of the collocation (Gries 2017:12). This suggests that, like entrenchedness, psychological reality is modulated by frequency. However, the psycholinguistic measures are also commonly correlated with factors such as dispersion (Gries 2017:12), i.e. the extent to which the collocation is evenly distributed across a corpus, rather than being concentrated in a small number of texts (Baker et al. 2006:59-60). Gries (2017:12) points out that "dispersion might actually be more highly correlated with reaction times" than frequency of occurrence, and goes on to state that "you are not only faster to react to a word if that word is more frequent in general but also if it is more widely used". This is unsurprising, as dispersion measures have been shown to have psychological validity outside the domain of corpus linguistics. For example, Ambridge et al. (2006:175) state that:

> Given a certain number of exposures to a stimulus, or a certain amount of training, learning is always better when exposures or training trials are distributed over several sessions than when they are massed into one session. This finding is extremely robust in many domains of human cognition.

Whereas psychological entrenchedness is determined solely by frequency of occurrence, psychological reality is determined by frequency in addition to other measures such as dispersion throughout the corpus. Frequency is just one of the measures of collocation strength that I investigate in Experiment 4 (Chapter 8). Therefore, in this thesis, I focus on psychological validity more generally, rather than the more narrowly defined concept of psychological entrenchedness.

2.11 Association measures and psychological validity

As mentioned in Chapter 1 (section 1.2), association measures are statistical scores which allow us to distinguish between words which co-occur due to chance, and words which co-occur due to true statistical association (Evert 2008:32). There are many different association measures, and the one chosen depends on the purpose of the study. For example, although the Dice coefficient (Weisstein 1999) cannot be used to identify words which are negatively associated, this association measure is particularly useful for identifying fixed collocations such as idioms (Smadja et al. 1996:7, 12; Evert 2008:1241).

Two of the most commonly used association measures in corpus linguistics are mutual information and log-likelihood (Gries 2014a:37). Log-likelihood (Dunning 1993) is an example of a *significance* statistic, meaning that it measures whether or not the "observed frequency" of a pair of words (i.e. a *bigram*) is higher than the "expected frequency" (Hoffman et al. 2008:150). Essentially, using a significance statistic is asking the question "how much evidence is there for a positive association between the words?" (Evert 2008:1233). By contrast, mutual information (Church & Hanks 1990) is an *effect size* statistic, defined as the "ratio between observed and expected frequency" (Hoffman et al. 2008:150). Essentially, using an effect size statistic is asking the question "how strongly are the words attracted to each other?" (Evert 2008:1233). Significance measures are known to assign high scores to high-frequency words, such as grammatical words, while effect size measures are known to assign high scores to low-frequency words (Evert 2008:1233; Gries 2017:118).

While some association measures are classified as significance statistics, and others are classified as effect-size statistics[4], there also exist *hybrid measures* which combine both significance and effect size, or significance/effect size and frequency (Hoffman et al. 2008:151). Examples of hybrid measures include the z-score (Dennis 1965), t-score[5] (Church et al. 1991), Dice coefficient (Weisstein 1999), and MI3 (Hoffman et al. 2008:151). Hoffman et al. (2009:151) explain that "t-score is a hybrid between frequency and significance, z-score is a hybrid between effect size and significance" and "both MI3 and Dice balance frequency and effect size, but do so in diametrically opposed ways".

In this thesis, rather than identifying collocations using one of the aforementioned association measures, I instead calculate the *transition probability* of adjacent words. Transition probability is an effect size statistic that is calculated by dividing the number of times the bigram X-then-Y occurs in a corpus by the number of times X occurs in the corpus

---

[4] Statistical measures which use only one dimension, whether that be significance, effect size, frequency, or transition probability, are referred to as "pure" measures (Hoffmann et al. 2008:154).

[5] Note that, in statistics, the z-score and t-score measures are seen as being measures of statistical significance rather than hybrid measures, but they work differently when applied to co-occurrence data (Hoffman et al. 2008:151).

altogether (McEnery and Hardie 2012:195). The justification for using this measure of collocation strength is provided in Chapter 4 (section 4.2.1).

Millar (2010:275) argues that "there is a need for more studies to validate the numerous statistical measures available to corpus linguists". Indeed, Gries (2013:153) states that:

> psycholinguists, who have been more eager to problematize and test our corpus-linguistic lexical association measures than we have ourselves, have produced an array of results that are not always easy to reconcile: sometimes, bidirectional measures such as co-occurrence frequency or *MI* predicts subjects' or speakers' patterning well, but sometimes unidirectional transitional probabilities (e.g. a/a+b) fare better.

Gries (2013:153) does not cite which psycholinguistic studies he is referring to. However, although the method of calculating transition probability cited by Gries is different from the method used in this thesis (see Chapter 4, section 4.2.1), it is interesting to see that he mentions that transition probabilities sometimes outperform mutual information (MI).

Gries has conducted some psycholinguistic experiments with the sole aim of investigating which measures of collocation strength can be seen as having the most psychological reality. In one study involving a sentence-completion task, Gries et al. (2005) find that an effect size measure of collocation strength (specifically, collostruction strength) strongly outperforms raw frequency. This result is replicated in a reading-time study by the same researchers (Gries et al. 2010). Similarly, Ellis and Simpson-Vlach (2009) produce experimental evidence to suggest that the mutual information score of *n*-grams is more indicative of their psycholinguistic validity than their raw frequency or length. In contrast, in a much larger-scale study, Wiechmann (2008) finds that raw frequency outperforms mutual information. In light of the inconclusive nature of the findings, Wiechmann (2008:283) argues that "there is still a strong need for empirical evaluations of competing measures of collocativity". This is exactly what I aim to achieve in Experiment 4.

2.12 Chapter summary and conclusion

In this chapter I have reviewed the different ways in which researchers discuss the notion of collocation, including the neo-Firthian approach, the syntax-based approach, and the usage-based approach. I have reviewed some of the theoretical and practical issues associated with these approaches, and I have outlined the key parameters of difference between them. Furthermore, I have argued that, across the different approaches, there is either an explicit or an implicit assumption that collocations are represented in the brain as transitions across a network.

On this basis, and on the basis of established knowledge of how neurons work in the brain, I have proposed *the network of language processing*, and I have decided to use this model as the basis for my conceputalization of collocation and the language system as a whole. Specifically, I have argued that collocations do not constitute a subset of the language and they do not have clear beginning and end points; rather, every word in a language exists in collocational patterns, and the strength of the collocational links between words for a given individual depends upon that individual's prior exposure to that particular collocational pattern. I have also concluded that I am investigating 'psychological validity' rather than 'psychological entrenchedness', and I have discussed the need to evaluate the psychological validity of the different association measures. In the next chapter, I review the psychological evidence for collocation from previous experimental work.

CHAPTER 3: A REVIEW OF PRIOR WORK ON THE PSYCHOLOGY OF COLLOCATION

3.1 Overview of the chapter

The first aim of this chapter is to review the previous studies that have investigated the psychology of collocation. As mentioned in the previous chapter, a huge array of terminology is used to refer to the notion of collocation, with different terms usually referring to different aspects of the same phenomenon (Wray 2002:8; Wray & Perkins 2000:3). In this chapter, in order to report the results of previous studies as accurately as possible, I will use the terminology that is used by the authors of each paper. Occasionally, however, I will suggest that a different term might have been more appropriate in capturing the particular type of collocation that is being studied.

The second aim of this chapter is to review the advantages and disadvantages of different experimental methods, namely self-paced reading, eye-tracking, and the ERP technique. I focus on collocation studies when discussing self-paced reading and eye-tracking; but, aside from in section 3.4.6 and 3.4.7 which specifically focus on ERP studies of collocation and predictability, I discuss language studies more generally when discussing the ERP technique. This is for two reasons. First, due to the complexity of the ERP technique, it is important to establish the fundamentals of language-based ERP studies before introducing work that is of more direct relevance to this thesis. Second, since the ERP technique is relatively new in the field of linguistics, there are at present very few studies which address the phenomenon of collocation, either explicitly (section 3.4.6.1) or non-explicitly (section 3.4.6.2). Nevertheless, this is likely to change in the near future, as the study of collocation and related notions (especially idioms and other fixed or semi-fixed multi-word expressions) is growing in cognitive psychology. Indeed, there is a growing literature of studies investigating multi-word expressions in a wide range of contexts, from language disorders (e.g. van Lancker

Sidtis & Fromkin 2017; Wray 2017) to language acquisition and learning (e.g. Theakston & Lieven 2017; Arnon & Christiansen 2017; McCauley & Christiansen 2017; Ellis & Ogden 2017). The importance of multi-word units across these different contexts is discussed in depth in Christiansen and Arnon (2017).

In terms of structure, I begin this chapter by discussing self-paced reading studies in section 3.2 and eye-tracking studies in section 3.3, before focusing on electrophysiological studies in section 3.4. Within section 3.4, I provide background information on EEG (section 3.4.1) and ERPs (section 3.4.2), and I introduce the electrophysiological markers of particular aspects of language processing in section 3.4.3 through to section 3.4.8. I then finish by outlining the advantages and disadvantages of the ERP technique in section 3.4.9 and providing a chapter summary and conclusion in section 3.5.

3.2 Self-paced reading studies

In self-paced reading experiments, participants are required to silently read either individual words or short phrases on a computer screen, and then press a specified key on the keyboard to reveal the next word or phrase. The amount of time taken to read each word or phrase is assumed to reflect the cognitive load associated with processing that element (McDonough & Trofimovich 2012:119).

Schmitt and Underwood (2004) conduct a word-by-word self-paced reading experiment in order to find out whether or not the final word of formulaic sequences is read more quickly than the final word of non-formulaic sequences by 20 native speakers and 20 non-native speakers of English. The results show that final words are processed “slightly faster” in formulaic sequences (albeit not statistically significantly so) than in non-formulaic sequences, for both native and non-native speakers. Schmitt and Underwood (2004:180-181,187) attribute the non-significance of the results to the nature of the self-paced reading experiment: “[t]he task, with its manual manipulation of the keyboard, might not be sensitive enough to measure the true differences in processing time inherent in the use of formulaic sequences". Moreover, they suggest that their reading time data might have been disrupted by the use of different reading strategies. For instance, some participants might focus on the meaning of each individual word, while others might quickly press the spacebar to get to the end of a sentence and then compute the meaning of the sentence as a whole. They also suggest that "it may be that the word-by-word nature of the task disrupts the holistic processing of formulaic sequences" (Schmitt & Underwood 2004:187). These are all clear disadvantages of the self-paced reading technique.

Conklin and Schmitt (2008) conduct a line-by-line self-paced reading experiment in order to find out whether or not formulaic sequences are read more quickly than non-formulaic sequences by 19 native speakers and 20 non-native speakers of English. Their results show that both participant groups read the formulaic sequences significantly more quickly than the non-formulaic sequences. However, in their study, the formulaic sequences consist of idioms; and the non-formulaic sequences used in their experiment are formed by rearranging the content words in the idioms and changing the function words. For example, *hit the nail on the head* becomes *hit his head on the nail* and *everything but the kitchen sink* becomes *everything in the kitchen sink*. Idioms "represent a special case of formulaicity" (Millar 2010:59) because their meaning cannot be derived from the meanings of the component words (Taylor 2002:540) (see Chapter 2, section 2.2). This suggests that idioms must be processed as holistic units (Wray 2002:18), but this is not necessarily the case for other formulaic sequences or collocations. This therefore limits the generalizability of the results.

In a later publication, Conklin and Schmitt (2012:50) themselves point out that "[m]uch of the research on formulaic sequence processing has been focused on idioms, and this is problematic for a number of reasons". They argue that, since idioms are not very frequent, they are not very useful for testing whether learners are sensitive to the statistical properties of language. Furthermore, since many idioms can be interpreted both figuratively and literally, the processing system has the added task of selecting between these competing interpretations. The results of Conklin and Schmitt's (2008) study are therefore not likely to be generalizable to non-idiomatic collocations.

Millar (2010) conducts three word-by-word self-paced reading experiments in order to study how native speakers of English process learner collocations. A *learner collocation* is operationalized by Millar as a collocation which occurs in a learner corpus but not the BNC1994 and which is intuitively unacceptable from a native speaker's perspective. Millar's (2010) experimental design is based on collocational learner bigrams (e.g. *best partner*) alongside bigrams with an equivalent meaning (*ideal partner*) and a mutual information (MI) score above 3 found in the BNC1994, embedded into sentences used for self-paced reading. The results of Millar's first experiment show that native speakers of English read the native speaker bigrams found in the BNC1994 significantly faster than the learner bigrams. Furthermore, the words which appear two tokens to the right of the bigram are also read significantly faster if they follow a native speaker bigram compared to if they follow a learner bigram. Finally, the results of Millar's (2010) third experiment shows that nouns in native speaker adjective-noun bigrams are read significantly faster than nouns in learner adjective-noun bigrams, even for learner bigrams that are *also* found in the BNC1994. In this case, the learner bigrams are weaker bigrams (MI score below 3 in the BNC1994) whereas the native speaker bigrams are relatively stronger bigrams (MI above 8). These results demonstrate that word-by-word self-paced reading experiments *are* sensitive enough to measure the differences in processing times associated with reading collocations of different strengths.

In a more recent self-paced reading experiment which builds on the work of Millar (2010), Hughes and Hardie (forthcoming) investigate whether or not adjective-noun bigrams with a high transition probability (e.g. *complex process*) are processed more quickly than adjective-noun bigrams with a low transition probability (e.g. *complex question*) by 20 native speakers of English and 20 non-native speakers of English. The results show that the stronger bigrams are processed significantly more quickly than the weaker bigrams by both the native and non-native speaker groups. Moreover, there is a significant interaction between the reading times for the nouns in each condition and whether the participant is a native or non-native speaker.

Interestingly, the difference in median and mean reading times is larger for the non-native speaker group compared to the native speaker group, suggesting that the non-native speakers are more sensitive to the transition probabilities between words than the native speakers. This is explained by assuming that the weaker bigrams are not entrenched in the minds of the non-native speakers to the extent that they are for the native speakers. Specifically, the non-native speakers might not have encountered the weaker bigrams before or, if they have, they might not have encountered them frequently enough for them to become entrenched. This would increase the difference in noun reading times between the two conditions relative to the native speakers, who would almost certainly have encountered all of the experimental bigrams – even the weaker ones – frequently enough for them to become at least somewhat entrenched.

Following Millar (2010), Hughes and Hardie (forthcoming) also look for spillover effects in the data by investigating whether or not the difference in processing time between the bigrams across conditions is sustained to the words that follow the bigrams. The results show that, while spillover effects are apparent on 4/5 of the experimental sentences for the non-native speakers, they are apparent on only 2/5 of the experimental sentences for the native speakers. Furthermore, while the spillover effects are sustained for up to 5 words for the non-native speakers, they are sustained for only 1 word for the native speakers. This, again, is attributed to differential entrenchment, due to the fact that the weaker bigrams might have been encountered less frequently by the non-native speakers than the native speakers. These findings are used as the basis for Hypothesis 4 in Experiment 3 (see Chapter 7, section 7.2), which states that the ERP responses for the non-native speakers will be larger than those demonstrated by the native English speakers from Experiment 2 (thereby reinforcing the finding that non-native speakers are more sensitive to the transition probabilities between words).

A major problem with the reading that takes place during self-paced reading experiments is that it does not accurately reflect the "normal" reading process (Kaiser 2014:141). Normal reading does not involve reading words one at a time, and it does not involve pressing a key on a keyboard to progress through the text. This slows the pace of reading, "giving the reader extra time that could potentially be used to engage processes that do not normally take place" (Rayner et al. 2012:221). Eye-tracking causes comparatively less disruption to the reading process, but the same cannot be said for electrophysiological studies of language because, as with self-paced reading experiments, electrophysiological studies require sentences to be presented word-by-word. This is so that any effects can be time-locked to the experimental words (see Chapter 3, section 3.4.2)

3.3 Eye-tracking studies

In eye-tracking experiments, participants are required to read text on a computer screen whilst an eye-tracking camera tracks their pupil and corneal reflection (Holmqvist et al. 2011:181). The measures most frequently used to gain an indication of the cognitive load required to process each word are related to the number and duration of fixations, i.e. the points at which the eyes are almost still (Rayner 1998:373).

McDonald and Shillcock (2003a; 2003b) find via eye-tracking experiments that participants fixate on the second word of a bigram for a significantly longer period of time if they are reading a bigram with a low transition probability (e.g. *avoid discovery*), compared to a bigram with a high transition probability (e.g. *avoid confusion*). However, in a replication study, Frisson et al. (2005) find that the difference in fixation times is more likely to be a product of the contextual predictability of the bigrams as opposed to their transition probabilities. Indeed, words which are highly predictable based on the semantics of the preceding words (see Chapter 4, section 4.5.3) receive fewer and shorter fixations than words which are less contextually constrained (Ehrlich & Rayner 1981:652).

Underwood et al. (2004) conduct an eye-tracking experiment examining whether or not the processing of the final word of a sequence is facilitated if the word is part of a formulaic sequence as opposed to a non-formulaic sequence. An example target word is the word *basket* in the formulaic sequence *put all your eggs in one basket* and the non-formulaic sequence *it was you who dropped my flower basket*. For both native and non-native speakers of English, they find significantly less fixations when the final word is part of a formulaic sequence. For the native speakers, the fixations are also significantly shorter when the final word is part of a formulaic sequence. However, as with the Conklin and Schmitt study discussed above, since Underwood et al. (2004) focus on fixed idioms and other restricted formulaic sequences, the results of this study cannot be generalised to account for less restricted collocations.

In a more recent study, Huang et al. (2012) demonstrate via an eye-tracking experiment that both native and non-native speakers of English process the final word of a sequence significantly faster if the word is part of a sequence with high transition probabilities as opposed to a sequence with low transition probabilities. As far I am are aware, there are no other studies using any psycholinguistics methods which investigate whether learners of English are sensitive to the transition probabilities between words.

In the foregoing review, I have illustrated the general pattern of much psycholinguistic experimental research into collocation: measuring the cognitive load of processing collocational/formulaic word sequences versus non-collocational/non-formulaic sequences, via a proxy measure for cognitive load, most commonly speed of word reading or degree of eye fixation. A limitation of these measures is that, by being proxy measures, they are only able to provide an indirect measure of cognition. As a result, to reach any conclusions about language processing, we need to draw inferences from the behavioural and eye-tracking data, and we are therefore only able to speculate about what might be happening in the brain. I will now discuss electrophysiological studies, which provide a more direct measure of cognition by measuring how the electrical activity of the brain changes in response to specific experimental stimuli.

3.4 Electrophysiological studies: Electroencephalography and event-related potentials

3.4.1 Background information on EEG

3.4.1.1 Overview

Electroencephalography, or EEG, is a non-invasive neuroimaging method which uses scalp electrodes to measure the electrical activity of the brain (Harley 2008:17, 494). The recorded signal is the difference in electrical potential between two electrodes (Martin 2006:22), where *electrical potential* is the energy expanded by an electromagnetic force in moving electrically charged particles (i.e. *ions*) from point A to point B (Thompson 2000:67; Crowell 2003:101). Electrical potential is measures in *volts* (V) (Duffin 1968:xv). However, since the electrical activity in the brain is very small (Darby & Walsh 2005:73), neural electrical potentials are measured in microvolts (μV) (Martin 2006:21-22). These signals are then amplified by the EEG equipment before they can be analysed (Martin 2006:22). 1 microvolt is equivalent to 1 millionth of a volt (Tyner et al. 1983:37).

The electrical activity in the brain that is detected in EEG experiments is displayed as a continuous waveform. This waveform appears as “a series of positive and negative peaks that vary in polarity, amplitude, and duration as the waveform unfolds over time” (Kappenman & Luck 2012:4). The width of the wave at a given point in time indicates the frequency of the electrical activity in hertz (Hz) at that millisecond, while the height of the wave indicates the amplitude of the electrical activity in microvolts (μV) (Sternberg 2009:41). This waveform display of brain signals is known as an *electroencephalogram*, while the recording device is known as an *electroencephalograph*. The abbreviation EEG can be used to refer to electroencephalography, electroencephalogram, or electroencephalograph. However, in this thesis, I use this abbreviation only when referring to the method (electroencephalography) as opposed to the recording device or the waveform display.

Since the electroencephalogram displays the frequency of the electrical activity at each millisecond of the EEG experiment, the electrical activity can be analysed in either the time domain or the frequency domain. When analysing the electrical activity in the frequency domain, the frequency spectrum is divided into at least four distinct bands: namely *delta* (below 4 Hz), *theta* (4-8 Hz), *alpha* (8-13 Hz), and *beta* (above 13 Hz) (Kiloh et al. 1981:53). Different brain states are associated with each frequency band. For example, high beta activity is associated with anxiety and overarousal while alpha activity is associated with relaxation, fatigue, and lack of concentration (Zillmer et al. 2008:41; Tiago-Costa 2016:90).

Frequency domain analysis is often carried out in clinical settings as it can help to assess levels of consciousness (Kiloh et al. 1981:53). However, it is less useful in cognitive neuroscience. This is because, while time domain analysis can be conducted on transient bursts of neural activity that occur in response to a particular stimulus, frequency domain analysis requires at least two complete cycles in a waveform (Kiloh et al. 1981:53), i.e. given that EEG experiments focus on frequencies on the order of 10 Hz, two cycles would require at least 1/5 of a second of stimulus-bound neural activity. This length of stimulus-bound activity would not typically be elicited in an EEG experiment. Therefore, in this thesis I focus on the time domain as opposed to the frequency domain. This is in line with previous language-related EEG studies, almost all of which focus on the time domain (Kutas & Van Petten 1994:88). Time domain analysis centres on *event-related potentials* (ERPs), which are introduced in section 3.4.2.

3.4.1.2 Source of electrical activity

To explain the source of the electrical activity detected by scalp electrodes, it is necessary to first describe some of the major divisions and structures of the human brain. As illustrated in figure 3.1, the human brain is divided into five distinct layers, namely the telencephalon, the diencephalon, the mesencephalon, the metencephalon, and the myelincephalon (Pinel 2014:68).

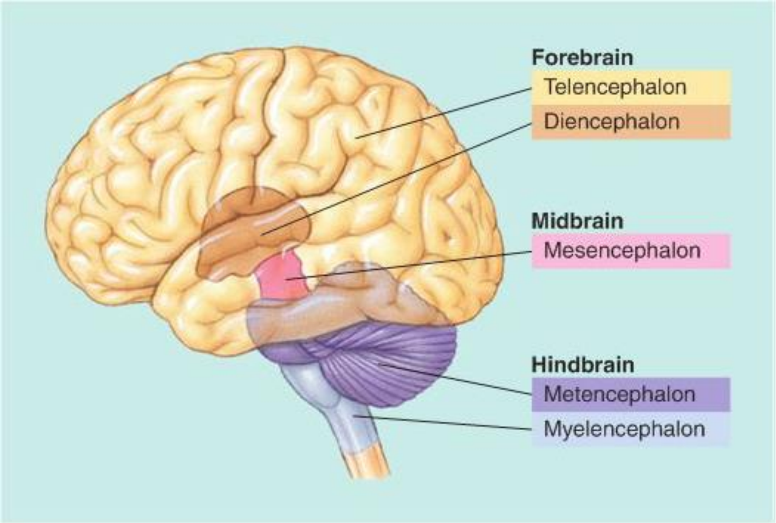


Figure 3.1: The divisions of the human brain (from Pinel 2014:68)

The lower four divisions of the brain are collectively referred to as the *brain stem*. The brain stem is responsible for regulating basic physiological functions such as heart rate and respiration (Pinel 2014:559). By contrast, the highest and largest division of the brain, the telencephalon, is responsible for mediating complex cognitive functions such as language, problem solving, and voluntary movement (Pinel 2014:70). The telencephalon is also known collectively as the cerebral hemispheres, and each hemisphere is divided into four lobes, namely the temporal, parietal, occipital, and frontal lobes. These are schematized in figure 3.2.

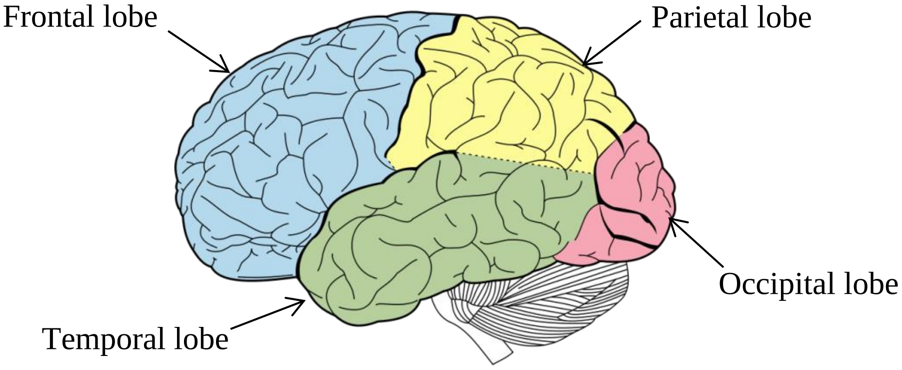


Figure 3.2: The lobes of the cerebral hemispheres (from Gray 1918)

Each lobe is thought to be broadly associated with a specific set of functions. For example, the occipital lobe is related to vision while the temporal lobe is related to language and hearing (Pinel 2014:72). However, these are very broad generalisations. For instance, as mentioned in Chapter 2 (section 2.8), many areas of the brain are now believed to be involved in language processing (Kutas & Dale 1997:228; Thompson-Schill et al. 2005:220). The lobes should therefore be considered to be structural regions of the brain that are not necessarily associated with particular functions (Pinel 2014:71).

As well as containing the four lobes of the cerebral hemispheres, the telencephalon also contains structures such as the limbic system, which itself contains distinct structures. The limbic system is associated with a broad range of functions including the regulation of sleep and sexual behaviour. Two important structures within the limbic system include the basal ganglia, which is associated with voluntary muscle movements (Pinel 2014:74) and the amygdala, which is associated with emotion (Pinel 2014:74). Importantly, as well as containing the structures which comprise the limbic system, the outer part of the telencephalon contains a 1-3 mm layer of neural tissue known as the *cerebral cortex* (Pinel 2014:70; Sternberg 2009:58). The cerebral cortex contains three types of cell, namely stellate, spindle, and pyramidal cells (Kiloh et al. 1981:24).

There are over 200 distinct types of nerve cell in the cerebral cortex, yet they all share some basic structural properties (Rosenzweig et al. 1996:36-37). Specifically, nerve cells (or neurons) have a *cell body* (also known as a *soma*) which contains the nucleus. They also have multiple *dendrites*, which are the receptive surfaces of the cell, as well as a single *axon*, which transmits signals on to neighbouring cells (Rosenweig et al. 1996:37). These basic structural properties are illustrated in figure 3.3.

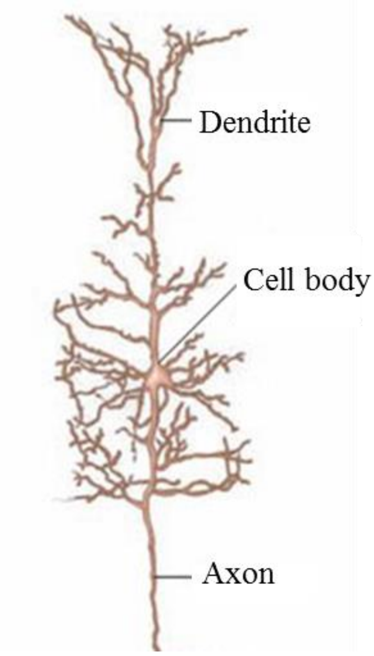


Figure 3.3: Basic structure of a cortical nerve cell (from Nolen-Hoeksema et al. 2009:37)

The signals from a single neuron are too weak to be detected at the scalp (Kutas & Dale 1997:203). Yet, when thousands of neurons are simultaneously activated, the signal is often large enough to be detected by scalp electrodes (Kappenman & Luck 2012:5). However, there are other prerequisites for signal detection. Specifically, the neurons must have the same orientation in order to avoid signal cancellation, and they must be aligned perpendicularly to the scalp (Kutas & Dale 1997:202). Furthermore, they must have dendrites that are long enough to form effective dipoles (Kiloh et al. 1981:21). A *dipole* is defined as "a small region of electrical current with a relatively positive end and a relatively negative end" (Banich 1997:77-78).

The only cells that meet all of these requirements are cortical *pyramidal cells* (Kappenman & Luck 2012:5). These cells have a pyramid-shaped soma (Martin 2010:927) and a large dendrite known as an *apical dendrite* which extends towards the cortical surface (Pinel 2014:72). Furthermore, they are considered to be "the main input-output cells of the cerebral cortex" (Luck 2014a:40). The electrical activity that is detected by scalp electrodes arises mainly from these cortical pyramidal cells (Kutas & Van Petten 1994:87; Kappenman & Luck 2012:5). Thus, because of this, EEG provides only a partial measure of brain activity (Kiloh et al. 1981:46).

It is also important to note that scalp electrodes only detect "the inputs to a group of neurons rather than the outputs of those neurons" (Kappenman & Luck 2012:5). The "inputs" refer to *postsynpatic potentials* whereas the "outputs" refer to *action potentials*. As mentioned in Chapter 2 (section 2.8), an *action potential* is a momentary change in a neuron's membrane potential from -70 mV to +50 mV (Pinel 2014:557). The action potential causes neurotransmitters (i.e. electrochemical signals) to be released and, when these neurotransmitters bind to receptors on neighbouring cells, this causes a change in voltage across the receptive cells. This change in voltage is the *postsynaptic potential* (Luck 2014a:39). Action potentials cannot typically be detected at the scalp because it is uncommon for sufficient numbers of neurons to fire at exactly the same time (Luck 2014a:39). Therefore, the electrical activity detected at the scalp usually reflects postsynaptic potentials (Kingsley 2000:526; Kappenman & Luck 2012:5).

3.4.2 Background information on ERPs and ERP components

Event-related potentials (ERPs) are "the momentary changes in electrical activity of the brain when a particular stimulus is presented to a person" (Ashcraft & Radvansky 2010:61).

An *event* refers to any stimulus or response that is under investigation by the researcher, and the electrical brain response is said to be *time-locked* to the event (Kutas & Van Petten 1994:87). This means that, every time the event occurs in the experiment, an *event code* (also known as a *trigger code* or a *timing marker*) is added to the EEG data so that the researcher can analyse the electrical activity that occurs in the milliseconds following the event. The events can be marked as belonging to different conditions of the experiment. In this way, the researcher can investigate the electrical activity that is associated with the different experimental conditions.

An *ERP component* can provisionally be defined as “a scalp-recorded voltage change that reflects a specific neural or psychological process” (Kappenman & Luck 2012:4). The voltage change associated with a particular component has a unique scalp distribution and occurs at a particular time point following the onset of the stimulus (Kutas & Dale 1997:205). The amplitude and latency of ERP components is traditionally quantified by measuring the amplitude and latency of the *peaks* in a waveform, i.e. “the points in an ERP waveform where the voltage reaches a maximally positive or maximally negative value” (Luck 2014a:54). However, this method of identifying ERP components will be problematized in section 3.4.8.

There are various ways of labelling ERP components. The most common approach is to begin the label with a <P> or <N> to indicate whether the component has a positive or negative amplitude (Kolb & Whishaw 1996:142). This <P> or <N> is then followed by a number which indicates the latency of the component in relation to the stimulus onset (Kutas & Van Petten 1994:86). For example, the term *P600* refers to an ERP component which is characterized by a positive shift in the ERP waveform occurring roughly 600 ms after the onset of the stimulus. Similarly, the term *N400* refers to an ERP component which is characterized by a negative shift in the ERP waveform occurring roughly 400 ms after the onset of the stimulus. Alternatively, the P or N can be followed by a number which indicates the component’s ordinal position in the waveform (Kutas & Van Petten 1994:86). For example, the P300 is also known as the *P3* because it is the third major positive peak in a waveform (Picton 1993:456). The P300 can also be referred to as the *late positive component*, or LPC (Polich 2007:2128).

As well as being labelled according to their latency or ordinal position in a waveform, ERP components can also be given functional labels (Brown & Hagoort 2000:217). For example, the *stimulus preceding negativity* (SPN) refers to an ERP component that is elicited while subjects are pre-empting a stimulus that is expected to appear within a few seconds (Brunia et al. 2012:190). These labels can sometimes include both functional information and information about the scalp distribution of the component. For example the *lateralized readiness potential* (LRP) refers to an ERP component that is elicited immediately before a participant presses a response button with one hand. This component is described as being lateralized (i.e. occurring in a particular hemisphere of the brain) because the voltage shift is observed over the motor cortex that is contralateral (i.e. on the opposite side of the body/brain) to the response hand (Smulders & Miller 2012:209).

Both domain-general and domain-specific ERP components have been identified. An example of a domain-specific component is the N170, which is associated with face recognition (Rossion & Jacques 2012:117). By contrast, the P300 is an example of an ERP component that has been shown to be associated with a wide range of cognitive functions spanning across the domains of attention, memory, and probability sensitivity (Polich 2007:2132, 2137). The relationship between the P300 and probability sensitivity will be discussed further in section 3.4.7.

In this thesis I focus mainly on two language-related ERP components, namely the N400, which is broadly associated with lexical/semantic processing (Kutas & Hillyard 1980), and the P600, which is broadly associated with syntactic processing (Osterhout & Holcomb 1992). The peaks roughly corresponding to these two components are shown in figure 3.4.

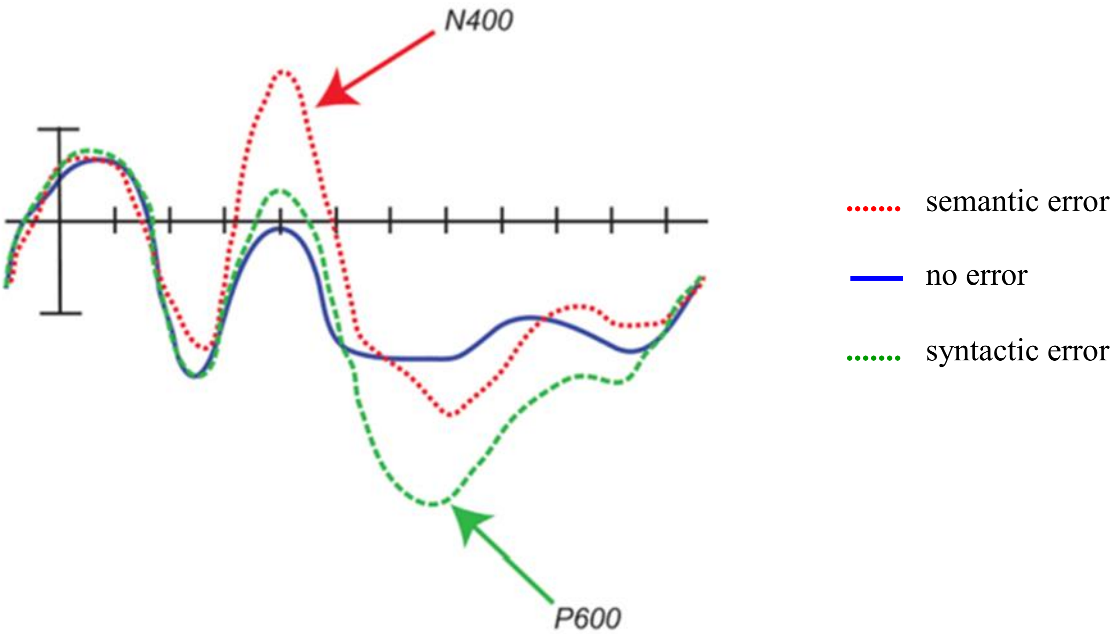


Figure 3.4: An ERP waveform displaying the N400 and P600 (adapted from Swaab et al. 2012:422)

The N400 is a negative shift in the ERP waveform occurring roughly 400 ms after stimulus onset that is elicited when a participant reads a semantic error in a sentence (Kutas & Hillyard 1980); the P600 is a positive shift occurring roughly 600 ms after stimulus onset that is elicited when a participant reads a syntactic error in a sentence (Osterhout & Holcomb 1992). Note that, although the ERP waveform in figure 3.4 is plotted “negative up”, it is common in the more recent ERP literature to plot waveforms “positive up” (Keil et al. 2013:8). I will use the “positive up” alternative throughout this thesis.

3.4.3 The N400 as an electrophysiological marker of semantic processing

The most widely studied language-related ERP component is the N400 (Swaab et al. 2012:399). The N400 is a negative shift in the ERP waveform that is largest over central-posterior scalp regions (Swaab et al. 2012:399) and that peaks at roughly 400 ms after stimulus onset (Kutas & Hillyard 1980). The N400 was first discovered by Kutas and Hillyard (1980), who conducted two experiments which each involved presenting 160 word-by-word sentences to participants. The sentences were all 7 words long and, in 25% of the sentences, the seventh word was semantically inappropriate but syntactically correct. The semantic incongruity between the last word in the sentence and the rest of the sentence was considered to be either "moderate" (e.g. *He took a sip from the* *waterfall*) or "strong" (e.g. *He took a sip from the transmitter*).. Kutas and Hillyard (1980) found that, in every participant, the semantically inappropriate word elicits an N400. Furthermore, the amplitude of the N400 is directly proportional to the strength of the semantic incongruity.

Since Kutas and Hillyard's (1980) seminal article, the N400 effect has been widely replicated across different experimental paradigms. For instance, in a word pair experiment conducted by Kutas and Hillyard (1989), each trial consists of a single word followed by another single word, followed by a single orthographic letter. After reading each word pair, participants have to decide whether or not the letter is present in either or both of the preceding words. Some of the words in the word pairs are highly semantically related, others are moderately semantically related, while others are not semantically related. The results show that the amplitude of the N400 is inversely related to the level of semantic relatedness, with the amplitude of the N400 being significantly larger in the condition containing unrelated words compared to the conditions containing moderately related or highly related words (Kutas & Hillyard 1989:41). This shows that the N400 effect is not restricted to semantic incongruities in sentences.

The N400 effect is also not restricted to studies of written English, as the N400 is elicited in response to semantic incongruities in spoken English as well as in a variety of languages (Kutas & Dale 1997). This is demonstrated in a study by Kutas et al. (1987), which investigates the N400 effect in written English, spoken English, and American Sign Language.

In this study, three groups of participants are presented with 100-135 sentences. In half of the sentences, the last word is semantically congruent with the rest of the sentence; in the other half, the last word is semantically incongruent with the rest of the sentence. After reading each sentence, participants have to press a button to indicate whether or not they think that the sentence makes sense. One group of 12 participants read the sentences word-by-word on a computer screen. Another group of 12 participants use headphones to listen to the sentences spoken at a normal speaking rate, and a group of 10 participants whose first language is American Sign Language view the signed sentences on a computer screen. The results show that the N400 is elicited in the semantically incongruent condition in the written English task, the spoken English task, and the American Sign Language task. This shows that the N400 is not specific to one linguistic modality, suggesting that it is related to semantic processing more generally.

3.4.4 The P600 as an electrophysiological marker of syntactic processing

Another widely studied language-related ERP component is the P600, which is also known as the *syntactic positive shift* or *SPS*. The P600 is a positive shift in the ERP waveform that peaks at roughly 600 ms after stimulus onset (Swaab et al. 2012:418). While some researchers cite the P600 as having a posterior scalp distribution (e.g. Swaab et al. 2012:22), with no distinct laterality, others state that it has a central-posterior scalp distribution (e.g. Ingram 2007:323; Friederici & Männel 2013:186).

The P600 was first discovered by Osterhout and Holcomb (1992), who conducted a study which involved presenting 180 word-by-word sentences to 15 participants. After reading each sentence, participants had to press a button to indicate whether they thought that the sentence was “acceptable” or “unacceptable” in terms of whether or not the sentence was “semantically coherent and grammatically correct”. Of the 180 sentences, 30 were grammatically incorrect as they contained a transitive verb without a direct object, e.g. **The* *woman advised to see the movie*. These grammatically incorrect sentences were each paired with a grammatically correct sentence that was identical up to but not including the target verb, e.g. *The woman agreed to see the play*. The results show that the verbs in the grammatically incorrect sentences elicited a P600, whereas the P600 was not elicited by the verbs in the grammatically correct sentences. This demonstrates that the P600 is sensitive to syntactic violations.

As Osterhout and Holcomb (1992:798) pointed out, it is problematic to suggest that the P600 is elicited uniquely in response to syntactic violations. This is because using a transitive verb without a direct object automatically causes a semantic error in addition to a grammatical error. Moreover, structural violations in domains other than language have been shown to elicit the P600. For instance, Núñez-Peña and Honrubia-Serrano (2004) demonstrate that the P600 is elicited in response to reading an error in a mathematical sequence such as *4, 7, 10, 13, 16, 19,* <u>*23*</u>. Furthermore, in a direct comparison of language and music, Patel et al. (1998) demonstrate that the P600 is elicited in response to encountering incongruities in both linguistic and musical structure (where a structural incongruity in music is operationalised as an out-of-key chord in a musical phrase). These results show that the P600 is sensitive to structural violations not only in language but in other domains as well.

3.4.5 Relationship between semantic and syntactic processing

The aforementioned studies suggest that distinct neuropsychological processes are involved in syntactic and semantic processing. If there *are* separate components for syntactic and semantic processing, this provides evidence against theories of language which argue that syntax and semantics are inseparable (see Chapter 2, section 2.9). However, other studies have shown that the P600 is actually sensitive to semantic errors in the absence of any syntactic errors (Kuperberg 2007:23). For instance, Geyer et al. (2006) find that the P600 is elicited in response to reading implausible sentences such as *Tyler cancelled the* <u>*tongue*</u> but not in response to reading plausible sentences such as *Tyler cancelled the* *<u>subscription</u>*. Similarly, Kuperberg et al. (2006) find that the P600 is elicited in response to semantic role animacy violations in sentences such as *to make good documentaries* *<u>cameras</u>* *must interview…*. Based on these results, Kuperberg et al. (2006:527) conclude that "the neural systems supporting syntactic and semantic processing may be linked".

While some studies have shown that the P600 can be sensitive to semantic errors, other studies have shown that the N400 can be sensitive to morphosyntactic errors (Tanner & van Hell 2014:298). For instance, Severens et al. (2008) find that an N400 is elicited by reading errors in subject-verb agreement, and Nieuwland et al. (2013) find that an N400 is elicited in response to reading animate object nouns with incorrect case marking.

Swaab et al. (2012:26, 28) state that "[r]esults such as these … raise serious and interesting questions about the relationship between semantic and syntactic processes in the brain", arguing that "the interaction between semantic and syntactic processes in the brain may be more dynamic than was previously suggested". This suggests that there is not a separate component specifically reflecting syntactic processing and another component specifically reflecting semantic processing. Therefore, if the distinction between syntax and semantics is not as clear-cut as it is often suggested to be (see Chapter 2, section 2.9) this provides evidence in support of theories of language which argue that syntax and semantics are partly or wholly inseparable.

3.4.6 ERP studies of collocation

3.4.6.1 The explicit approach

Some ERP studies explicitly state that they focus on collocation; however, they conceptualize or operationalize collocation differently from how it is conceptualized and operationalized in this thesis. In one such study, collocation is defined as "a sequence of words or terms which co-occur more often than would be expected by chance" and the collocations are extracted from a Spanish corpus (Molinaro & Carreiras 2010:176). However, although this corpus-based approach seems similar to the approach taken in this thesis, Molinaro and Carreiras (2010) focus on whether or not the collocations are literal and figurative and they explicitly state that all of the collocations that they extract from the corpus are either "idioms or clichés". They therefore focus on expressions which are much more fixed and non-compositional than those which are used as experimental stimuli in this thesis.

In this study conducted by Molinaro and Carreiras (2010), 36 native speakers of Spanish are asked to read 56 Spanish sentences containing literal idioms or clichés, and 56 Spanish sentences containing figurative idioms or clichés. In one condition, the original idioms and clichés are used, and the last word of each of these idioms and clichés has a high cloze probability. In the other condition, the last word of the idioms and clichés is replaced with another word that is either synonymous with the original word or semantically related. The final word in the string is therefore semantically appropriate but unexpected in the context.

The results of Molinaro and Carreiras' (2010) experiment show that the literality of the string modulates the N400, while the predictability of the final word modulates the P300, i.e. a positive shift in the ERP waveform that is largest over temporal-parietal scalp regions and that peaks roughly 300 ms after stimulus onset (Polich 2007:2128). Collocations with a high cloze probability elicit a larger P300 than the same word strings with unexpected endings. These results are reinforced by Vespignani et al. (2010), who also find that expected endings in idioms elicit a P300. However, these findings are very unusual in the context of P300 research because, as will be explained in section 3.4.7, the P300 is traditionally thought to be elicited in response to reading or viewing *un*expected or unpredictable stimuli.

In a more recent study conducted by Molinaro et al. (2013), 88 of the stimuli items from their previous experiment are re-used. However, whereas in the previous experiment they refer to the stimuli as collocations, and sub-divide this notion of collocation into idioms and clichés, in this study they refer to the same stimuli as “multi-word expressions”. Furthermore, they claim that “in corpus linguistics, a sequence of words that co-occur more often than would be expected by chance is defined as a multi-word expression and the relation between single words as ‘collocation’” (Molinaro et al. 2013:124). However, in their previous study, they use this exact definition of multi-word expressions to define collocation (Molinaro & Carreiras 2010:176). This inconsistent use of linguistic terminology can potentially be attributed to the fact that the researchers are primarily psychologists, not linguists, and so they do not use the term *collocation* in precisely the same way as it is used in corpus linguistics; or, rather, they have picked up on two contradictory operationalisations of the general concept. This highlights a wider problem whereby, although the terminology surrounding collocation is used inconsistently in linguistics, the terminology is used even more inconsistently in psychology, making it more difficult to compare results across different studies.

The Molinaro et al. (2013) study involves 36 native speakers of Spanish reading 254 sentences presented word-by-word. The experiment contains three conditions. The target word is kept the same across each condition but the preceding context is different in each condition. In the “high-constraining collocational” condition, the target word forms the last word of a fixed corpus-derived multi-word expression. In the “high-constraining semantic compositional” condition, the target word is preceded by a sentence stem that creates a high semantic contextual constraint. Finally, in the “low-constraining control” condition, the target word is not part of a fixed multi-word expression and it is not preceded by a semantically constraining sentence stem.

The results of this study show that a very early component known as the N1 is smaller when the target word is expected as part of a fixed multi-word expression compared to when the target word is expected due to sematic constraints. The N1 is a negative shift in the ERP waveform that is largest over anterior scalp regions and that peaks roughly 100 ms after stimulus onset (Luck 2014a:76). The N1 is traditionally associated with early perceptual processing and spatial attention (Mangun & Hillyard 1991; Mangun 1995). Molinaro et al. (2013:130) explain their finding that there is a smaller N1 in the high-constraining collocational condition compared to the high-constraining semantic compositional condition by suggesting that reading words in collocational patterns reduces demands on the visual-orthographic processing system. This suggests that reading the collocational bigrams in my experiments will place fewer processing demands on participants than reading the non-collocational bigrams.

However, by focusing on multi-word expressions, Molinaro et al. (2013) conceptualize and operationalize collocation somewhat differently from how it is conceptualized and operationalized in this thesis. In particular, they state that "[i]n the collocational context only one possible candidate is pre-activated and its activation is not competing with any other item" (Molinaro et al. 2013:131). By contrast, by focusing on transition probabilities, in this thesis I am assuming that the activation of a word is always competing with the activation of other words. However, I would argue that, even in fixed multi-word expressions, there are still options for what word might come next; there is just a very high transition probability predicting each word in the expression.

While the previous studies conceptualize and operationalize collocation differently from how I have conceptualized and operationalized collocation in this thesis, Tremblay and Baayen (2010) actually use a comparable probability-based operationalization of collocation. Tremblay and Baayen (2010) extract 432 four-word sequences (ABCD) from the BNC1994 and determine the probability of word D coming immediately after word string ABC. Under their method, the probability "is equal to the frequency of the whole string divided by the sum of the frequencies of every four-word sequence that share the first three words minus the frequency of that string" (Tremblay & Baayen 2010:162). Tremblay and Baayen (2010:168) find that "the N1 component is more positive for lower probability sequences and more negative for higher probability ones”. In other words, reading sequences of words which are highly likely to occur together elicits a larger N1 compared to reading sequences of words which are highly unlikely to occur together. These results are in opposition to those of Molinaro et al. (2013) as, while Tremblay and Baayen (2010) find a larger N1 for high probability sequences, Molinaro et al. (2013) find a smaller N1 for high probability sequences.

This discrepancy in results could be attributed to differences in experimental setup (Brandeis & Lehmann 1986:157), but it could also be related to Luck’s (2014a:33) criticism that, when a shift in the ERP waveform occurs so soon after stimulus onset, it is problematic to attribute this voltage shift to an experimental effect. Luck (2014a:75) explains that many different processes and brain regions are activated within the first 100 ms after the onset of a visual stimulus, so it is difficult to disentangle the experimental effect from effects that are associated with the physical properties of the stimulus.

This criticism could also be applied to Tremblay and Baayen’s (2010) other finding concerning the P1: another very early component. The P1 is a positive shift in the ERP waveform that is largest over lateral occipital scalp regions and that peaks roughly 100 ms after stimulus onset (Hillyard & Anllo-Vento 1998). Tremblay and Baayen (2010:21) find that the amplitude of the P1 is inversely related to the probability of the collocation. They argue that, due to the speed at which the P1 and N1 voltage deflections appear after stimulus onset, the P1 and N1 effects found in their experiment provide evidence that the “four-word sequences are retrieved in a holistic manner … rather than computed online via rule-like processes”. Moreover, they claim that their results “are exactly in line with usage-based accounts of grammar” (see Chapter 2, section 2.7). However, for reasons outlined above, it is important to be cautious when interpreting voltage shifts that occur very soon after the onset of the stimulus.

In ERP research, there are technical terms for early versus late components. Components which peak within 100 ms of stimulus onset are known as *sensory* or *exogenous* components, and they "depend largely on the physical parameters of the stimulus"; components which peak later than 100 ms are known as *cognitive* or *endogenous* components, and they are more likely to reflect cognitive information processing (Sur & Sinha 2009:70). In my experiments I focus only on endogenous components so that I can make more reliable claims about cognition.

Finally, in the most recent ERP study to explicitly focus on collocation, Siyanova-Chanturia et al. (2017) extract 40 binomial collocations such as *knife and fork* from the BNC1994. Then, for each binomial collocation, they manufacture 40 binomial non-collocations that are grammatically plausible and semantically congruous (e.g. *spoon and fork*), and 40 binomial non-collocations that are grammatically plausible but semantically incongruous (e.g. *theme and fork*). The content words in the binomial collocations share the same level of semantic association as the content words in the semantically congruous binomial non-collocations (according to the University of South Florida Free Association Norms database).

The binomials across all three of these conditions are presented word-by-word to 48 participants, and the EEG activity is time-locked to the final word. In Part 1 of the study, the binominals are presented in full (with the conjunction between the content words); in Part 2 of the study, the content words from the same binomials are presented *without* the conjunction.

The results show that, in Part 1, the N400 amplitude is smallest in the semantically congruous collocation condition (*knife and fork*) and largest in the semantically incongruous non-collocational condition (*theme and fork*). Interestingly, in the semantically congruous non-collocational condition (*spoon and fork*), the N400 amplitude is smaller than that in the semantically incongruous non-collocational condition (*theme and fork*), but larger than that in the semantically congruous collocation condition (*knife and fork*). This suggests that, as well as being sensitive to semantic violations, the N400 is also sensitive to collocational violations.

Moreover, the results from Part 2 of the experiment reveal that, when the content words from the binomials are presented *without* the conjunction, the difference in ERP response between the collocational and non-collocational conditions is no longer apparent. This suggests that it is the (non-)collocational status of the binomials, rather than the properties of the individual words, which causes the difference in N400 amplitude. I might therefore expect to find that an enlarged N400 is elicited by participants reading the non-collocations in my experiments. However, it is important to note that binomials are a very specific type of collocation, so the results of this study may not be applicable to studies such as mine which use a different type of collocation.

3.4.6.2 The non-explicit approach

In the previous sub-section, I discussed ERP studies which explicitly state that they are focusing on collocation, although many of these studies conceptualize and operationalize collocation differently from how it is conceptualized and operationalized in this thesis. In this sub-section, I discuss ERP studies which state that they are focusing on other aspects of language processing, yet are essentially studying collocation as it is defined in this thesis.

In one pair of studies by Kutas et al. (1984) and Kutas and Hillyard (1984), the focus is on the elicitation of an N400 by words that are semantically appropriate but unexpected in their contexts. In Kutas et al.'s (1984) study, 13 participants are presented with 321 word-by-word sentences. The sentences are taken from Bloom and Fisher's (1980:632) study of the cloze probability of 329 sentences. In Bloom and Fisher's (1980) study, 100 participants provide the terminal word to an incomplete sentence by using the first semantically appropriate word that comes to their mind as being a likely ending to that sentence. An example of a sentence with a high cloze probability is *She mailed the letter without a* <u>*stamp*</u>; an example of a sentence with a low cloze probability is *The bill was due at the end of the* <u>*hour*</u>. Kutas et al. (1984) find that an N400 is elicited in response to sentence-final words with a low cloze probability, showing that the N400 is related to expectancy and is therefore not elicited solely in response to semantic violations. However, the amplitude of the N400 is larger in response to semantic violations compared to words with a low cloze probability (Kutas et al. 1984).

In Kutas and Hillyard's (1984) study, the same 321 sentences were re-used from Bloom and Fisher's (1980) study, and the ERP response to the terminal word was compared in sentences with high, medium, and low cloze probabilities. Moreover, in this study, contextual constraint was introduced as an additional factor. Sentences with a high contextual constraint are defined as those that lead to highly predictable endings, meaning that the terminal word is primed by the semantics of the preceding words. This notion of contextual constraint was therefore operationalized in a non-empirical sense, whereby the researchers presumably used intuition to determine the level of contextual constraint of the sentence; by contrast, cloze probability was defined empirically as "the proportion of subjects using that word to complete a particular sentence" (Kutas & Hillyard 1984:161). The interaction of cloze probability and contextual constraint in the stimuli can be seen in the examples given in table 3.1.

Table 3.1: Examples of experimental sentences varying in degree of contextual constraint and cloze probability (table adapted from Kutas & Hillyard 1984:161)

Contextual Constraint	**Cloze Probability**	**Experimental sentence**
High	High	*He mailed the letter without a stamp.*
High	Low	*The bill was due at the end of the hour.*
Medium	High	*She locked the valuables in the safe.*
Medium	Medium	*Too many men are out of jobs.*
Medium	Low	*The dog chased our cat up the ladder.*
Low	High	*There was nothing wrong with the car.*
Low	Low	*He was soothed by the gentle wind.*

The results show that the amplitude of the N400 is inversely proportional to cloze probability, thus replicating the findings from Kutas et al. (1984). However, the amplitude of the N400 does not vary significantly as a result of differences in contextual constraint.

Interestingly, although the aforementioned studies focus on semantic expectancy, it could be argued that the sentences in Kutas and Hillyard's (1984) study which have a medium or low contextual constraint but a high cloze probability are essentially examples of collocation. Indeed, since these sentences have a low contextual constraint but are still likely to attract one particular word over another, this is not that different from how collocation is being operationalized in this thesis, i.e. as lexical associations which are not (necessarily) motivated by syntax/semantics. Therefore, since the sentences with a low cloze probability elicit an N400 in this study, it is plausible to hypothesize that the sentences with non-collocational bigrams might elicit an N400 in my experiments.

Some other studies which essentially seem to be investigating collocation state that they are focusing on the associative links between words. This notion of *association* has been variably defined in the psychology literature, with some researchers equating association with collocation. For instance, Molinaro et al. (2013:122) define "associative relations" as "links between items [which] are developed because their lexical items frequently co-occur in language". Furthermore, as mentioned in Chapter 2 (section 2.4.2), the term *associative priming* has traditionally been used in psychology to refer to words that frequently co-occur (e.g. Moss et al. 1994:413). However, although Rhodes and Donaldson (2008) do not explicitly state how they are defining association, they cite Postman and Keppel (1970) who discuss associations in terms of stimulus-response relationships between words/concepts. Moreover, Rhodes and Donaldson (2008:50) argue that "the presence of semantic relationships has often been confounded with associations, where one word calls to mind another based on free association". This does not imply that there is a syntagmatic (i.e. linear) relationship between associative words, thereby suggesting that Rhodes and Donaldson (2008) do not equate association with collocation.

In one study which attempts to dissociate semantic relationships from associative relationships, Rhodes and Donaldson (2008) ask 32 participants to read four sets of word pairs, the characteristics of which are summarised in table 3.2. In this study, words are considered to have a semantic relationship if they share “categorical or functional features”. This is therefore equivalent to the linguistic notion of paradigmatic relations, whereby there is a substitution relationship between elements of the same category (see section 2.4.2).

Table 3.2: Stimuli types used by Rhodes and Donaldson (2008:52)

	Semantic relationship between words?	**Associative relationship between words?**	**Example word pairs**
1st type of word pair	No	No	*beard-tower* *alarm-cloud*
2nd type of word pair	Yes	No	*cereal-bread* *violin-guitar*
3rd type of word pair	No	Yes	*traffic-jam* *fountain-pen*
4th type of word pair	Yes	Yes	*lemon-orange* *brother-sister*

Rhodes and Donaldson (2008) compare the ERPs elicited by the four different types of word pairs. The results show that an N400 is elicited by all types of word pair except the word pair that has a semantic relationship but not an associative relationship (i.e. the 2nd type of word pair in table 3.2). Rhodes and Donaldson (2008:56) therefore conclude that “it is the associations between words, rather than the semantic relationship between them, that is critical for the processing of meaning”. However, although Rhodes and Donaldson (2008) seem to be defining associations in terms of stimulus-response relationships between words/concepts, rather than as collocational links between words, I would argue that the word pairs used in this study that have associative relationships (i.e. the 3rd and 4th types of word pair) are essentially strong collocations. This makes sense as, while the words with semantic relationships have a paradigmatic relationship, the words with associative relationships have a syntagmatic relationship, as the link between the words is linear rather than categorical. This is interesting because, in their own definition of associations provided earlier, Rhodes and Donaldson (2008:50) give no indication that associative relations constitute syntagmatic links between words.

Since the strong collocations in this study elicit an N400, this could weigh against the previously mentioned hypothesis that the non-collocational bigrams in my experiments might elicit an N400. However, Rhodes and Donaldson (2008) use an entirely different task involving words rather than sentences. In this task, since an N400 is *not* elicited when there is a semantic relationship in the absence of an associative relationship (i.e. the 2nd type of word pair), but an N400 *is* elicited when there is neither a semantic or associative relationship, it seems that the elicitation of the N400 can be attributed to the participants reading words that they would not expect to find together. Moreover, since the 3rd and 4th types of word pair which I have suggested are collocations also elicit an N400, it could be the case that the participants are not expecting the words to be related syntagmatically. Thus, in this study, the N400 might be an electrophysiological marker of expectancy rather than association. This is plausible because it is well-known that ERP components can vary dramatically in response to different task demands (Brandeis and Lehmann 1986:157). Therefore, if this is the case, the results of this study do not necessarily undermine the previously mentioned hypothesis that the non-collocational bigrams in my experiments might elicit an N400.

In a much more recent study, Lau et al. (2016) aim to find out the extent to which the N400 is affected by semantic incongruity, independently of association strength and contextual predictability (the link between predictability and the N400 is discussed at length in the following sub-section). They do this with a series of three ERP experiments, using adjective-noun bigrams categorized in the following way:

Table 3.3: Stimuli types used by Lau et al. (2016:4)

	Congruous predictable	**Congruous unpredictable**
Predictability manipulation	*runny nose* *mashed potato*	*dainty nose* *shredded potato*
	Congruous unpredictable	**Incongruous unpredictable**
Congruity manipulation	*yellow bag* *healthy cat*	*innocent bag* *empty cat*

However, looking at the examples given in table 3.3, it is clear that the bigrams categorized as ‘congruous predictable’ could equally be described as collocations, while the bigrams categorized as ‘congruous unpredictable’ could equally be described as non-collocations.

The results show that reading an incongruous bigram elicits a larger N400 than reading a congruous bigram when controlling for predictability, and reading an unpredictable bigram elicits a larger N400 than reading a predictable bigram when controlling for semantic congruity. Since the predictable bigrams in this study are essentially strong collocations, while the unpredictable bigrams are essentially non-collocations, the results of this study show that there is a neurophysiological difference in the way that the brain processes collocational adjective-noun bigrams compared to non-collocational adjective noun bigrams, thereby suggesting that the phenomenon of collocation can be seen as having psychological validity. In this way, the Lau et al. (2016) study seems to achieve the aim of this thesis.

Indeed, the Lau et al. (2016) experiments are very similar to the experiments presented in this thesis, not only because both sets of experiments compare the ERP response to reading collocational and non-collocational adjective-noun bigrams, but also because the adjective-noun bigrams in both sets of experiments are extracted from a corpus according to their transition probabilities. Specifically, while I extract collocational adjective-noun bigrams with a minimum transition probability of 0.01011 from the BNC1994 (see Chapter 4, section 4.2.1), Lau et al. (2016) extract collocational adjective-noun bigrams with a minimum transition probability of 0.5 from the Corpus of Contemporary American English (COCA; Davies 2008).

However, there are key differences between the two sets of experiments. One difference is that Lau et al. (2016) vary the adjective in each condition and keep the noun constant (e.g. *mashed potato* vs. *shredded potato*), whereas I vary the noun in each condition and keep the adjective constant (e.g. *clinical trials* vs. *clinical devices*). Yet two more important differences relate to the corpus frequencies of the experimental bigrams, and whether the bigrams are presented in isolation or embedded into full sentences. Lau et al. (2016:5) justify their decisions in relation to these points in the following way:

> We note that in the current design, the bigram frequency of the high probability condition was much higher than in the control condition. This was not a primary concern here because the probability comparison in the current experiment was mainly aimed at replicating the effects of probability that have been observed in sentence paradigms, where an analogous confound between lexical probability and probability of the overall event being described also holds.

However, in trying to avoid this confound, and despite the overall aim of their study being to unpick the confounds that are commonly seen in the N400 literature (namely the confounds between semantic congruity and contextual predictability, and semantic congruity and semantic association), Lau et al. (2016) inadvertently create a confound between collocationality and frequency. This can be seen in the way that their bigram frequencies are higher in the predictable condition compared to the unpredictable condition, making it impossible to discern whether or not the larger N400 in the unpredictable condition can be attributed to predictability or frequency. Confounding collocationality with frequency is particularly problematic in ERP studies because there exists reliable evidence to show that the N400 is modulated by word frequency, with low frequency words eliciting a larger N400 than high frequency words (Davenport & Coulson 2011). In my experiments, I therefore try to ensure that the nouns in each condition are matched for frequency; where an exact match is not possible, I ensure that it is the non-collocational noun which has the highest frequency, so that the expected increase in N400 amplitude in the non-collocational condition cannot be attributed to the frequency of the noun (see Chapter 4, section 4.2.1). Finally, I also choose to embed my bigrams into sentences, as I want them to be read in naturalistic reading contexts (or as naturalistic as is possible in the context of an ERP experiment). This is discussed in Chapter 4 (section 4.2.2).

3.4.7 ERP studies of predictability

In sections 3.4.3 through to 3.4.5 I established that the N400 and P600 are sensitive to syntactic and semantic processing; but, in addition to this, the N400 and P600 have also been shown to be sensitive to predictability. In a study by Laurent et al. (2006), 30 native speakers of French are asked to read 240 French sentences. Half of the sentences are weakly salient non-idiomatic expressions and half of the sentences are idioms which vary in their level of semantic saliency. In this study, salient meanings are defined as “coded meanings foremost on our mind due to conventionality, frequency, familiarity, or prototypicality” (Laurent et al. 2006:151). Saliency is therefore essentially equated with conventionality, and the idioms that are considered to be highly salient are those that “enjoy a high degree of entrenchment or fixedness” (Laurent et al. 2006:153). The results of this study show that the amplitude of the N400 and P600 is significantly smaller for the last word of highly salient (i.e. conventional) idioms compared to the last word of weakly salient (i.e. unconventional) expressions. Thus, this shows that words which are less predictable/probable elicit a larger N400 and P600.

Similarly, in a more recent study by Davenport and Coulson (2011), 18 participants read three types of sentences varying in cloze probability. The results mirror those of Laurent et al. (2006) by again showing that the amplitude of the N400 and P600 is smaller for sentences with a high cloze probability compared to sentences with a low cloze probability. Taken together, the results of these studies suggest that the N400 and the P600 are both elicited in response to reading words that are unpredictable even though they are semantically and grammatically appropriate. I therefore hypothesize that the N400 and the P600 will be elicited by the second word of the non-collocational bigrams in my experiments.

Another component that is well known for its association with predictability is the P300. The component that later became known as the P300, or P3, was first discovered by Sutton et al. (1965) and has since become one of the most widely studied ERP components. As mentioned in section 3.4.2, psychologists are still uncertain about how the P300 relates to cognition (Polich 2007:2137), but it has been broadly associated with attention, memory, and the processing of probability sequences (Polich 2007:2132, 2137).

Donchin and Coles (1988:363) state that "the rarer the event, the more likely it is to elicit a P300", and that "the amplitude of the P300 elicited by such rare events will be inversely related to the probability of the event". However, it is important to note that the P300 is typically only considered to be sensitive to *task-relevant* probability (Donchin & Coles 1988:367). This means that, in a task which involves reading individual orthographic letters, the letter <E> will elicit a P300 if it is rare in the context of this task even though the letter <E> is the most frequent letter in the English language (Luck 2014a:97). This suggests that the P300 is unlikely to be consistently sensitive to measures of collocation strength such as transition probability or mutual information, as these measures are based on how the words pattern throughout the whole corpus, rather than in the context of the experimental task.

In recent years, the P300 has come to be recognized as a family of ERP components rather than as an individual component, with each sub-component having a distinct scalp distribution and reflecting slightly different (though overlapping and highly contested) cognitive processes. The P300 sub-components include the P3a, the P3b, the novelty P300, and the go/no-go P300 (Polich 2007:2134). Polich (2007:2134) considers the latter three sub-components to be variants of the same component, as their scalp topography tends to vary only in response to different task demands. Thus, the two main sub-components are the P3a and the P3b.

The P3a has a frontal scalp distribution and it tends to peak earlier than the P3b, which has a temporal-parietal scalp distribution (Polich 2012:2137). Both subcomponents are traditionally elicited during what is known as an *auditory oddball task* (Polich 2012:160). In this commonly used experimental paradigm, a target auditory stimulus (e.g. a tone with a fundamental frequency of 1000 Hz) referred to as the ‘oddball’ is randomly interspersed with a repetitive non-target auditory stimulus (e.g. a tone with a fundamental frequency of 500 Hz) which has a lower probability of occurrence (e.g. 0.8 vs. 0.2) in the context of the experiment (Polich 2012:160). Participants are instructed to focus on the low probability non-target items (Polich 2012:160). The act of discriminating the low probability target items from the high probability non-target items elicits a P3b (Polich 2012:160), showing that the P3b is elicited in response to attending to low probability items. Moreover, if participants are instructed to simultaneously engage in a secondary task, the amplitude of the P3b decreases in a way that is inversely proportional to task difficulty (Kramer et al. 1985; Polich 2012:163). This shows that the P3b is modulated by attentional demands. By contrast, the P3a is not modulated by attentional demands as it is elicited when participants listen to the stimuli passively, without being instructed to engage in the task of attending to the low probability items (Polich 2007:2133).

Ignoring for now the claim that the P300 is only elicited in response to task-relevant stimuli (Donchin & Coles 1988:367), it is plausible to suggest that the P300 might be elicited in my experiments in response to participants reading the second word of the non-collocation bigrams. This is because, since the non-collocational bigrams are deviant in the sense that they have a low transition probability, they can perhaps be seen as ‘oddballs’ that are not frequently encountered. Moreover, since the elicitation of the P3b requires the participant to consciously attend to the low probability stimuli, whereas the elicitation of the P3a does not require this conscious attention, I am perhaps more likely to find a P3a than a P3b. This is because the participants in my experiments are not given any indication of the presence of the non-collocational bigrams, or indeed the focus on the probabilities between words.

Interestingly, the P600 is often considered to be somehow related to the P3b (Coulson et al. 1998a:655). This is because the two components share a similar scalp distribution (Coulson et al. 1998b:31) and, although the P600 is traditionally associated with syntactic processing (see section 3.3.2.3), both the P3b and the P600 have been shown to be sensitive to probability (Coulson et al. 1998a). In a study by Coulson et al. (1998b), 16 native speakers of English were asked to read 200 English sentences. The aim of the experiment was to elicit both the P600 and the P3b in order to test the relationship between the two components. In order to elicit the P600, sentences with morphosyntactic violations were presented as well as matched sentences containing no morphosyntactic violations; in order to elicit the P3b, the 200 sentences were split across two blocks of 100 sentences, and the proportion of grammatical to ungrammatical sentences varied across each block.

The results show that the improbable stimuli elicit both a P600 and a P3b (Coulson et al. 1998b). Interestingly, Coulson et al. (1998b:45) interpret the P600 results as not reflecting ungrammaticality directly but, rather, as reflecting “the relatively rare linguistic event of ungrammaticality”. They thereby categorize the P600 as being “a member of the P300 family of components” (Coulson et al. 1998b:45). This could be important to keep in mind when interpreting the P600 results of my own experiments.

A summary of the key ERP studies of collocation and predictability introduced in sections 3.4.6 and 3.4.7 is given in table 3.4. I will return to these studies in the discussion in Chapter 9.

Table 3.4: Summary of previous ERP studies of collocation and predictability

Author (s)	**Date of publication**	**Number of participants**	**Language studied**	**Number of trials per condition**	**Collocation/predictability measure**	**Observed effects**
Laurent et al.	2006	30	French	20	The authors do not state how the stimuli were selected.	N400 and P600 amplitude is larger in response to reading the last word of an unconventional expression compared to the last word of a conventional idiom.
Rhodes & Donaldson	2008	32	English	102	Words extracted from MRC Psycholinguistic database, according to their level of "association". It is unclear which database parameters were used.	N400 is enlarged in response to reading words in a word pair which are not likely to occur together.
Molinaro & Carreiras	2010	36	Spanish	56	Corpus-derived idioms and clichés; cloze probability	P300 is enlarged in response to reading the final word of collocations with a high cloze probability compared to the final word of collocations with a low cloze probability.
Vespignani et al.	2010	50	Italian	29	Predictability of idioms is determined by the use of pre-experiment questionnaires	P300 is enlarged in response to reading the final word of collocations with a high cloze probability compared to the final word of collocations with a low cloze probability.

Tremblay & Baayen	2010	10	English	215	Frequency of string ABCD divided by frequency of four word sequences containing ABC minus the frequency of that string	N1 and P1 amplitude is larger in response to reading low probability sequences compared to high probability sequences
Davenport & Coulson	2011	18	English	80	Cloze probability	N400 and P600 amplitude is larger in response to reading sentence-final words in sentences with a low cloze probability compared to sentence-final words in sentences with a high cloze probability.
Molinaro et al.	2013	36	Spanish	88	Corpus-derived idioms and clichés; cloze probability	N1 amplitude is larger in response to reading the last word of a non-fixed multi-word expression compared to the last word of a fixed multi-word expression.
Lau et al.	2016	28	English	30	The authors do not state how probability was calculated from the corpus data.	N400 amplitude is larger in response to reading an unpredictable adjective-noun bigram compared to predictable adjective-noun bigrams.
Siyanova-Chanturia et al.	2017	48	English	40	Binomial collocations and their frequencies extracted from the BNC1994, presumably based on intuition.	N400 amplitude is smaller is response to reading binomial collocations compared to matched non-collocations.

3.4.8 Advantages and disadvantages of the ERP technique

The major advantage of the ERP technique is its high temporal resolution (Brandeis and Lehmann 1986:152). Brandeis and Lehmann (1986:151) state that the ERP technique is "the only noninvasive method which resolves the dynamic patterns of events in the human brain down to the millisecond range", and Kutas and Van Petten (1994:93) add that the ERP technique "is as close to immediate and on-line as is now technically possible". Furthermore, Rhodes and Donaldson (2008:51) argue that "ERPs (derived from EEG) are an ideal method of studying the processes underlying language comprehension because they provide a real-time record of neural activity, allowing processes to be identified and dissociated on the basis of fine-grained temporal information". The accurate temporal information provided by EEG data is an advantage that is shared by behavioural techniques such as self-paced reading; yet the ERP technique has the added advantage of providing a more direct measure of brain activity (Kutas & Van Petten 1994:92).

The high temporal resolution provided by the ERP technique is in contrast with neuroimaging techniques such as *functional magnetic resonance imaging* (fMRI), which provide very low temporal resolution but high spatial resolution (Swaab et al. 2012:22). Thus, a disadvantage of the ERP technique is that it does not provide the high spatial resolution provided by fMRI (Swaab et al. 2012:22). Although ERP components are partly characterised by their scalp distribution (Kappenman & Luck 2012:16), the electrical activity associated with that ERP component may actually be generated in a different part of the brain. Luck (2014a:48) explains that it is mathematically impossible to accurately disentangle the source of the electrical activity by only knowing the scalp location. Therefore, a major disadvantage of the ERP technique is that it cannot be used to draw any conclusions about the location of mental processes in the brain (Hillyard 2000:25). This is known as the *inverse problem* (Swaab et al. 2012:418) and it will be explained further in Chapter 4.

Another advantage of the ERP technique is that it enables psycholinguistic data to be gathered without having to impose many additional task demands (Kutas & Van Petten 1994:94). While, for instance, self-paced reading experiments require participants to press a button after reading every word or sentence, the ERP technique requires participants to simply read the words that are being presented to them (Kutas & Van Petten 1994:95). However, one disadvantage that the ERP technique shares with self-paced reading is that presenting sentences word-by-word does not reflect "normal" reading conditions (Luck 2014a:32). Normal reading does not involve reading words one at a time (Tremblay et al. 2011:581), and normal reading involves skipping words and making *regressions* in which the eyes move back to words that have previously been read (Rayner 1998:375). Moreover, while the typical reading rate under normal reading conditions is around 200 ms per word (Luck 2014a:32), the typical presentation rate used in language-based ERP experiments is 500 ms per word (Swaab et al. 2012:32). This longer presentation rate is needed in order to minimise the overlap between the ongoing brain response to the previous word and the brain response to the next word that is presented (Kutas & Van Petten 1994:99). However, as with self-paced reading experiments, slowing the pace of reading gives "the reader extra time that could potentially be used to engage processes that do not normally take place" (Rayner et al. 2012:221).

As already mentioned in section 3.4.1, another limitation of the ERP technique is that it provides only a partial measure of brain activity (Kiloh et al. 1981:46). This is due to the fact that the electrical activity that is detected by scalp electrodes mainly arises from one type of cell, namely cortical pyramidal cells (Kutas & Van Petten 1994:87; Kappenman &Luck 2012:5). Furthermore, a large number of these cells need to be spatially aligned and simultaneously active in order for their electrical activity to be detected at the scalp (Kappenman & Luck 2012:3). This means that, even if no experimental effect is found with the ERP technique, the null hypothesis cannot be rejected because it is possible that the neural activity that would give rise to the experimental effect is just not detectable by scalp electrodes (Otten & Rugg 2005:10; Kappenman & Luck 2012:3).

Further limitations relate to the difficulty in defining and identifying ERP components. In section 3.4.2, I provisionally defined an ERP component as “a scalp-recorded voltage change that reflects a specific neural or psychological process” (Kappenman & Luck 2012:4). However, Kappenman and Luck (2014:4) themselves point out that the words “reflect” and “process” refer to “loose concepts without clear definitions”; thus, “it is no easy task to find a simple, concise, and widely accepted definition of the term ERP component”. Some researchers focus primarily on the function of a component (i.e. the neural or psychological processes that the component reflects) while other researchers focus primarily on the scalp region in which the component is found (Coulson et al. 1998:658). This is problematic because, if the same psychological process is observed across different scalp regions, researchers will differ as to whether or not they consider this to be a single component (Coulson et al. 1998:658).

Kappenman and Luck (2012:4) note that the term ERP component “is rarely defined or conceptualized beyond the peaks in the observed ERP waveform”. The focus on peaks can partially be attributed to their visual salience (Kappenman & Luck 2012:4). Moreover, peak measures were used in early ERP studies when modern analysis techniques were not yet available, and this tradition has continued despite the improvement in analytical software and techniques (Donchin & Heffley 1978:557). However, it is highly problematic to equate the notion of an ERP component with a peak in a waveform because “peaks do not map onto distinct ERP components in a simple one-to-one manner” (Kappenman & Luck 2012:8). Each peak is typically associated with more than one component because, since the electrical activity related to each cognitive process often persists for hundreds of milliseconds, different ERP components overlap as different cognitive processes are active within the same latency range

(Kappenman & Luck 2012:8). This means that “the peaks in the observed ERP waveforms can provide a misleading representation of the underlying components” (Luck 2014a:54).

Even if peaks and components *did* have a one-to-one relationship, the amplitude and latency of the peak would still not have much theoretical significance because, as pointed out by Luck (2014a:54), “theories of cognition and brain processes do not usually say much about when a process peaks. Instead, these theories usually focus on the onset of a process, the duration of the process, or the integrated activity over time”. Although most ERP studies still report peak measures, other methods are available for quantifying the amplitudes and latencies of ERP components which focus more on the onset or duration of the component (Kappenman & Luck 2012:23). In my experiments, instead of reporting peak measures, I will report the mean amplitude between two time points and I will also report the point at which the waveforms in each condition begin to diverge. The advantages of this combined approach will be discussed in Chapter 4 (section 4.6.7).

Even when researchers avoid focusing on peak measures, they still face the *superposition problem*, i.e. the fact that “it is very challenging to isolate and measure the internal underlying components on the basis of the data that you actually record from the scalp” (Luck 2014a:35). The phrase *underlying component* refers to the internal brain activity that is not directly observable in ERP waveforms (Kappenman & Luck 2012:9). Kappenman and Luck (2012:9) explain that “[w]hen multiple ERP components are simultaneously generated in different brain areas, the voltages from these components sum together”, and so “[a]n infinite number of underlying components could sum together to give rise to a given ERP waveform”. This is problematic for my experiments because, if I find what seems to be an increase in the amplitude of the N400 in the non-collocational bigram condition, it will be unclear whether this is truly an N400 effect or whether it is actually a P300 effect (Kappenman & Luck 2014:3). After all, even the polarity of a component at a particular electrode site can vary across different participants depending on the shape and structure of their cortex (Kappenman & Luck 2012:14). The superposition problem therefore has the potential to drastically alter the conclusions that are drawn from an experiment.

In order to minimise the superposition problem, it is necessary to take a holistic approach to identifying ERP components. This involves combining evidence from all of the different characteristics of a component, including polarity, latency, amplitude, scalp distribution, and response to experimental manipulations (Kappenman & Luck 2012:16). Kappenman and Luck (2012:16) point out that "[w]hen this converging evidence approach is taken, it is important to consider both the strength of the evidence that a given component has a specific property and the degree to which other components might have that same property". Alternatively, in some experiments it is possible to remove the reliance on identifying specific ERP components by using a component-independent design which focuses purely on the time course of the processes (Kappenman & Luck 2012:17). In my experiments, I combine the advantages of a component-based design with the advantages of a component-independent design. This will be discussed in detail in Chapter 4 (section 4.6.7).

3.5 Chapter summary and conclusion

In this chapter I have reviewed prior work on the psychology of collocation, covering experimental methods as diverse as self-paced reading, eye-tracking, and EEG. I have outlined the advantages and disadvantages of using each of these methods, arguing that EEG/ERP studies provide a more direct measure of cognition than is possible with self-paced reading and eye-tracking experiments. Moreover, I have summarized a large body of language-related ERP research, as well as the smaller body of ERP studies which address collocation (although it is often not explicitly acknowledged that collocation is being studied).

From reviewing prior work on the psychology of collocation, I have concluded that psycholinguistic experimental methods such as self-paced reading and eye-tracking rely on proxy measures of cognitive load, and therefore only allow us to speculate about what might be happening in the brain during language processing. Collocation has also been investigated using the ERP technique, which provides a more direct measure of brain activity than what is obtained from psycholinguistic experiments. However, previous ERP studies tend to take a very narrow view of collocation by focusing on idioms or other fixed expressions, or they have serious methodological issues which undermine their results. There is therefore a clear gap in the literature for ERP studies of corpus-based collocations. In the following chapter, I outline the methodological steps and decisions that are common to each of the four ERP experiments of corpus-based collocations presented in this thesis.

CHAPTER 4: METHODOLOGY

4.1 Overview of the chapter

The aim of this chapter is to describe the methodological steps and provide justification for the methodological design decisions that are central to all of the ERP studies presented in this thesis. I start by describing the selection of stimuli in section 4.2 and the presentation of stimuli in section 4.3. I then describe the electrode placement in section 4.4 before outlining the experimental setup and procedure in section 4.5. Then, in section 4.6, I describe each pre-processing, processing, and analysis step that is necessary in ERP research, before providing a chapter summary and conclusion in section 4.7.

The experimental procedure is reported in line with the recent guidelines specified by Keil et al. (2013). This report provides recommendations and publication guidelines for EEG and MEG studies, and "is the result of the collaborative effort of a committee appointed by the Society for Psychophysiological Research" (Keil et al. 2013:1). Following these guidelines therefore helps to ensure that my ERP experiments are conducted in a way that is up to the standards of published work.

4.2 Selection of stimuli

4.2.1 Selection of bigrams

In this thesis, collocations are operationalized in terms of *bigrams*, i.e. adjacent word pairs. This is because it is more feasible to conduct a controlled experiment with a collocation that contains a single transition than with a collocation that contains multiple transitions. Furthermore, by operationalizing collocation in terms of bigrams, I avoid the theoretical assumptions that are carried by some alternative ways of operationalizing collocation. For example, operationalizing collocation in terms of formulaic language carries Wray's (2002:15) assumption that there are two linguistic processing systems and that there is therefore a qualitative difference in the way that formulaic sequences and non-formulaic sequences are processed. Although I do agree that non-compositional idioms must be processed holistically, I do not necessarily agree that all compositional formulaic sequences are processed holistically, and that the processing mechanism used for compositional formulaic sequences is qualitatively different from the processing mechanism used for non-formulaic sequences. Instead, I am assuming that both formulaic and non-formulaic sequences (to the extent that they can be dichotomized) are processed in terms of transition probabilities. I therefore assume that, in both a theoretical and methodological sense, there is not a qualitative difference but a quantitative difference in the way that collocations and non-collocations are processed.

This is in line with the network view of collocation discussed in Chapter 2 (section 2.8), as it suggests that there are no clear beginning and end points that mark the boundaries between sequences of words that are processed holistically and sequences of words that are not processed holistically. Rather, collocation is considered to be a fluid phenomenon whereby, whenever an individual hears or produces any word, other words that have previously followed that word for that individual will be mentally activated. If word Y frequently follows word X in an individual's language experience, the connections between that word pair will strengthen so that there is an increased probability that word Y will follow word X whenever that individual produces word X. Similarly, there is an increased probability that the individual will expect word Y to follow word X whenever that individual hears word X.

My experimental hypothesis is that, when an individual reads word X followed by a word that does *not* frequently follow word X, their brain will respond in a quantitatively different way compared to when reading word X followed by a word that *does* frequently follow word X. I used two experimental conditions to test this hypothesis. The collocational condition contains *collocational bigrams*, where word Y has a high probability of following word X; the non-collocational condition contains *non-collocational bigrams*, where word Y has a very low probability of following word X.

I extracted the collocational bigrams from the written section of the BNC1994 using BNC*web* (Hoffman & Evert 2006; Hoffman et al. 2008). I chose to use the BNC1994 because it contains 90 million words of written English from a variety of text types and is therefore fairly representative of written British English (Aston and Burnard 1998:5, 31-32). Furthermore, I chose to use only the written section because my experiments involve reading as opposed to listening. This is important because the collocations that are found in writing are "fundamentally different" from those found in speech (Biber 2009:299).

The first step in extracting bigrams from BNC*web* was to use part of speech tags to search for bigrams containing particular grammatical categories. In this thesis I focus solely on adjective-noun bigrams, and extracting these involved using the following search term: _{ADJ} _{N}. The underscores indicate that I was searching for part-of-speech tags as opposed to words, and the simple tags ADJ and N in combination with the braces indicate that I was searching for any word whose simple tags are ADJ and N. This corpus query therefore allowed me to retrieve all instances of any type of adjective followed by any type of noun.

The next step was to *thin* (i.e. reduce) the query results from 6,270,939 to the maximum value of 250,000. The need to do this was due to a software limitation in BNC*web*, which does not allow any functions to be carried out on more than 250,000 concordance lines at any given time. I then performed the *frequency breakdown* function on these 250,000 random instances. This provided me with a downloadable frequency list of adjective-noun bigrams from the written BNC1994. The collocational bigrams in my pilot study were all selected from this list. By contrast, although the non-collocational bigrams contained the same adjective as the matched collocational bigrams, the non-collocational bigrams did not appear in the BNC at all. I manufactured the non-collocational bigrams and ensured that they do *not* occur in the

BNC1994 based on the assumption that absence of a particular bigram from the BNC1994 can serve as a proxy for the identification of non-collocational bigrams. Of course, it is possible that a collocation could be “accidentally absent” (Stefanowitsch 2006:62) from the BNC1994, since any corpus just contains a sample of the language as a whole. Therefore, following Miller (2010), I used my intuition as a native speaker of English as an additional check on whether a particular bigram that is absent from the BNC1994 is or is not a collocation[6].

I chose to manufacture the non-collocational bigrams for a number of reasons. Following Millar (2010), I could have compared native speaker bigrams found in the BNC1994 with bigrams found in a learner corpus. As mentioned in Chapter 3 (section 3.2), the purpose of Millar’s experiments was to investigate how native speakers of English process these learner collocations, which are not considered to be collocational from a native speaker perspective. Across his three self-paced reading experiments, Millar consistently finds that learner bigrams are read significantly more slowly by native speakers than matched collocational bigrams. However, since I am not focusing on non-native language use, there is less justification for using this method of non-collocational bigram selection in my thesis.

Alternatively, I could have used weak collocations that are found in the BNC1994 and compared the brain response to processing these weaker collocations to the brain response to processing the stronger collocations in the collocational condition. As mentioned in Chapter 3 (section 3.2), Hughes and Hardie (forthcoming) compared stronger and weaker bigrams and found that stronger bigrams are processed significantly more quickly than weaker bigrams. However, if I found that the brain as observed via EEG does not respond differently to processing stronger and weaker bigrams, this would not confirm whether or not the brain

---

[6] In retrospect, asking a group of people to rate the bigrams according to their perceived collocationality would have been more reliable than using my own intuition.

responds differently to processing collocational and non-collocational bigrams. Indeed, if there is a neural correlate of collocation, it is more likely to be found when comparing collocational bigrams with non-collocational bigrams than when comparing stronger bigrams with weaker bigrams. However, if I find significant results when comparing collocational bigrams with non-collocational bigrams, this justifies carrying out Experiment 4, which compares stronger and weaker bigrams and therefore explores the sensitivity and gradience of the brain response.

It is important to note that even the non-collocational bigrams are still collocations in a very broad sense. For instance, any adjective-noun bigram constitutes a colligation because it juxtaposes two grammatical categories. Therefore, reading the adjective of an adjective-noun bigram pair would automatically activate the ‘noun’ grammatical category. However, it would only strongly activate the nouns that are also collocates in their own right, i.e. the nouns in the collocational condition as opposed to the non-collocational condition.

Although the non-collocational bigrams are absent from the BNC1994, I had to ensure that they are still semantically plausible. After all, if the non-collocational bigrams are also semantically deviant, we would expect this to automatically cause an N400 brain response. This would create a confound which would make it impossible to disentangle the brain’s response to a semantic error from the brain’s response to a collocational error in the absence of any other semantic or grammatical errors.

When selecting the bigrams, I also had to ensure that the nouns in both conditions were matched for frequency. This is because not only is there psycholinguistic evidence to show that high frequency words are processed more quickly than low frequency words (Conklin & Schmitt 2008:79), there is also evidence from ERP studies to show that the N400 is sensitive to word frequency (Kutas & Dale 1997:222; Swaab et al. 2012:5; Grainger et al. 2012:613). I used the *frequency list* function in BNC*web* to obtain a frequency list of the nouns in the written

BNC1994. I searched for the noun from a potential collocational bigram in this list, and then found a noun with a similar frequency to form a potential non-collocational bigram.

I could not always find nouns with exactly the same frequency; if I only accepted bigrams which were exactly matched for the frequency, I would not have been able find enough to produce my experimental stimuli. However, the biggest difference that I allowed was a difference of 4002, in the case of the bigram pair *classic example/company*. I thought that the large difference was justified in this case because the nouns both had a very high frequency (34,127 for *example*; 38,129 for *company*) compared to the other nouns that I used as part of experimental bigrams.

When the matched nouns did not have exactly the same frequency, I ensured that it was always the noun in the non-collocational bigram that was more frequent than the noun in the collocational bigram. This meant that, if I found a quantitative difference in brain response between the two conditions, the difference could not be attributed to word frequency. Finally, once I had found nouns that were suitably matched for frequency, I searched the whole BNC1994 to ensure that the non-collocational bigrams did not occur in the BNC1994 at all.

As well as being matched for frequency, the nouns were also matched for length (in terms of the number of letters and the number of syllables). However, when I could not find bigrams that were matched for both length and frequency, I prioritized frequency over length. This is because, while there is psycholinguistic evidence for a quantitative difference in the processing of words with different frequencies, “a length effect independent of word frequency has been hard to find” (Conklin & Schmitt 2008:79). Following Biber (2009:291), I also ensured that the collocational bigrams occur in at least 5 different texts. This is because linguistic co-occurrences with a very low level of dispersion are less likely to be representative of how those linguistic items behave in the entire corpus and in the language as a whole (Gries 2008a:406).

In addition, I tried to ensure that I did not choose text from part of a longer collocation that is a stronger unit than the bigram itself. For example, the bigram *regular basis* is part of the longer collocation *on a regular basis*. If I embedded *regular basis* into a sentence without preceding it with *on a*, it might seem out of place in the sentence. The brain response might therefore reflect the unexpected presence of the whole bigram instead of the expected presence of the noun. Similarly, if I embedded the entire collocation *on a regular basis* into a sentence, I would be unable to find a matched bigram with an equal contextual constraint. This notion of contextual constraint will be discussed in the following sub-section.

Due to the small-scale nature of the pilot study, I used 15 bigram pairs and presented them just once to each participant; for Experiments 2 and 3, I used the same 15 bigram pairs but repeated the stimuli lists twice. This is because prior work on large language-related ERP components such as the N400 tends to use a minimum of 30 trials in each condition (Luck 2014a:262) (see Chapter 5, section 5.5.6, and Chapter 6, section 6.3.3, for full justification). After selecting the bigrams, I calculated the *transition probability* of each bigram by dividing the number of times the bigram X-then-Y occurs in the written BNC1994 by the number of times X occurs in the written BNC1994 altogether (McEnery & Hardie 2012:195). The higher the transition probability, the stronger the bigram. In McDonald and Shillcock's (2003:648) eye-tracking experiments (see Chapter 3, section 3.3), the stronger bigrams have a mean transition probability of 0.01011. I therefore used this as the minimum value for the transition probability of my collocational bigrams.

Since the non-collocational bigrams do not occur in the BNC1994 at all, for the purposes of this thesis, I state that they have a transition probability of zero. However, as the non-collocational bigrams are, by definition, semantically plausible word pairs, they cannot actually have a transition probability of zero; rather, they have a transition probability that is lower than can be measured in the written BNC1994.

I chose to use transition probability as a measure of collocation strength because it can be seen as a point-estimate of the psychological transition probability inherent in the network model of language processing. Furthermore, transition probability is a directional measure. This means that it takes into account the asymmetry of bigrams, i.e. the fact that the strength of the association is not equal in both directions. For example, the bigram *of course* is asymmetric because the set of words that are likely to follow *of* is much larger than the set of words that is likely to precede *course* (Gries 2013:144). Gries (2013:144, 148) argues that bidirectional measures such as mutual information "conflate two pieces of information that should probably not be conflated" because they conceal the fact that (in many cases) "the association is in fact only high in one direction". It is particularly important to avoid this conflation in my experiments because, since sentences will be presented to participants word-by-word (see section 4.3), participants will only be able to predict the next word. Therefore, only the forward transition probability as opposed to the backward transition probability is relevant in this context.

4.2.2 Selection of sentences

Once I had selected the bigrams, I embedded them into sentences so that the participants are exposed to the bigrams in a relatively natural reading context, instead of having to read two words that would not normally be encountered outside a wider context. The experimental sentences used for Experiments 1, 2, and 3 are shown in Chapter 5 (table 5.3), while the experimental sentences used for Experiment 4 are shown in Chapter 8 (table 8.3).

In order to find appropriate sentences for the collocational bigrams, I searched for each collocational bigram in the written section of the BNC1994 and then analysed the concordance lines in which they occur. This enabled me to find semantically plausible sentences in which the bigrams have actually occurred. In order to find appropriate sentences for the non-collocational bigrams, I searched for the second word in the bigram so that I could see the typical contexts for that word. I also searched for a bigram that had a similar meaning even though the first word was different from the non-collocational bigram. For example, when identifying plausible sentences for the non-collocational bigram *crucial night* for experiments 1, 2, and 3, I searched for the semantically similar bigram *important night*. This provided an indication of the kinds of sentences that the non-collocational bigrams could plausibly be found in.

Having found appropriate sentences, I edited the sentences (e.g. by removing deictic expressions) to ensure that they were semantically coherent as standalone units. According to Spöttl and McCarthy (2004:197):

> unedited concordance lines often do not provide enough context to enable an item to be used successfully as experimental prompts. Learners [or any experiment participants] are most likely to meet multi-word items in the real world in adequate contexts, not out of context or in impoverished or (to the outside observer) impenetrable contexts.

In some cases the sentences were identical in both conditions except for the experimental bigram. This occurred whenever both of the bigrams were equally semantically plausible within the sentence. For example, this was the case for the collocational bigram *clinical trials* in the experimental sentence *The high cost of* ***clinical trials*** *is delaying progress in medical research*, and the non-collocational bigram *clinical devices* in the experimental sentence *The high cost of* ***clinical devices*** *is delaying progress in medical research*. In most cases, however, only the preceding contexts were identical in both conditions. The preceding contexts needed to be identical in both conditions so that any experimental effects could be attributed to the bigram itself rather than any of the preceding words.

In addition, following Millar (2010:137), I also ensured that the bigrams in each condition were matched for grammatical function. For instance, if the collocational bigram is the subject of the sentence then the matched non-collocational bigram is also the subject of the sentence.

Finally, the preceding contexts needed to create an equally “low contextual constraint” for the bigrams in each condition (Millar 2010:108). If the sentences stems preceding the bigrams have a low contextual constraint, this ensures that the bigrams in each condition are not differently primed by the semantics of the preceding words. For example, Millar (2010:109) states that there is a high contextual constraint for the word *ticket* in the sentence *He got on the train without a* ____ whereas there is a low contextual constraint for the same word in the sentence *He gave her a* ____. This is because the word *ticket* is semantically related to *train*, whereas the sentence stem *He gave her a* does not set up any particular semantic expectations for the noun. In practise, this meant that the experimental bigrams usually needed to be placed very near to the beginning of the experimental sentences so that there were not many preceding words which *could* set up any semantic expectations. This was actually beneficial for another reason, as I will now explain.

I always placed the bigrams near the beginning or middle of the sentence, partly to reduce the possibility of there being a high contextual constraint, but also to avoid *sentence wrap-up effects*. This refers to the phenomenon whereby readers process sentence-final words more slowly than sentence-initial or sentence-medial words because processing a sentence-final word also entails processing and integrating the semantic and syntactic information in the sentence as a whole (Just & Carpenter 1980:345-346).

Once I had edited the sentences I wrote a true/false statement for each sentence, which the participants had to respond to by pressing either the ‘T’ or the ‘F’ key on the keyboard (see section 4.6). Half of the judgements were true; the other half of the judgements were false. Including a language comprehension task such as this encourages the participants to read for meaning, thereby helping to ensure that they process the sentences in the same way as they would if they were reading in a naturalistic environment. Without being encouraged to read for meaning in this way, the participants might just observe the words without actively comprehending their meaning, or without linking the meanings of the individual words together to form a cohesive mental representation of the meaning of the sentence as a whole. Another benefit of making the participants actively engage with the experiment by giving a response is that it encourages the participants to concentrate throughout the duration of the experiment, and it also makes them feel as though they are actively taking part in an experiment (Kutas & Dale 1997:95).

I could have used a between-groups design, where each participant is exposed to only one condition (i.e. either a full set of sentences containing collocational bigrams, or a full set of sentences containing non-collocational bigrams), as this would have allowed me to control for *repetition priming*. This refers to the phenomenon whereby "an initial presentation of a stimulus affects the person's response to the same stimulus when it is presented later" (Goldstein 2011:405). Repetition priming would be problematic in the context of my experiments. For instance, if a participant encounters the non-collocational bigram *crucial night* before the collocational bigram *crucial point*, their brain might then expect the word *night* to follow the word *crucial* on the second encounter of *crucial* even though *crucial point* has a much higher transition probability; conversely, if a participant encounters the collocational bigram *crucial point* before the non-collocational bigram *crucial night*, and their brain subsequently expects the word *point* on the second encounter of the word *crucial*, it would be difficult to disentangle the repetition priming effect from any potential experimental effect.

Although using a between-groups design, where each participant is exposed to just one condition, would be the optimal way of reducing repetition priming, Luck (2014b:5) recommends using a within-subject design for ERP studies, where each participant is exposed to both conditions. This is because, despite the adverse effects caused by repetition priming, the endogenous ERP components that I am focusing on in this thesis are influenced by individual differences between participants (Luck 2014b:5). Examples of individual differences include differences in their level of arousal or differences in the amount of cognitive effort they are exerting in concentrating on the stimuli (Luck 2014b:5). I therefore used a within-subject design but, in order to minimize the effects of repetition priming, I included filler items (listed in Chapter 5, table 5.4) and then ordered the sentences so that the experimental bigrams in each condition were at least 5 sentences apart. This is schematized in figure 4.1.

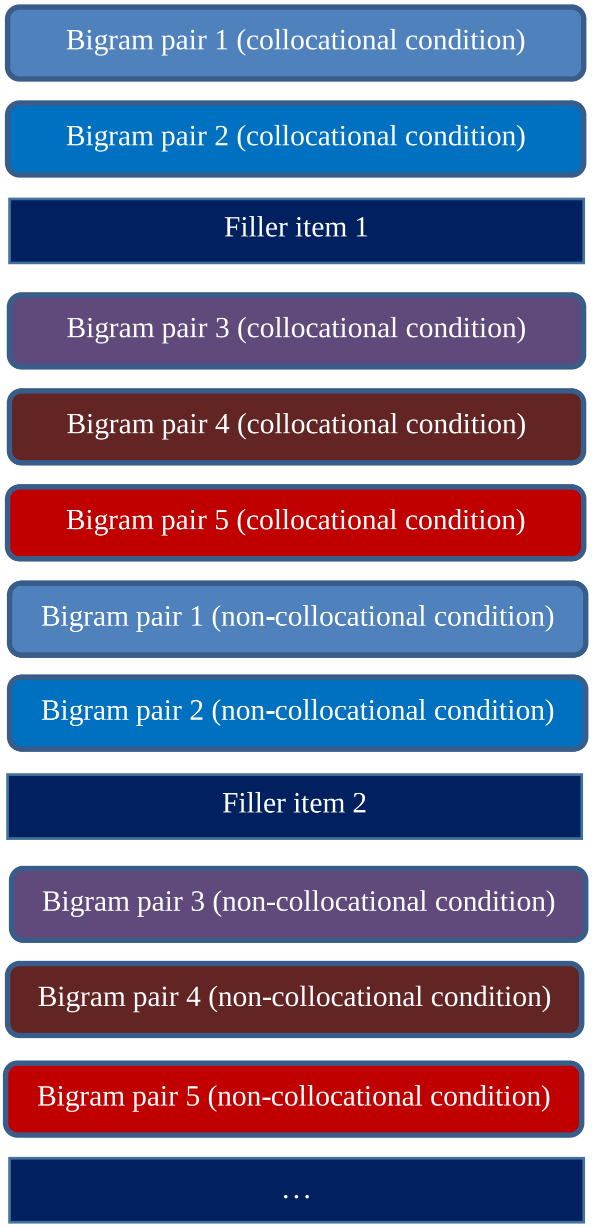
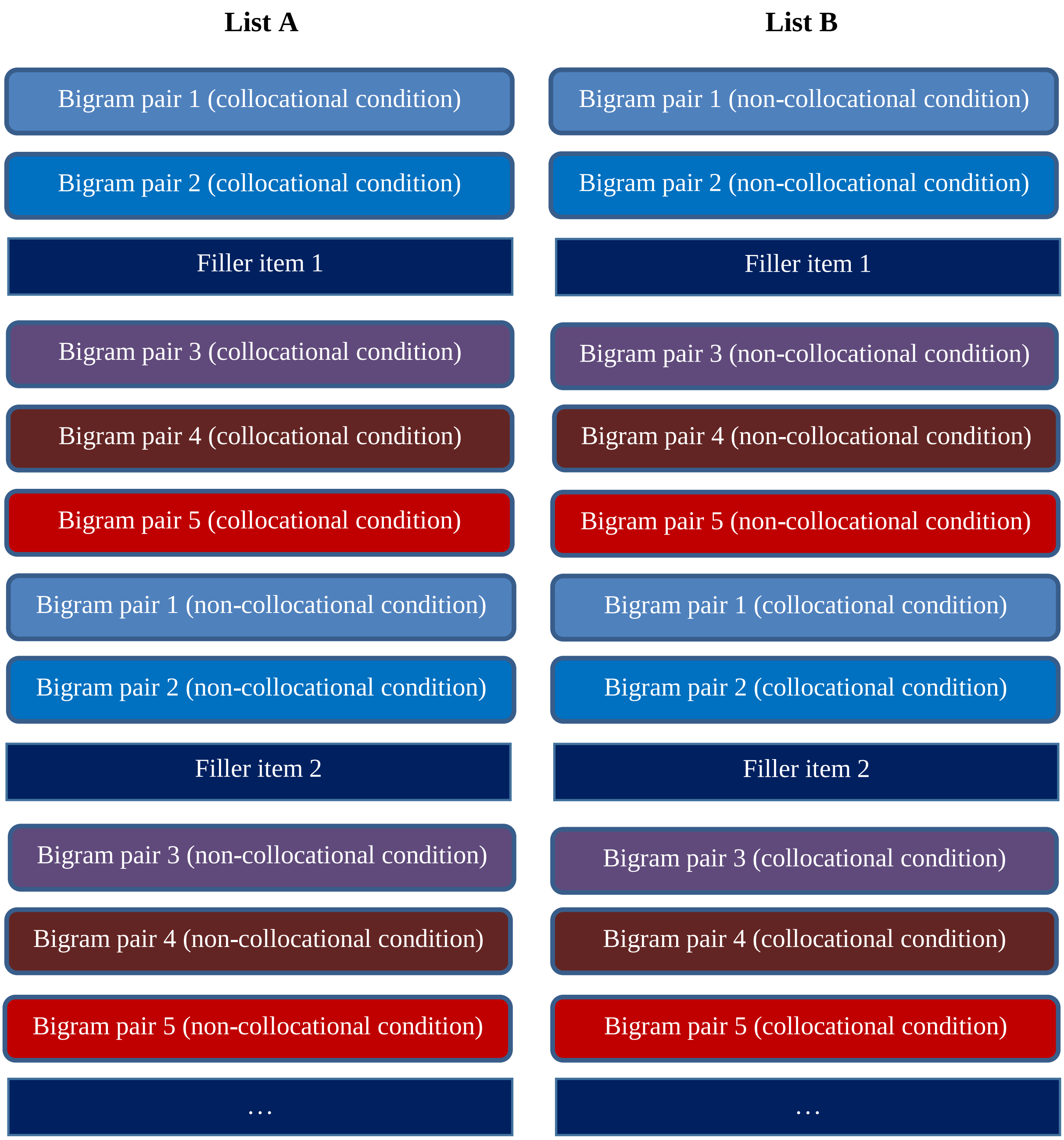


Figure 4.1: Example stimuli list; matched experimental sentences presented at least 5 sentences apart in order to reduce the memory trace and therefore limit repetition priming

Ordering the stimuli in this way helps to limit repetition priming because the intervening stimuli could reduce the participant's memory trace of the bigram that they first encountered.

Closely related to the notion of repetition priming is the concept of *order effects* (Mai et al. 2005). Referring to order effects describes the situation where participation in one condition affects the participants' performance in the other condition. Order effects can be limited by putting the stimuli into differently ordered lists and counterbalancing the stimuli. For the stimuli set used in Experiments 1, 2, and 3 (see Chapter 5, table 5.3), as well as the stimuli set used in Experiment 4 (see Chapter 8), I put the stimuli into 4 differently ordered lists, and randomly assigned each list to an equal number of participants in each experiment. I counterbalanced the lists so that the collocational items came first in List A, and the non-collocational items came first in List B. In List C, half of the collocational items preceded the non-collocational items and, in List D, the other half of the collocational items preceded the non- collocational items. This is schematized in figures 4.2 (List A and B) and 4.3 (Lists C and D). All four lists used in Experiment 1 are reproduced in full in appendix 1, all four lists used in Experiments 2 and 3 are reproduced in full in appendix 2, and all four lists used in Experiment 4 are reproduced in full in appendix 3.



List A	List B
Bigram pair 1 (collocational condition)	Bigram pair 1 (non-collocational condition)
Bigram pair 2 (collocational condition)	Bigram pair 2 (non-collocational condition)
Filler item 1	Filler item 1
Bigram pair 3 (collocational condition)	Bigram pair 3 (non-collocational condition)
Bigram pair 4 (collocational condition)	Bigram pair 4 (non-collocational condition)
Bigram pair 5 (collocational condition)	Bigram pair 5 (non-collocational condition)
Bigram pair 1 (non-collocational condition)	Bigram pair 1 (collocational condition)
Bigram pair 2 (non-collocational condition)	Bigram pair 2 (collocational condition)
Filler item 2	Filler item 2
Bigram pair 3 (non-collocational condition)	Bigram pair 3 (collocational condition)
Bigram pair 4 (non-collocational condition)	Bigram pair 4 (collocational condition)
Bigram pair 5 (non-collocational condition)	Bigram pair 5 (collocational condition)
…	…

Figure 4.2: Ordering of experimental sentences in counterbalanced lists (collocational items come first in List A; non-collocational items come first in List B)

List C	List D
Bigram pair 1 (collocational condition)	Bigram pair 8 (collocational condition)
Bigram pair 2 (collocational condition)	Bigram pair 9 (collocational condition)
Filler item 1	Filler item 1
Bigram pair 3 (collocational condition)	Bigram pair 10 (collocational condition)
Bigram pair 4 (collocational condition)	…
Bigram pair 5 (collocational condition)	Bigram pair 8 (non-collocational condition)
Bigram pair 1 (non-collocational condition)	Bigram pair 9 (non-collocational condition)
…	Bigram pair 1 (non-collocational condition)
Bigram pair 8 (non-collocational condition)	Bigram pair 2 (non-collocational condition)
Bigram pair 9 (non-collocational condition)	Bigram pair 3 (non-collocational condition)
Filler item 7	Filler item 7
Bigram pair 10 (non-collocational condition)	Bigram pair 4 (non-collocational condition)
…	Bigram pair 5 (non-collocational condition)
Bigram pair 8 (collocational condition)	Bigram pair 1 (collocational condition)
Bigram pair 9 (collocational condition)	Bigram pair 2 (collocational condition)
…	…

Figure 4.3: Ordering of experimental sentences in counterbalanced lists (half of the collocational items come first in List C; the other half of the collocational items come first in List D)

The fillers (listed in Chapter 5, table 5.4) were drawn from the set of fillers used by Millar (2010:110); each contains either an adjective-noun bigram or a verb-noun bigram. Since not every sentence contains an adjective-noun bigram, this should make the precise purpose of the study more covert. Furthermore, participants should not be able to distinguish between the experimental sentences and the fillers because the sentences are all approximately the same length. For instance, the experimental sentences used in experiments 1, 2, and 3 have a mean length of 14.97 words while the fillers and practice sentences (which were also drawn from Millar's (2010) set of fillers and are listed in Chapter 5, table 5.4) have a mean length of 15.33 words.

4.3 Presentation of stimuli

The stimuli were presented to participants on a computer screen using the psychology software tool *E-Prime 2.0* (Schneider et al. 2002). Each trial of each experiment consisted of the presentation of a fixation cross (+), followed by a sentence which was presented word-by-word, followed by a true/false statement which was not presented word-by-word. In cognitive neuroscience, the technique of presenting words one at a time in fairly quick succession is known as *rapid serial visual presentation* (RSVP).

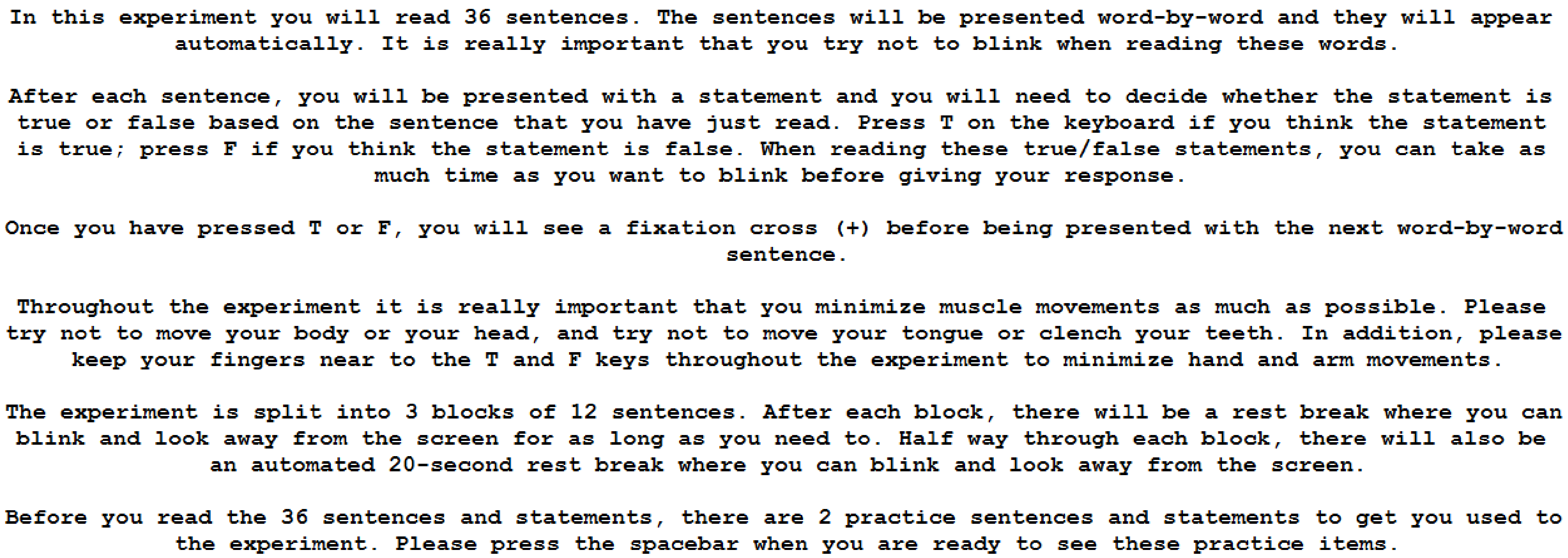
The fixation cross was presented in the centre of the screen for 1000 ms, as it is standard in language-based ERP experiments to present a fixation cross for 1000-2000 ms (Swaab et al. 2012:2), and this was then replaced by the first word of the sentence. The purpose of the fixation cross was to minimize the participants' eye movements. This is very important in EEG experiments because there is a 100 millivolt (mV) potential difference between the cornea and the retina in the eye known as the *corneal-retinal potential*, so any eye movements cause voltage fluctuations that are detected by electrodes (Kiloh et al. 1981:58).

The words in the sentences were each presented in the centre of the screen for 500 ms. As mentioned in Chapter 3 (section 3.4.8), this is the duration that is typically used for words presented in language-based EEG experiments (Swaab et al. 2012:32). It was important to present the words individually, not only to minimize eye movements, but also to ensure that any experimental effects could be time-locked to the second word of each bigram. The true/false statements were not presented word-by-word. This introduces eye movements which could potentially contaminate the data for future sentences. Nevertheless, since the true/false statements are immediately followed by a fixation cross, eye movements should become minimal again before the participants read the experimental bigram in the next sentence.

Participants were instructed to try to avoid blinking while reading the word-by-word sentences, because blinking causes large voltage fluctuations that are detected by electrodes (Kiloh et al. 1981:58). Explicitly instructing participants to avoid blinking is slightly problematic because, as Luck (2014a:211) points out, this means that "you are essentially giving them two tasks to perform at the same time (the task you are explicitly studying and the task of suppressing blinks)". Ochoa and Polich (2000:97) find that this additional task of suppressing blinks decreases the amplitude of the P300, so it perhaps has the potential to modulate other components too. Nevertheless, since participants have the additional task of suppressing blinks in *both* conditions of my experiments, this should not affect whether or not I find a processing difference between conditions.

Although participants were instructed to try to avoid blinking while reading the word-by-word sentences, they were made aware that they could blink while reading and responding to the true/false statements. Since the true/false statements require a response from the participant in order for the experiment to continue, this meant that the participants could take as long as they needed to blink and move their eyes before moving on to the next sentence. In addition, the experiments were organized into blocks varying from 12 experimental sentences in each block in Experiment 4 (see Chapter 8, section 8.3.2) to 15 experimental sentences in each block in Experiments 1, 2, and 3 (see Chapter 6, section 6.3.1). This number of trials was selected to correspond to one block because it takes roughly 5 minutes to complete 12-15 trials, and Luck (2014c:2) states that 5 minutes is the optimal length of time for one block. In between each block, participants were instructed to take a break. This gave the participants a chance to blink and move their eyes and, according to Luck (2014a:194), breaks also help "to keep the subjects more alert and focused on the task". In addition, all experiments were preceded by two practice sentences. This gives the participants a chance to become familiar with the task and practice conscious control of their blinks and eye movements.

To respond to the true/false statements, participants were instructed to press 'T' on the keyboard if they thought that the statement was true in relation to the previous sentence, or 'F' if they thought that the statement was false. These keys are located in close proximity to each other on the keyboard. This is important in minimizing muscle activity because, like eye movements and blinks, muscle activity causes voltage fluctuations that contaminate the EEG recording (Keil et al. 2013:5). To further reduce muscle activity, participants were explicitly instructed to sit as still as possible throughout the experiment and to keep their fingers close to the 'T' and 'F' keys. The main instruction screen presented to participants is shown in figure 4.4.



In this experiment you will read 36 sentences. The sentences will be presented word-by-word and they will appear automatically. It is really important that you try not to blink when reading these words.

After each sentence, you will be presented with a statement and you will need to decide whether the statement is true or false based on the sentence that you have just read. Press T on the keyboard if you think the statement is true; press F if you think the statement is false. When reading these true/false statements, you can take as much time as you want to blink before giving your response.

Once you have pressed T or F, you will see a fixation cross (+) before being presented with the next word-by-word sentence.

Throughout the experiment it is really important that you minimize muscle movements as much as possible. Please try not to move your body or your head, and try not to move your tongue or clench your teeth. In addition, please keep your fingers near to the T and F keys throughout the experiment to minimize hand and arm movements.

The experiment is split into 3 blocks of 12 sentences. After each block, there will be a rest break where you can blink and look away from the screen for as long as you need to. Half way through each block, there will also be an automated 20-second rest break where you can blink and look away from the screen.

Before you read the 36 sentences and statements, there are 2 practice sentences and statements to get you used to the experiment. Please press the spacebar when you are ready to see these practice items.

Figure 4.4: Screenshot of main instruction screen taken from Experiment 4

Finally, it was important to ensure that the words, sentences, and fixation crosses in both conditions were matched for psychophysical properties, including font size, colour, style, and luminosity. This is because differences in the physical characteristics of stimuli can elicit distinct ERP responses (Woodman 2010:2038). Similarly, since the size of the stimuli is affected by the participants’ distance from the computer screen, I ensured that all participants were sitting at the same distance from the screen throughout the whole experiment. I placed the participants’ seat 70 cm from the screen, as this is the minimum distance recommended by Luck (2014c:4). I also raised or lowered the monitor according to the height of the participant, as the muscles of the neck can cause unwanted voltage deflections if the head is not straight (Reis et al. 2014).

4.4 Electrode placement

In EEG experiments, electrodes are traditionally arranged according to the *International 10-20 System*. This system, devised in 1953, places electrodes at 10% and 20% points horizontally and vertically across the scalp. The most recently updated version of this 10-20 system is illustrated in figure 4.5. This updated version is sometimes referred to as the 10-10 system due to the addition of electrodes at intermediate points.

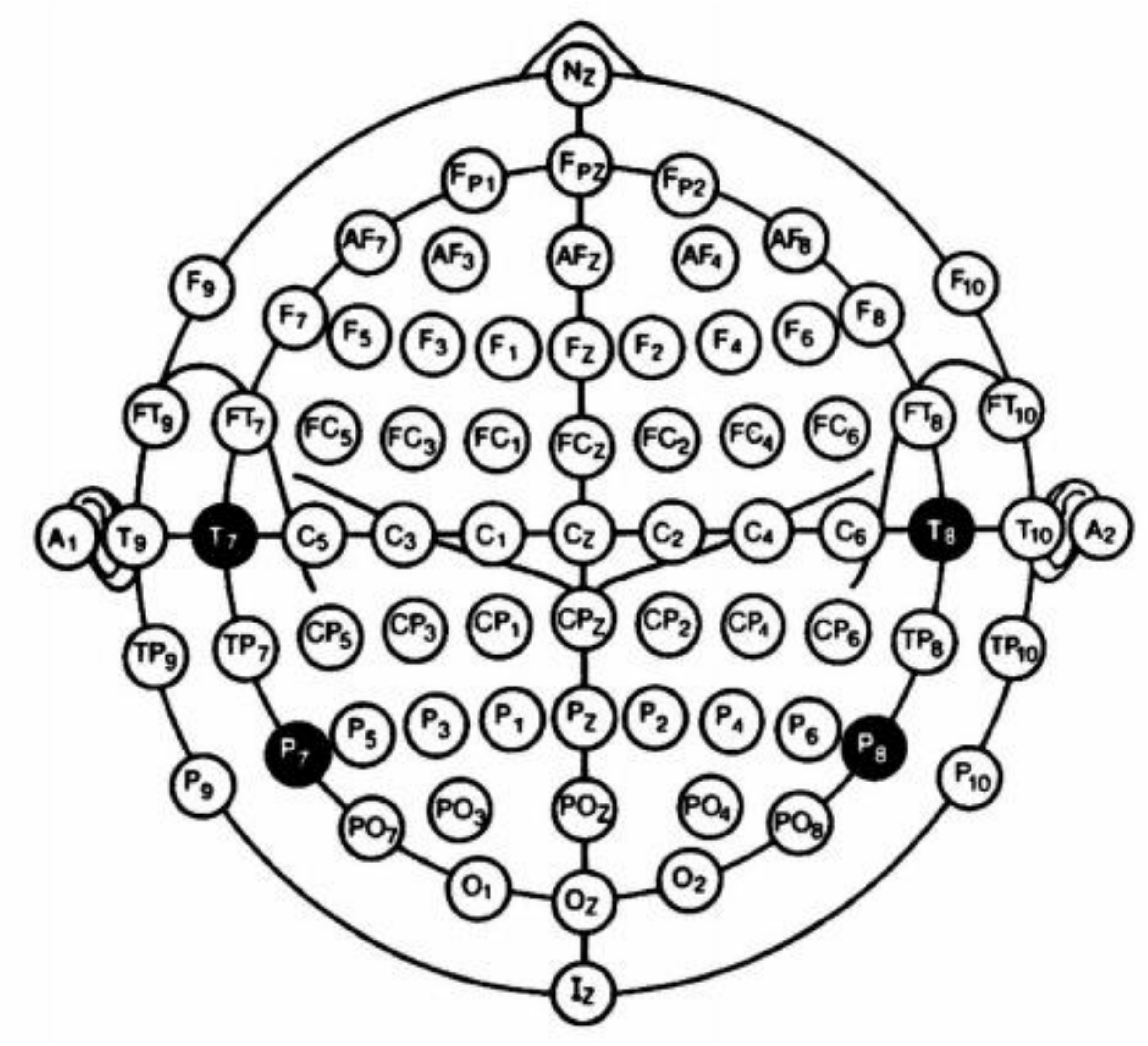


Figure 4.5: Recently updated version of the International 10-20 (or 10-10) System (from American Clinical Neurophysiology Society 2006:223)

The electrodes are labelled according to the lobe or other structural region of the brain that they are placed over (F = frontal; P = parietal; O = occipital; T = temporal; C = central; and Fp = frontal pole), as well as their distance from the midline. Electrodes that are placed over the right hemisphere are given even numbers, and electrodes that are placed over the left hemisphere are given odd numbers. The electrodes that are closer to the midline are given a smaller number than those further away from the midline, though electrodes that are placed along the midline are labelled with a “z” to avoid confusion between the number 0 and the letter <O> for “occipital”. The label N is short for *nasion*, which is the indentation between the eyes and the top of the nose; the label I is short for *inion*, which is the lump at the back of the scalp; and the label A is short for *pre-auricular points*, which are the indentations in front of the ears. The black background on four of the electrodes in figure 4.5 indicates that their labels have changed from earlier versions. Specifically, the labels T7 and T8 replaced the earlier labels T3 and T4, while the labels P7 and P8 replaced the earlier labels T5 and T6 (American Clinical Neurophysiology Society 2006:224).

The EEG equipment that I used for the experiments in this thesis is the BioSemi ActiveTwo system, imcluding a BioSemi headcap with 64 plastic electrode holders. Although this headcap uses a different labelling system from the updated version of the International 10-20 System described earlier (figure 4.5), they are actually arranged in the same layout. This is illustrated in figure 4.6, which gives the International 10-20 System labels (inside the circles) alongside the equivalent BioSemi labels (outside the circles).

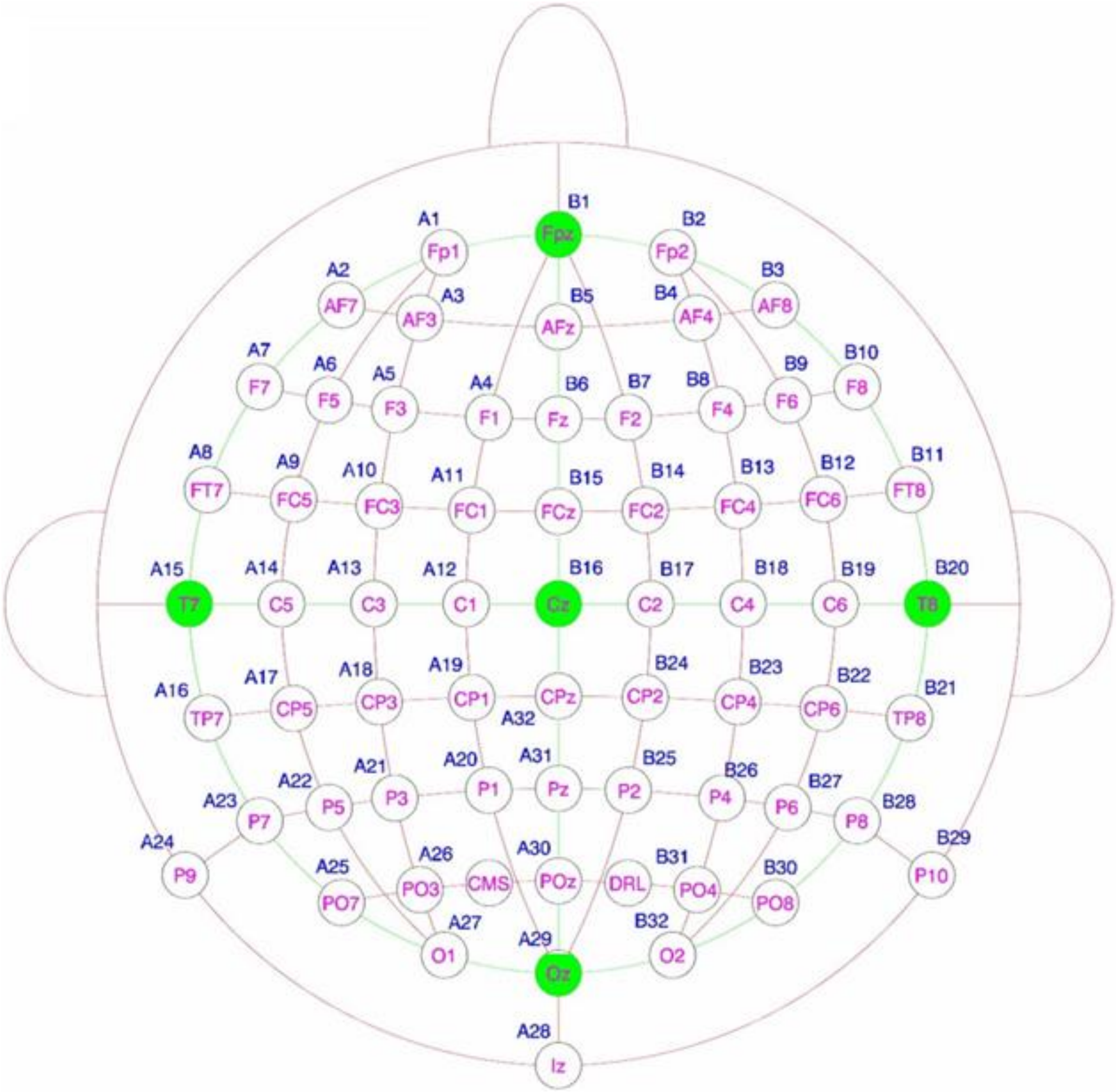


Figure 4.6: BioSemi electrode positions (from the BioSemi website: http://www.biosemi.com)

Despite the general similarities, there are a few differences between the International 10-20 system described in figure 4.5 and the arrangement of electrodes in the BioSemi headcap (figure 4.6). A clear difference is that the BioSemi headcap contains no electrodes holders on the nasion or pre-auricular points. This does not have any implications for the setup of my experiments. However, an important difference that does have implications for the setup of my experiments is the addition of two electrodes adjacent to POz, labelled CMS (*common mode* *sense*) and DRL (*driven right leg*). The CMS is designated as the *ground* electrode, which is used in calculating the voltages of all of the other active electrodes, while the DRL forms a feedback circuit with the ground to ensure that the average potential of the participant is close to the average potential of the equipment (Luck 2014a:151, 153). In the International 10-20 system, any electrode can be selected as the ground, though it is typically Fpz (Berry & Wagner 2015:2). However, in my experiments I will not need to select an electrode location for the ground due to the presence of these specific extra electrodes.

Another important difference that has implications for the setup of my experiments is that the BioSemi headcap contains no electrode holders for the electrodes that are labelled TP9 and TP10 in the International 10-20 System. These electrodes are placed on the *mastoids*, which are the bones behind the ears, and the average of the mastoids is widely used as the *reference* in EEG experiments (Lei and Liao 2017). Theoretically, the reference is an “inactive” site that the voltage at each “active” electrode is measured against (Kutas and Van Petten 1994:85). However, in practice, there is no site on the head that does not pick up any electrical activity from adjacent brain areas (Kiloh et al. 1981:46). More details about the reference electrodes will be given in section 4.6.3. Since the BioSemi headcap does not contain electrode holders for the mastoid electrodes, I will add an additional electrode on each mastoid so that the average of the mastoids can be used as a reference.

The 64 BioSemi electrodes which fit into the electrode holders on the BioSemi headcap are known as *pin-type active electrodes*. In addition to these, I also used 8 *flat-type active electrodes* which can be applied to any surface of the body using BioSemi adhesive disks. Both of these sets of BioSemi electrodes are silver chloride (Ag-AgCl) electrodes. I placed two of the flat-type electrodes on the mastoids, and I placed the remaining six around the eyes. Specifically, following Rhodes and Donaldson (2008:53), I placed an electrode above, below, and adjacent to the outer corner of each eye. The purpose of this was to help me to detect the participant’s eye blinks in the data, which is very important because blinking causes large voltage fluctuations that render affected portions of the data unusable (Kiloh et al. 1981:58). Blinks and other unwanted voltage deflections are discussed further in section 4.5.

It is important to note that placing an electrode at a particular scalp site does not mean that the electrode will only detect the electrical activity that occurs in the brain area below that particular scalp site (Kappenman & Luck 2011:7). Instead, any given electrode placed at any location on the scalp can detect electrical activity from distant areas of the brain. This is partly due to the fact that the electrical activity needs to travel through layers of skull and skin before it can be detected by scalp electrodes (Andrewes 2001:31; Kappenman & Luck 2011:7). As explained by Kappenman and Luck (2011:7), “the high resistance of the skull relative to the low resistance of the underlying brain and overlying scalp causes the voltage to spread laterally as it travels”. It is actually mathematically impossible to accurately disentangle the source of the electrical activity by only knowing the scalp location (Luck 2014a:48). This is known as the *inverse problem* (Swaab et al. 2012:418), and it explains why EEG cannot be used to draw any conclusions about the location of mental processes in the brain (Hillyard 2000:25). Nevertheless, although EEG has very low spatial resolution, it has very high temporal resolution (Brandeis & Lehmann 1986:152), and this is what is important for my research.

4.5 Experimental setup

Before describing the experimental setup and procedure used for my ERP studies, it is important to first point out that I explain the experimental setup in a much more concrete and basic way than would be expected in a psychology thesis. This is because the primary audience for my thesis is those who have a background in linguistics rather than psychology.

The setup of the EEG equipment is illustrated in figure 4.7. Every time a word is presented to a participant on the stimulus presentation computer, an event code is sent from the stimulus presentation computer’s parallel port along the BioSemi presentation cable to the trigger port on the USB2 receiver. The event code is then sent via a USB cable to the data acquisition computer, which displays the electroencephalogram and event codes while the data is being recorded. The data acquisition software used is *ActiView*, a program designed specifically for the BioSemi ActiveTwo system.

All of the electrodes are connected to a battery powered AD-box, which amplifies the incoming signals and converts them from analog to digital format. The amplified signals are then transmitted to the USB2 receiver via a fibre-optic cable, before being relayed onto the data acquisition computer. The AD-box also contains the power switch for the EEG system.

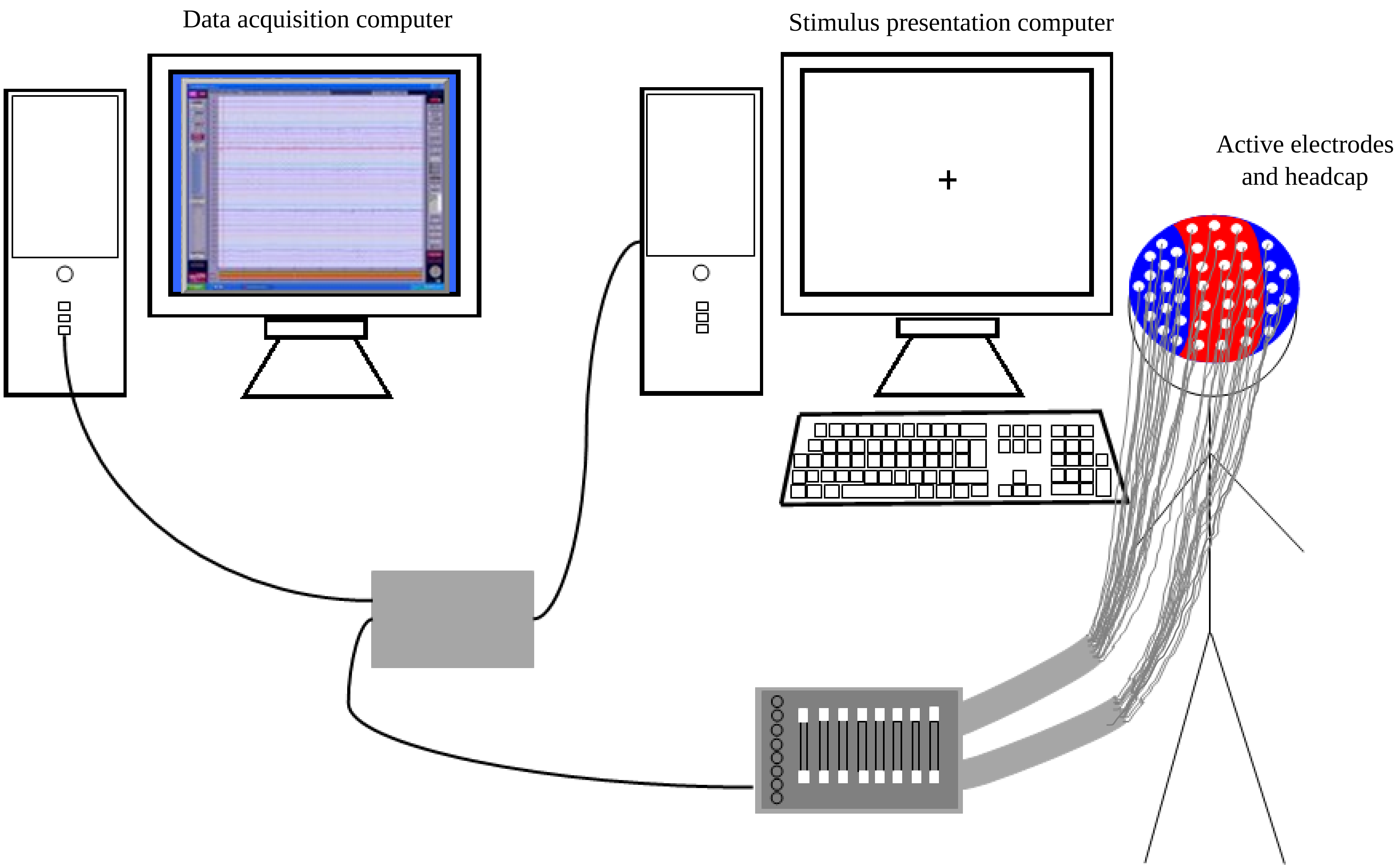


Figure 4.7: Experimental setup

The first step in applying the electrodes was to measure the head circumference of the participant with a tape measure. This allowed me to select the correct headcap size according to the measurements given in the *ActiveTwo System Operating Guidelines* (Smith 2009:8). Once I had selected the appropriate headcap for the participant, I then measured the distance between their nasion and their inion (see section 4.4) and divided this distance by two to locate the *vertex*. The vertex is the central point at the top of the scalp which is labelled Cz in the 10/20 system (Duffy et al. 1989:114).

I placed the headcap onto the participant's head from front to back, pulled out the label from underneath the headcap, and then adjusted the position of the cap to ensure that the central electrode holder was located on the vertex. This involved measuring the nasion to the vertex and adjusting the headcap position from front to back, as well as measuring the distance between the pre-auricular points and adjusting the headcap position from left to right. I also visually inspected the cap from the front and the back of the participant to ensure that the cap was not rotated.

Once the headcap was in place and the chin strap of the headcap had been secured, I used a plastic syringe to insert *SignaGel* into the 64 electrode holders. *SignaGel* is a highly conductive electrolyte gel which allows the electrodes to make contact with the participants' skin (Smith 2009:5; Parker Labs 2017). In order to ensure that the gel was making contact with the participants' skin, I asked each participant whether or not they could feel the gel on their scalp when I first started to insert the gel into the electrode holders. If the participant could not feel the gel, I used the blunt plastic tip of the syringe to part their hair. I then asked the participant again and, if they still could not feel the gel, I added more gel and continued moving their hair until they could feel it. However, I also ensured that I did not use too much gel by making sure that the level of the gel did not rise above the plastic electrode holder. This was important because, if too much gel is used, it can make create a 'bridge' (i.e. make contact) with the gel from neighbouring electrode sites; this results in identical measurements at both sites due to the conductivity of the gel. To further reduce the likelihood of this happening, I ensured that I moved the syringe away from the participant’s head when inserting the gel as opposed to holding the syringe against the scalp. This is because, if the syringe is held firmly against the scalp, the gel is more likely to spread laterally over the scalp rather than filling the electrode holder vertically (Smith 2009:41).

Once the gel had been inserted into each electrode holder, I inserted the pin-type active electrodes into the appropriate electrode holders and then plugged the electrode ribbon cables into the AD-box. I also used double-sided adhesive disks to attach the flat-type active electrodes (also with *SignaGel*) to the mastoids and to the skin around the participants’ eyes (see section 4.4). Finally, I switched the EEG system on using the power button on the AD-box, and ensured that the light on the AD-box indicating ‘CM in range’ was blue and was not flashing. This indicates that the CMS and DRL electrodes are properly connected to the participant (Smith 2009:47). The positioning of the electrodes used in my experiments is shown in figure 4.8.

Figure 4.8: Positioning of the electrodes on a test-run participant (Gillian Smith, PhD student in LAEL, Lancaster University)

Having attached all of the electrodes to the participant, the next step was to display the raw electroencephalogram in *ActiView* and inspect the *Electrode offset* tab to check the signal quality. This view displays the average voltage between the CMS and each active electrode (Smith 2009:53). The variations in voltage should be within +/-40 mV for each electrode channel and they should not change rapidly when the participant moves (Smith 2009:53). If the variation in a particular electrode channel was much greater than +/- 40 mV, I inserted more gel into that particular electrode holder; if the variation across all of the channels was much greater than +/- 40 mV, I inserted more gel into the CMS and DRL electrode holders (Smith 2009:53). To further check the signal quality, I moved to the *Monopolar display* tab in *ActiView* and ensured that there was no low frequency instability in any channels, as this indicates that an electrode might need more gel. I also ensured that the signals had an amplitude of less than 100 µV. This is because 100 µV is the maximum amplitude found in the electroencephalogram for a typical adult (Teplan 2002:2), so any signals with an amplitude of above 100 µV is unlikely to reflect brain activity.

Once I had ensured that the signal quality was acceptable, I could start presenting the stimuli to the participant and start recording the data. When recording the data in *Actiview*, I ensured that the software was saving the incoming signals from the 64 pin-type active electrodes as well as the 8 flat-type active electrodes (section 4.4). I also paused the recording during breaks between experimental blocks, and then continued saving the recording immediately after each break. This ensured that the data from each block for each participant was saved in one data file, so it could be easily transferred to the data analysis software (see section 4.6.7).

Throughout the recording session, I continuously monitored the signal quality on the *Monopolar display*. Examples of problematic waveform patterns that I looked for are shown in figure 4.9.

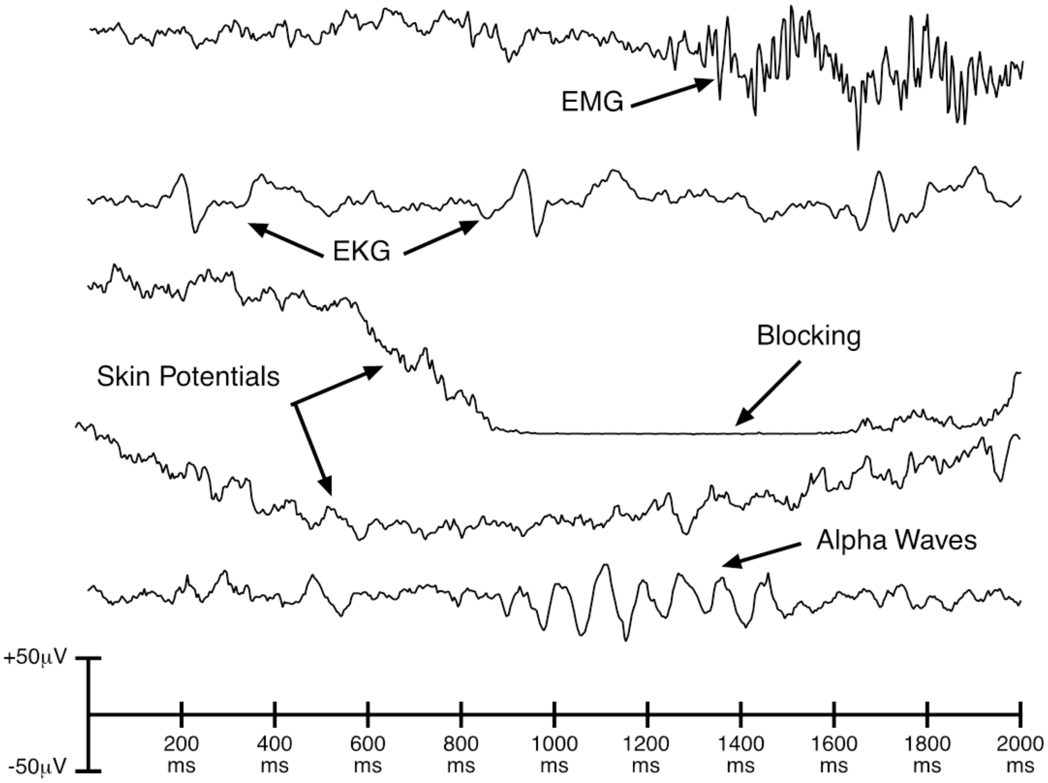


Figure 4.9: Problematic waveforms (from Luck 2014a:203)

EMG, or electromyogram, refers to the rapidly fluctuating waveform that is produced as a result of muscle activity (Longstaff 2005:230; Andreassi 2007:11). It is often possible to identify the source of the muscle activity by looking at the distribution of the EMG noise across the different channels. For instance, the channels labelled T7 and T8 in the 10/20 system are located directly above the *temporalis muscles*, which are the muscles that are used to contract the jaw (Luck 2014a:205). The EMG noise at T7, T8, and neighbouring channels can therefore be reduced by pausing the recording and the stimulus display and asking the participant to relax their jaw and avoid clenching their teeth (Luck 2014a:205). Similarly, the channels labelled Fp1 and Fp2 are located directly above the muscles of the forehead. EMG noise at these channels can therefore be reduced by asking the participant to relax their forehead muscles and ensuring that the headcap is not pulling on the participant's forehead (Luck 2014a:206).

EKG, or electrocardiogram, refers to the distinctive waveform that is produced by the electrical activity of the heart, which is about 100 times greater than the electrical activity of the brain (Darby & Walsh 2005:73). EKG noise can be picked up by the mastoid electrodes, as these are located directly above the *carotid arteries*, where the heart rate can be measured. Since the EKG is a periodic waveform, with the same cycle repeated roughly once per second, it should not cause an artificial difference between experimental conditions (Luck 2014a:206). However, EKG could potentially be problematic in my experiments because I am using the mastoids as the reference (see section 4.4 and 4.6.3), and any signal that is picked up by the reference is seen in inverted form in all other channels (Luck 2014a:206). Nevertheless, EKG noise can sometimes be reduced by slightly moving the position of the mastoid electrodes.

When a channel appears as a flat line, this can be the result of *blocking*, i.e. the situation where the AD-box becomes saturated and does not amplify all of the signals (Luck 2014a:202). However, this happens very rarely, and a channel is more likely to appear as a flat line due to an electrode having become completely disconnected (Luck 2014a:202-203). If an electrode becomes *partially* disconnected, this causes slow voltage shifts in that channel (Luck 2014a:202). These slow voltage shifts also arise when sweat accumulates in the sweat glands on the participant's scalp, as this can alter the conductivity of the channels (Majumdar 2018). Voltage shifts that arise due to the properties of the skin are known as *skin potentials*.

Another problematic waveform is the *alpha wave*. As mentioned in section 4.2, alpha activity refers to EEG oscillations between 8 and 13 Hz (Zillmer et al. 2008:41). Alpha waves are typically more prominent at occipital electrode sites, and high alpha activity indicates that the participant is tired and fatigued (Tiago-Costa 2016:90). It is therefore important to ensure that the participant remains alert throughout the duration of the recording. One way of doing this is to pause the stimulus presentation and the EEG recording to give the participant additional breaks.

Finally, additional waveform noise that is not illustrated in figure 4.11 is caused by blinking. The participant's blinks appear on the electroencephalogram as 50-100 µV voltage deflections that are sustained for roughly 200-400 ms (Luck 2014a:194). Blinks can be easily identified on the electroencephalogram, as they cause the waveform corresponding to the electrode below the eye to be opposite in polarity to the waveforms corresponding to the electrodes above the eyes (Luck 2014a:194). This can be seen is figure 4.10, where the VEOG waveform, corresponding to the channel underneath the eye, is opposite in polarity to the waveforms corresponding to the scalp electrodes (Fz, Cz, and Pz).

Figure 4.10: How a blink appears on the electroencephalogram (from Luck 2014a:195)

As is evident from figure 4.12, the amplitude of the blink response is largest at the channels nearer the eyes (VEOG, Fz) and smaller at the channels further away from the eyes. This is known as *ocular artifact propagation*, where the voltage deflections from eye blinks propagate backwards over the scalp (Haas et al. 2003:19). Following the practice recommended by Luck

(2014a:194), if a participant was blinking excessively, I would remind the participant of the need to control their blinking during the break of the experiment.

Blinks, muscle movements, skin potentials, and anything else that interferes with the quality of the data are known as *artifacts*. Participants have to be excluded from the data analysis entirely if their data has excessive artifacts. Luck (2014a:210) suggests excluding participants from the data analysis if more than 25% of their trials are rejected due to artifacts. It is possible to remove these artifacts during the data analysis (see section 4.6.5). However, Luck (2014a:186) points out that “it is always better to minimize the occurrence of artifacts rather than to rely heavily on rejection or correction procedures, which always have a cost”. Monitoring the signal quality therefore allowed me to maximise the amount of usable data obtained from each recording session and from the participant group as a whole.

4.6 Data analysis and interpretation

4.6.1 Overview

One method of analysing EEG data, and the method taken in this thesis, is to compute event-related potentials (ERPs). As mentioned in Chapter 3 (section 3.4.2), ERPs are “the momentary changes in electrical activity of the brain when a particular stimulus is presented to a person” (Ashcraft & Radvansky 2010:61). ERPs are displayed on graphs with the ERP waveform from one condition overlaid on the ERP waveform from the other condition. This shows the differences in brain response across both conditions in terms of amplitude (µV) and latency (ms).

I processed the EEG data and computed the ERPs in MATLAB (The MathWorks 2015) using two toolboxes, namely EEGLAB (Delorme and Makeig 2004) and ERPLAB (Lopez-Calderon and Luck 2014). I used EEGLAB to import the data and carry out the initial pre-processing steps. This involved scrolling through the waveforms to ensure that a signal had been recorded from each channel and to ensure that the event codes were visible in the data. It also involved specifying the scalp locations of the channels.

Having done this, I used ERPLAP to carry out the subsequent pre-processing steps, namely editing the event labels and assigning the events to bins. The slices of EEG waveform corresponding to the second word of each collocational bigram had to be labelled differently from the slices of EEG waveform corresponding to the second word of each non-collocational bigram. These waveform slices also had to be labelled differently from those corresponding to the non-experimental words, i.e. the words in the fillers (for Experiment 1), and the words in the experimental sentences that were not being analysed as part of a collocational or non-collocational bigram.

Once the events had been labelled correctly, I could assign the events to separate bins. In ERPLAB, a *bin* refers to "a set of averaged ERP waveforms, one for each electrode site" (Luck et al. 2011:12). By assigning the events for the collocational bigrams to one bin and assigning the events for the non-collocational bigrams to another bin, ERPLAB is made aware of the separate conditions of the experiment so that averaged ERP waveforms can be created for each condition.

In order for the averaged ERP waveforms to be created, I carried out the following data processing operations in ERPLAB:

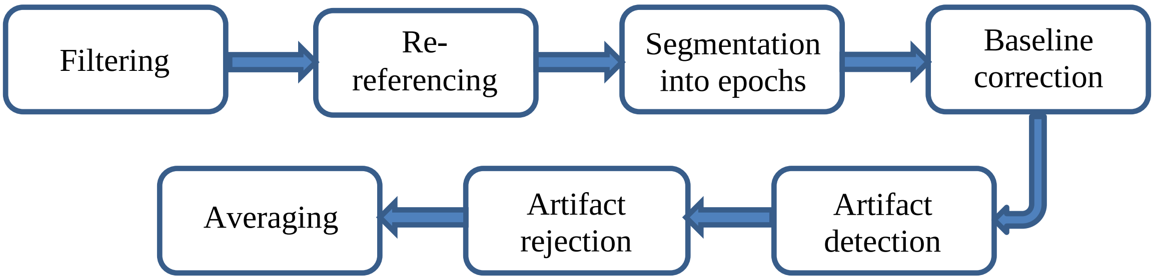


Figure 4.11: Data processing operations

Each processing operation is explained in detail in the subsequent sub-sections. The order of operations is not entirely fixed as the end result would be the same if artifact detection and rejection was carried out before segmentation, for example. However, it was important to carry out filtering first because this needs to be done on the continuous EEG rather than the segmented EEG or the averaged ERP waveforms (Tanner et al. 2015:1004). Furthermore, I chose to segment the data immediately after filtering. This makes the whole process more efficient as it means that only the event-related data is inspected and processed.

4.6.2 Filtering

The first data processing operation is filtering. The purpose of filtering is to decrease the amount of noise in the data by removing unwanted frequencies. Low-pass filters remove high frequencies such as EMG noise and 50 Hz electrical line noise, while high-pass filters remove low frequencies such as skin potentials and other slow voltage drifts (Luck 2014a:245). Following Luck’s (2014a:232, 245) recommendation, I applied a half-amplitude high-pass cutoff of 0.1 Hz and a half-amplitude low-pass cutoff of 30 Hz, with a roll-off slope of 12 dB/octave. These settings decrease the amount of noise in the data without removing the frequencies that reflect cognition. This is because, as Luck (2014a:227) points out, “most of the relevant portion of the ERP waveform in a typical cognitive neuroscience experiment consists of frequencies between approximately 0.1 Hz and 30 Hz”.

Although filters remove unwanted frequencies and therefore increase statistical power (Kappenman & Luck 2010:901), it is important to note that all filters distort the data to some extent (Tanner et al. 2015:997). High-pass filters are known to be particularly problematic because, not only can they reduce the amplitude of late ERP components (Duncan-Johnson and Donchin 1979), they can also cause artifactual peaks to appear in the data (Tanner et al. 2015:997). Tanner et al. (2015) provide concrete evidence for this by studying the effects of different high-pass filter settings on the N400 and P600 in a typical language processing experiment. They find that increasing the high-pass filter cutoff to 0.3 Hz decreases the amplitude of the P600 and induces a significant artifactual N400-like peak. They also find that, when the high-pass filter cutoff is increased to 0.7 Hz, the artifactual N400 is still significant but the P600 (the true effect in the data) is no longer significant. Using a high-pass filter of 0.3 Hz or above could therefore cause researchers to reach false conclusions (Tanner et al. 2015:1007).

According to Luck (personal communication), filtering at around 0.5 Hz is likely to be the cause of the apparent N400 in studies which claim to elicit an N400 in response to a syntactic anomaly. Nevertheless, the studies which I cited in Chapter 3 (section 3.4.5) as eliciting an N400 in response to morphosyntactic errors both use a high-pass cutoff of 0.1 Hz (Severens et al. 2008:143; Martin et al. 2012:1862), and are therefore immune to this critique. Tanner et al. (2015:1997) conclude that high-pass cutoffs of 0.1 Hz are optimal as they are "likely to attenuate low-frequency noise … while providing minimal distortion to the underlying ERP effect".

Filters can be applied during data acquisition (i.e. an online filter) or during data processing (i.e. an offline filter). With the BioSemi system, any online filters that are applied are not saved to disk. This is useful because it means that the signal can be viewed with different filter settings during data acquisition before applying the final filters offline (Luck 2014c:2). Following the recommendations given by Smith (2009:54), I disabled the filters during data acquisition so that I could monitor the overall signal quality. Yet I also occasionally applied a 0.1 Hz to 30 Hz half-amplitude cutoff in order to more closely monitor the signal quality within the frequencies of interest (Luck 2014c:2).

4.6.3 Re-referencing

The next data processing operation is re-referencing. As mentioned in section 4.4, the reference is theoretically an "inactive" site that the voltage at each "active" electrode is measured against (Kutas & Van Petten 1994:85). In practice, there is no truly “inactive” site on the head that does not pick up any electrical activity from adjacent brain areas (Kiloh et al. 1981:46). However, I chose to use the average of the mastoids as the reference as this is widely used in EEG experiments (Lei & Liao 2017). Occasionally, I found that re-referencing to the average of the mastoids inadvertently increased the noise level of the data, as any noise that is present at the reference site spreads to the rest of the electrode sites during the process of re-referencing. Specifically, any signal that is picked up by the reference is seen in inverted form in all other channels (Luck 2014a:206). In cases such as this, I would instead use a different widely used reference, such as the average of P9 and P10. These are the electrode sites that are adjacent to/closest to the mastoids, yet they often pick up less noise from the neck muscles than the mastoids (Luck 2014a:164). Another alternative reference that I used was the average reference, which is the average of all of the scalp electrodes. See Chapter 6 (section 6.3.4) for examples of this scenario.

The reference that is used at the point of data acquisition (i.e. the *online reference*) is not necessarily the reference that is used for data analysis (i.e. the *offline reference*) as it is always possible to re-reference the data (Keil et al. 2013:5). This means that the voltage at each active electrode site is given in relation to the voltage at an inactive site that is *not* the same inactive site used during data acquisition. Having the option to re-reference is useful when a researcher is interested in the electrical activity at a particular scalp location. For instance, a researcher who is interested in the electrical activity of the left temporal lobe could use the right mastoid as the reference during data acquisition. This is because the reference should not be placed near the area of interest (Noirhomme & Lehembre 2012:56). However, the polarity and amplitude of the data changes depending on which reference is used (Kutas & Van Petten 1994:85; Keil et al. 2013:5). Therefore, in order to avoid biasing the results towards the left hemisphere, the researcher could then re-reference the data using the average of the mastoids (Lopez-Calderon & Luck 2014:6).

In the BioSemi system, the data *always* needs to be re-referenced because, like with offline filters, the reference that is used during data acquisition is not saved to disk (Luck 2014c:2). Instead, the data is imported into EEGLAB as unreferenced EEG signals. I chose to view the raw unreferenced signals during data acquisition, as opposed to selecting an online reference. This is because, according to Smith (2009:54), "[t]he unreferenced view is helpful in identifying noisy channels". I then used the average of the mastoids as the offline reference (in most cases).

4.6.4 Segmentation and baseline correction

The next data processing operations are segmentation into epochs followed by baseline correction. An *epoch* or *segment* is defined as a portion of the data that is event-related. The epoch consists of the event code, a period of 100-200 ms prior to the event code, and a period of 500-1500 ms after the event code (Luck 2014a:250). The time window to be included after the event code is determined by the hypothesis of the experiment (Luck 2014a:250). For instance, an experiment focusing solely on the N400 component would need a time window up to 500 ms whereas an experiment focusing on the P600 would need a larger time window up to 800 ms (Tanner et al. 2015:1001). Since I am interested in both of these components, I used a time window from -200 ms pre-stimulus to 800 ms post-stimulus.

Once the continuous EEG waveform had been segmented into epochs, the portion of the waveform that is not event-related was discarded and a procedure known as *baseline correction* was performed on the remaining segments. According to Kutas and Van Petten (1994:90), "the baseline is presumed to be a time of inactivity and has most often been some period (50-200 ms) prior to stimulus onset". The post-stimulus activity was then analysed in relation to the baseline (Kutas & Van Petten 1994:85).

In language experiments where the experimental words are embedded into sentences, the period prior to the experimental word is not a time of inactivity. This is because the words preceding the experimental word will also cause EEG activity (Kutas & Van Petten 1994:90). It is therefore typical for language researchers to select the period prior to the onset of the sentence as the baseline as opposed to the period prior to the experimental word (Luck 2014a:255). However, this is not necessary in my experiment because the preceding words are identical in both conditions. After all, Keil et al. (2013:5) point out that the baseline "should ideally be chosen such that it contains no condition-related differences". I therefore used the -200 ms pre-stimulus window as the baseline.

The purpose of baseline correction is to bring the baseline to zero in order to isolate and quantify the experimental effect. This is achieved by subtracting the average voltage of the baseline from the remainder of the epoch (Keil et al. 2013:5). In ERPLAB, baseline correction is performed on all channels by specifying the baseline period and then running the baseline correction function on the pre-stimulus window.

4.6.5 Artifact detection and rejection

The next data processing operations are artifact detection and rejection. As mentioned in section 4.5, blinks, muscle movements, skin potentials, and anything else that interferes with the quality of the data are known as *artifacts*. ERPLAB offers several different methods for automatically detecting artifacts in epoched data. The method that I used for Experiment 1, which is also the method used in the ERPLAB tutorial (Luck et al. 2011:19), is known as the *moving window peak-to-peak amplitude* method. This method calculates the difference in amplitude between the two points that have the highest and lowest amplitude values within a specified time window (Luck 2014a:191). The calculation is first made within a time window at the beginning of the epoch, but the window then shifts to the right and the peak-to-peak amplitude is calculated within this new window. This process then continues throughout the whole epoch and, if the largest peak-to-peak amplitude value found within the epoch exceeds a particular threshold, the epoch is marked for rejection. I used the default settings in ERPLAB when carrying out the moving window peak-to-peak amplitude method. This means that the windows had a width of 200 ms, the window shifted by 100 ms each time, and the rejection threshold was set at 100 μV.

Note that, while in Experiment 1 I only used the moving-window peak-to-peak amplitude method of artifact detection, in Experiments 2, 3, and 4 I expanded the artifact detection procedure by carrying out three separate artifact detection routines on each dataset. This allowed me to more accurately detect different types of artifacts. See Chapter 6 (section 6.3.4) for a description and explanation of these additional artifact rejection techniques.

Once the automatic artifact rejection procedure was complete, I scrolled through the epoched data to ensure that no obvious artifacts had been missed by the algorithm(s). If any obvious artifacts *had* gone unnoticed, following standard practice in ERP research, I adjusted the artifact detection settings so that they were more suited to the properties of the particular dataset being processed. For example, when using the moving window peak-to-peak amplitude method in Experiment 4, I halved the voltage threshold for participants 9 and 14, as the original was not sensitive enough to capture some clear artifactual voltage deflections. However, it is important to note that, although the moving window peak-to-peak amplitude method is particularly good at detecting large voltage deflections such as blinks (Luck 2014a:191), algorithms are not as reliable when detecting other artifacts which cause smaller voltage deflections (Kutas & Van Petten 1994:97). I therefore manually inspected the epoched data for evidence of smaller artifacts such as skin potentials (see figure 4.11), and manually rejected any epochs containing these artifacts. Manual artifact detection/rejection can be problematic, as researchers differ in terms of what they reject and how much they reject. However, I minimised this problem by following Luck's (2014a:188) suggestion to only reject problematic artifacts that would cause false differences between conditions.

4.6.6 Averaging

The final data processing operation is averaging. The purpose of averaging is to compute the average ERPs at each electrode site for each participant. Luck (2014a:249) states that "averaging simply consists of summing together a set of EEG epochs and then dividing by the number of epochs". Through this process, variability between trials is reduced and any effects not resulting from the stimulus are eliminated (Anderson 2005:28). This therefore allows the researcher to isolate the experimental effect. Importantly, epochs that were marked for rejection during the artifact detection process are excluded from the averaging procedure, so that these trials do not contaminate the final results. I first carried out the averaging procedure on an individual level and then, when I was ready to produce waveforms representing the results of the whole participant group (see section 4.6.9), I carried out the averaging procedure on the group level. The average of the whole group of participants is known as the *grand average*, and it is the grand average waveforms that are presented in the analysis. Finally, it is important to note that, conceptually, averaging is a distinct processing step but, in ERPLAB, averaging is done at the same time as computing the ERPS, which I will now describe.

4.6.7 Computation and quantification of ERPs

Once the data has been processed, ERPs can be computed via the measurement tool in ERPLAB (Lopez-Calderon & Luck 2014:10). The component-based approach that I am taking is "the most common form of ERP analysis, where the amplitude and latency of specific ERP components … are quantified as a function of the specific experimental condition" (Handy 2005:33). Luck (2005:23) advises that a given experiment should focus on a maximum of two components. This is because the more components analysed in a given experiment, the higher the *experimentwise error rate* (Luck 2014a:312). The *experimentwise error rate* is "[t]he probability that at least one *p* value among all the *p* values in the statistical analyses for a given experiment will be a false positive" (Luck 2014a:356). When there is a false positive (also known as a *Type I error*), the null hypothesis is rejected even though the null hypothesis is actually true (Luck 2014a:360). Studying more components thus leads to a decrease in statistical power (Luck 2014a:333). Therefore, although in Chapter 3 I discussed five components which are potentially relevant to the study of collocation, namely the P1, the N1, the P300, the N400, and the P600, I focus on just two of these components in my experiments.

I chose to focus on the N400 and the P600, partly because these are the most widely studied components in the literature on the ERP analysis of language, but also because they are endogenous rather than exogenous components. As mentioned in Chapter 3 (section 3.4.6), endogenous components peak later than 100 ms after stimulus onset and reflect cognitive information processing, whereas exogenous components peak within 100 ms of stimulus onset and reflect the physical properties of the stimulus (Sur & Sinha 2009:70). Luck (2014a:333) points out that "[y]ou should be highly suspicious of any effects that begin within 100 ms of stimulus onset unless they reflect [physical] differences in the stimuli". Therefore, I avoid focusing on the P1 and N1 in my experiments. Likewise, although the P300 is an endogenous component, it is sensitive to task-relevant probability (Donchin & Coles 1988:367) and so it is unlikely to be sensitive to linguistic transition probabilities that are external to the task.

When quantifying the amplitude of the N400 and P600, I used a *mean amplitude measure* rather than a *peak amplitude measure* (Handy 2005:38). This is because, as noted by Picton et al. (1995:27), "picking peaks is a simplistic and often misleading view of the components of an ERP" (see Chapter 3); mean amplitude measures are also much less sensitive to noise in the ERP waveform than peak amplitude measures (Handy 2005:39). Furthermore, Luck (2014a:285-286) argues that mean amplitude measures are "superior" to peak amplitude measures because they do “a better job of treating an ERP component as something that is extended over time”. In order to compute the mean amplitude for each participant, I used the *mean amplitude between two fixed latencies* function in ERPLAB.

The two fixed latencies that I used to quantify the amplitude of the N400 were 300-500 ms after stimulus onset, while the two fixed latencies that I used to quantify the amplitude of the P600 were 500-800 ms after stimulus onset[7]. These are the time windows that are conventionally used in the analysis of the N400 and P600, respectively (Tanner et al. 2015:1001). It is crucial to determine the time windows, and thereby the components of interest, *before* looking at the ERP waveforms, in order to minimise the Type I error rate (Groppe et al. 2011:1711-1712; Keil et al. 2013:7). Ideally, before observing the ERP waveforms, researchers should also decide which electrode sites they will focus on, based on the typical scalp distributions of the components of interest (Handy 2005:36). However, Handy (2005:36) notes that “given the potential for extensive individual and group differences in ERP scalp topography, it may not always be the case that a component of interest will be maximal over the expected scalp sites”. This is especially the case for endogenous rather than exogenous components (Handy 2005:36). Therefore, although I specify time windows in advance, I do not focus on specific electrode sites (at least in Experiment 1).

The measurement tool in ERPLAB provides a variety of measurement algorithms that can be used to quantify ERP latencies (Lopez-Calderon & Luck 2014:10). Kiessel et al. (2008:250) note that “[t]he most straightforward procedure for determining latency differences, and indeed the one most often used, is to compare peak latencies”. However, as mentioned in Chapter 3 (section 3.4.8), the onset of a brain process has much more theoretical significance

---

[7] Note that these latency ranges apply only to Experiment 1. The latency ranges used in subsequent experiments are outlined in Chapter 6 (section 6.3.5).

than the peak latency of the process (Luck 2014a:54). I therefore do not use the commonly used peak latency measure. Instead, I measure *onset latency*, which is the point in time at which the waveforms in the two conditions begin to diverge (Kappenman & Luck 2011:23). As recommended by Luck (2014a:300), I measure onset latency in ERPLAB using the *fractional peak latency* measure (also known as *50% peak latency*). This technique involves ERPLAB finding the peak amplitude and then working backwards in the waveform to find the time point at which the amplitude is a given fraction (typically half) of the value of the peak amplitude (Luck 2014a:300). Rugg and Coles (1995:31) point out that, when interpreting the results from any onset latency measure:

> [i]t is important to note that the ERP difference only places an *upper* bound on the time by which processing is different. It is entirely possible that processing begins to differ at an earlier point in time, but that such a difference is not evident in the ERP.

This reflects the fact that EEG provides only a partial measure of brain activity (Kiloh et al. 1981:46).

By measuring onset latency as well as mean amplitude, I combine the advantages of a component-independent design with the advantages of a component-based design (Kappenman & Luck 2011:17). The mean amplitude measure to some extent enables me to determine whether or not the expected language-related ERP components are sensitive to the transition probabilities between words. Meanwhile, even if the expected language-related ERP components do *not* appear to be sensitive to transition probabilities, significant differences in onset latency between the two conditions provide evidence of an experimental effect that does not depend on the identification of specific components. This is important because, in order to answer my research question, it does not necessarily matter which components are or are not found to be sensitive to transition probabilities (Kappenman & Luck 2011:25). Rather, *any* significant difference between the waveforms in the two conditions constitutes evidence to suggest that the participants are implicitly sensitive to the transition probabilities between words (Otten & Rugg 2005:5). This combined approach is commonly used in ERP research (Kappenman & Luck 2011:24).

4.6.8 Statistical analysis

In ERP experiments, experimental effects are revealed once the amplitude and latency values have been tested for statistical significance (Brandeis & Lehmann 1986:153). The statistical analysis itself cannot be carried out in ERPLAB (Lopez-Calderon & Luck 2014:10). However, the ERP measurement tool used to compute amplitudes and latencies in ERPLAB allows the data to be formatted in a way that is suitable for transfer to statistical packages (Lopez-Calderon & Luck 2014:10). The statistical package used to perform the statistical analyses for my experiments is SPSS version 22 (IBM Corp. 2013).

The most common statistical test used in the analysis of ERP data is the repeated measures analysis of variance (ANOVA) (Dien & Santuzzi 2005:57), so this is the statistical test that I used to analyse the ERP data in my experiments. Separate repeated measures ANOVAs were conducted for amplitude and latency measurements (Luck 2014a:312). For each ANOVA, I used three factors, namely experimental condition (collocational bigrams vs. non-collocational bigrams), anterior-to-posterior electrode position, and left-to-right electrode position. Following Luck (2014a:314), I used three levels for the left-to-right electrode position factor (left hemisphere, right hemisphere, and midline), and three levels for the anterior-to-posterior electrode position factor (frontal, central, and posterior). This created nine electrode zones, as illustrated in figure 4.12.

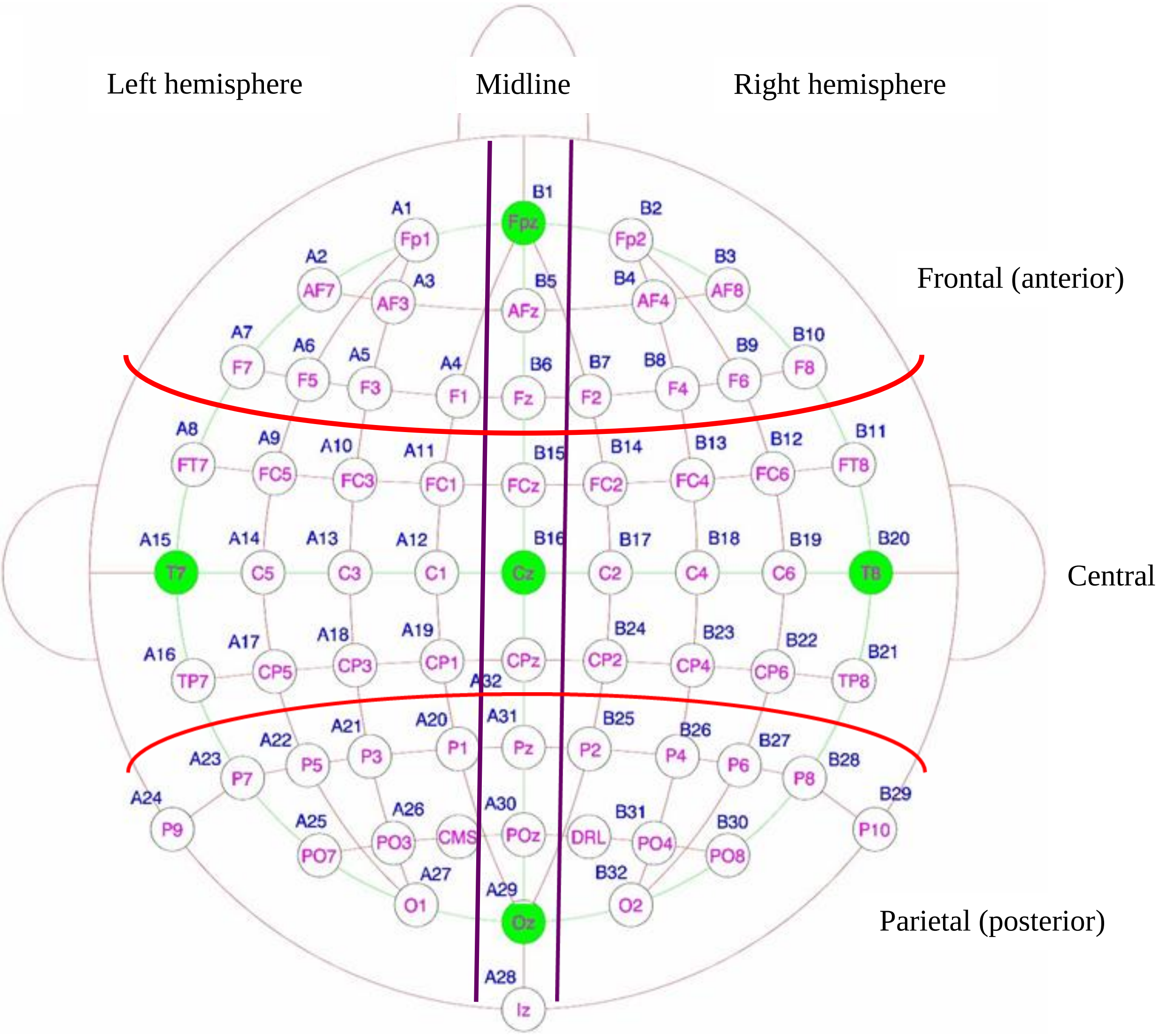


Figure 4.12: BioSemi electrode positions grouped for statistical analysis. Purple lines designate left-to-right electrode position while red lines designate anterior-to-posterior electrode position.

Note that, while I use the term *posterior* for the third level of the anterior-to-posterior electrode position factor, Luck (2014a:314) uses the term *parietal*. I chose to use *posterior* rather than *parietal* in order to clarify the fact that this level includes electrode sites that cover both the parietal and occipital lobes. It seems to be commonplace throughout the ERP literature for the terms *parietal* and *posterior* to be interchangeable, despite them technically designating different structural areas of the brain. Thus, throughout this thesis, I use the term *posterior* to refer to component scalp distributions which include the parietal and occipital lobes, even when the researchers that discovered the components may have used the word *parietal*. This is the case with the N400, which is typically said to have a central-parietal scalp distribution (Swaab et al. 2012:399), but actually has what is better described as a central-posterior scalp distribution.

When interpreting the output from the repeated measures ANOVA for both the component-based (N400 and P600) and the component-independent analysis stages of each experiment, I looked for a significant main effect as well as significant interactions. Specifically, I looked for a significant interaction between condition and the left-to-right electrode position factor, and condition and the anterior-to-posterior electrode position factor. In ERP studies, it is expected that there will *not* be a significant main effect; as Luck (2014a:336) explains, "[b]ecause the difference between conditions is likely to be large at a subset of the sites and small or even opposite at others, you probably won't see a significant main effect of condition". Significant effects are therefore expected to take the form of interactions. This is not a problem, because a significant interaction between condition and electrode site still constitutes a significant difference between conditions (Luck 2014a:336). In cases where I did find a significant interaction, which occurred in most of the ANOVAs conducted for this thesis, I carried out post hoc repeated measures ANOVAs to pinpoint the specific electrode zone(s) where the effect was maximal.

A problem with using the repeated measures ANOVA in ERP research is that "the assumptions of ANOVA are violated by almost every ERP experiment, so the *p* values that we get are only approximations of the actual probability of a Type I error" (Luck 2014a:309). Like any ANOVA, the repeated measures ANOVA works on the assumptions of *normality* and *homogeneity of variance* (Howell 2014:477). If a dataset adheres to the normality assumption, most of the data points cluster around the middle value and there is a roughly equal number of data points either side of the middle value, with the number of data points decreasing with increasing distance from the middle value. If a dataset adheres to the homogeneity of variance assumption, the different conditions display equal variation. It is not particularly problematic if an ERP experiment violates these assumptions, as ANOVAs are fairly robust to violations of the assumptions of normality and homogeneity of variance (Dien & Santuzzi 2005:69; Howell 2014:424).

However, repeated measures ANOVAs work on the additional assumption of *homogeneity of covariance,* and it can be problematic if this assumption is seriously violated (Howel 2014:477). If a dataset adheres to the homogeneity of covariance assumption, the correlation for each pair of levels of a particular factor should be constant (Howell 2014:477). For instance, for the anterior-to-posterior electrode positions factor in my experiments, the data from frontal and central electrode sites should correlate to the same extent as the data from the central and posterior electrode sites and the data from the frontal and posterior electrode sites. However, ERP datasets that use electrode site as a factor in a repeated measures ANOVA commonly violate this assumption because “data from nearby electrodes tend to covary more than data from distant electrodes” (Luck 2014a:318-319). Nevertheless, violations of the homogeneity of covariance assumption can be corrected for by applying the Greenhouse-Geisser adjustment (Howell 2014:477). Luck (2014a:319) explains that this adjustment “counteracts the inflation of the Type I errors produced by heterogeneity of covariance”, so I used this adjustment when conducting the statistical analyses in my experiments.

4.6.9 Plotting ERP waveforms

When presenting the results of ERP experiments, as well as providing the results of the statistical analysis, it is important to provide a visual representation of the results in the form of grand average ERP waveforms. I plotted the waveforms in ERPLAB, making sure to overlay the wave for both conditions on the same graph. However, ERPLAB does not produce graphs which are suitable for publication (Lopez-Calderon & Luck 2014:9), so I then edited the graphs using the open source graphics programme Inkscape 0.91.

4.7 Chapter summary and conclusion

In this chapter I have outlined the methodological steps and explained the methodological decisions that are central to the ERP studies in this thesis. I first described the selection and presentation of stimuli, before describing the electrode placement and the experimental setup and procedure. I then described each of the steps required to process and analyse the EEG/ERP data, from the pre-processing steps of editing event values and assigning events to bins, to the initial processing steps of filtering and re-referencing, to the computation and quantification of ERPs and, finally, the statistical analysis and the plotting of ERP waveforms.

To conclude this chapter, in the four experiments conducted for this thesis, I operationalized collocations as adjective-noun bigrams with a transition probability of at least 0.01011 in the BNC1994. I embedded the bigrams into sentences which were presented word-by-word to participants using the psychology software E-Prime 2.0 (Schneider et al. 2002), and I used the ERP technique of measuring brain activity.

Brain activity was recorded using the BioSemi ActiveTwo system (with 64 scalp electrodes) and the ActiView data acquisition software. Having obtained the data, I then processed the data in EEGLAB (Delorme and Makeig 2004) and computed ERPs in ERPLAB (Lopez-Calderon and Luck 2014). All datasets were filtered offline using a half-amplitude high-pass cutoff of 0.1 Hz and a half-amplitude low-pass cutoff of 30 Hz, with a roll-off slope of 12 dB/octave. I used the average of the mastoids as the offline reference, and I used the 200 ms pre-stimulus window as the baseline.

In order to combine the advantages of a component-independent design with the advantages of a component-based design, I measured the onset latency as well as the mean amplitude of the N400 and P600 latency ranges. I then conducted repeated measures ANOVAs and planned post-hoc repeated measures ANOVAs in SPSS.

Methodological decisions that are specific to the individual experiments of this thesis will be outlined in the relevant chapters. In the next chapter, Chapter 5, I focus on Experiment 1 (i.e. the pilot study). In Chapter 6 I focus on Experiment 2 (the native speaker study), and in Chapter 7 I focus on Experiment 3 (i.e. the non-native speaker study). In the final experiment chapter, Chapter 8, I focus on Experiment 4 (i.e. the replication and correlational analysis). In each of these chapters, I start by stating the aims and hypotheses of the relevant study. I then outline any methodological steps or decisions that are specific to that experiment, before presenting the results and providing a discussion of the findings. Finally, I end each chapter with a short summary and conclusion.

CHAPTER 5: EXPERIMENT 1 – A pilot study

5.1 Overview of the chapter

In this chapter I focus on Experiment 1, which serves as a pilot study in this thesis. In section 5.2, I state the aims and hypotheses of the pilot study. Although a full description of the methodology is provided in the previous chapter, I describe the methodological decisions that are specific to this pilot study in section 5.3. In section 5.4 I show the results of the pilot study, first from the component-based experimental design (section 5.4.1) and then from the component-independent experimental design (section 5.4.2). I then discuss the results as well as the limitations of the pilot study in section 5.5, before concluding the chapter in section 5.6.

5.2 Aims and hypotheses: Experiment 1

Experiment 1 is an exploratory study with two main aims:

Aim 1: To pilot a procedure for determining whether or not there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams.

Aim 2: To refine the hypotheses and methods in preparation for Experiments 2 and 3.

Two hypotheses are associated with these aims:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will elicit a P600.

These hypotheses are based on the results of previous ERP studies which find an N400 and a P600 in response to violations in expectancy and predictability (see Chapter 3, sections 3.4.6 and 3.4.7).

5.3 Methodology: Experiment 1

5.3.1 Participants

Experiment 1 was carried out on 16 participants, as Luck (2014a:262) recommends using 12-16 participants per EEG experiment (though see Chapter 9, section 9.5, for associated limitations). All participants were native speakers of English with normal or corrected-to-normal vision, and no (history of) neurological disorders, and all were students at Lancaster University. Participants were recruited in line with the ethics procedures of the Department of Linguistics and English Language, Lancaster University (see appendices 4 and 6 for information sheet and consent form).

Luck (2014a:210) suggests excluding participants from the data analysis if more than 25% of their trials are rejected due to artifacts. Following this recommendation, 4 out of the 16 participants were excluded from my analysis, leaving just 12 datasets. This is a fairly high attrition rate, which I attempt to rectify in subsequent experiments (see Chapter 6, section 6.3.1, and Chapter 8, section 8.3.2).

5.3.2 Stimuli

The experimental bigrams used in this pilot study (as well as Experiments 2 and 3) are shown in table 5.1.

Table 5.1: Experimental bigrams for Experiments 1, 2, and 3

Collocational bigrams	Non-collocational bigrams
nineteenth century	nineteenth position
endangered species	endangered fish
foreseeable future	foreseeable weeks
wide range	wide city
minimum wage	minimum prize
vast majority	vast opportunity
head teacher	head character
classic example	classic company
profound effect	profound person
random sample	random content
clinical trials	clinical devices
chief executives	chief definitions
key issues	key costs
massive increase	massive statement
crucial point	crucial night

These bigrams were chosen because they have the specific features that are being targeted in the experiment, i.e. the collocational bigrams have a high transition probability whereas the non-collocational bigrams have a transition probability of zero. The bigrams in both conditions share the same adjective; furthermore, the nouns are well-matched for frequency and length (both in terms of the number of letters and the number of syllables); see table 5.2 for a full listing of these properties.

Table 5.2: Statistical properties of Experiment 1/2/3 bigrams extracted from written BNC1994

Adjective (X)	Noun (Y)	Frequency of X	Frequency of Y	Frequency of X-then-Y	Transition Probability	Number of letters in Y	Number of syllables in Y	In no. of texts
Nineteenth	Century	3057	19025	2614	0.855	7	3	569
	Position	3057	21010	0	0	8	3	0
Endangered	Species	369	9506	252	0.683	7	2	132
	Fish	369	9431	0	0	4	1	0
Foreseeable	Future	410	12933	278	0.678	6	2	234
	Weeks	410	13018	0	0	5	1	0
Wide	Range	9601	18799	2690	0.28	5	1	995
	City	9601	21329	0	0	4	2	0
minimum	wage	1413	2811	347	0.246	4	1	109
	prize	1413	2853	0	0	5	1	0
Vast	majority	4398	9296	803	0.183	8	4	484
	opportunity	4398	9314	0	0	11	5	0
Head	teacher	1227	7945	198	0.161	7	2	88
	character	1227	8037	0	0	9	3	0
classic	example	1280	34127	176	0.138	7	3	149
	company	1280	38129	0	0	7	3	0
profound	effect	1371	21640	123	0.09	6	2	111
	person	1371	22011	0	0	6	2	0
random	sample	2156	4140	123	0.057	6	3	60
	content	2156	4270	0	0	7	2	0
clinical	trials	2948	2056	142	0.048	6	2	65
	devices	2948	2113	0	0	7	3	0
Chief	executives	3213	1283	138	0.043	10	4	84
	definitions	3213	1285	0	0	11	4	0

The experimental bigrams were embedded into sentences and ordered into four counterbalanced lists (see appendix 1) in accordance with the methodological decisions outlined in Chapter 4 (section 4.2.2). These experimental sentences are shown in table 5.3.

Table 5.3: Experimental sentences for Experiments 1, 2, and 3

Experimental sentences - bigrams in **bold**; T/F statements indented and in *italics*
The **nineteenth century** was a time of religious revival and controversy.
There was no religious controversy in the nineteenth century.
The **nineteenth position** in the competition’s scoring system is second to last.
The competition has a scoring system.
The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.
The legislation that exists to protect endangered species is always adequate.
The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.
The legislation that exists to protect endangered fish lacks proper enforcement.
In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.
In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.
There is a **wide range** of new products on offer and this is attracting many new customers.
There are not many new products.
There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.
The wide city is on top of the hill.
The **minimum wage** is designed to help people in low pay service industries.
The minimum wage is designed to help people in high pay service industries.
The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.
The minimum prize included £500 plus four nights in a luxury hotel.
The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
Most people do not lose weight on 1,500 calories per day.
The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
The competition was in the March issue of the Clothes Show Magazine.
The **head teacher** in the local primary school was featured in the local newspaper.
The head teacher in the local primary school was featured in a national newspaper.
The **head character** in the local production of Annie is played by a girl from the local primary school.
A girl from the local primary school is involved in a local production of Annie.
It is a **classic example** of losing sight of the wood for the trees.
It is a classic example.
It is a **classic company** with an excellent reputation and a large workforce.
The company has an excellent reputation.

Amy was telling told John about the **profound effect** that the book had on her.
The book had a profound effect on Amy.
Amy told John about the **profound person** that she met on holiday many years ago.
John told Amy about a person who he met on holiday.
The **random sample** of 50 observations was sufficient to carry out statistical tests.
The random sample contains 65 observations.
The **random content** of the article made it difficult to identify the target audience.
The content of the article was random.
The high cost of **clinical trials** is delaying progress in medical research.
Clinical trials are expensive.
The high cost of **clinical devices** is delaying progress in medical research.
Clinical devices are expensive.
The **chief executives** are optimistic about the prospects of their companies.
More than one chief executive is optimistic about the prospects of their companies.
The **chief definitions** are listed at the top of each dictionary entry.
The top of each dictionary entry contains the chief definition.
The **key issues** associated with the development will be discussed in the monthly meeting.
The key issues will be discussed in the weekly meeting.
The **key costs** associated with the development will be discussed in the monthly meeting.
The key costs will be discussed in the monthly meeting.
The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.
There is an increase in traffic in rural areas.
The **massive statement** concerning the regional distribution of urban growth was followed by two other studies.
The statement concerned the regional distribution of rural growth.
It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.
A very important point was raised in parliament.
It was a **crucial night** for basketball in this country and there were many disappointed fans.
Basketball fans were very happy.

Some of these sentences were taken directly from concordance lines. Examples include *The* ***nineteenth century*** *was a time of religious revival and controversy* and *It is a* ***classic example*** *of losing sight of the wood for the trees*. However, although all of the sentences were based on concordance lines, most had to be edited to ensure that they met the experimental requirements outlined in Chapter 4 (section 4.2.2). Specifically, the sentences had to be semantically coherent as standalone units, the words preceding the bigrams had to create a low contextual constraint and be identical in both conditions, and the bigrams had to be placed near the beginning or middle of the sentences to avoid sentence wrap-up effects. As can be seen from table 5.3, there are always more words following the bigram than preceding the bigram, even in the sentences where the bigram is placed near the middle, e.g. *Amy told John about the* ***profound effect*** *that the book had on her*.

Table 5.4 shows the 13 fillers and 2 practice sentences that were drawn from the set of fillers used by Millar (2010:110). Each filler and practice sentence contains a collocational bigram (see Chapter 4, section 4.4.2).

Table 5.4: Fillers and practice sentences

Fillers - bigrams in **bold**
Claudia passed the interview, but had to take an **aptitude test** before she got the job.
Claudia failed the interview.
The scientific study showed that female smokers in older **age groups** were at greatest risk.
Young male smokers are at greatest risk.
Eighty five year old Mrs. Brown can still **recite poems** she learnt off by heart at primary school.
Mrs. Brown has forgotten all of the poems that she learnt at primary school.
The judge found Paul guilty and gave him one hundred hours of **community service** and a heavy fine.
The judge decided that Paul was not guilty.
After David lost his job, he just sat around **doing nothing** all the time.
David sat around doing nothing after losing his job.
Most of the people on the island **earned a living** through tourism or fishing.
There is very little tourism or fishing on the island.
The brochure said the hotel was a **five-minute walk** from the beach, but actually it was over a mile.
It took more than 5 minutes to walk from the hotel to the beach.
The government announced plans to build a **high speed** train link from the airport to the city.
The high speed train link from the airport to the city has already been built.
About 500,000 people stood in the **pouring rain** to listen to the Pope speak.
It was not raining when the Pope was speaking.
Andy’s company tried to encourage staff to use **public transport** instead of driving to work.
Andy’s company prefer it when people drive to work instead of using public transport.
All the sunbathing Sandy did as a teenager has caused **serious damage** to her skin.
Sandy’s skin is seriously damaged.
Nicole’s dad has very **strong opinions** on the subject of politics and immigration.
Nicole has very strong opinions about politics and immigration.
During the meeting with her boss, Linda listened carefully and **took notes** of the main points.
Linda did not listen during the meeting.
Practice sentences - bigrams in **bold**
The threat of job losses has caused **considerable alarm** among the workers.
The workers were worried about losing their jobs.
After Steve left his rucksack on the train, he contacted the **lost property** office.
Steve lost his rucksack.

As mentioned in section 4.2.2, the sentences were placed into 4 differently ordered lists to control for order effects. Moreover, whenever a word from an experimental bigram occurred elsewhere in the stimuli, the sentences with the experimental bigram came first to avoid unwanted priming effects. For instance, the word *city* is part of the non-collocational bigram *wide city*, but it is also part of the filler sentence *The government announced plans to build a high speed train link from the airport to the city*. I therefore ensured that, in all 4 lists, the experimental sentence containing the bigram *wide city* was presented before the filler containing this noun.

5.3.3 Procedure

Due to the small-scale nature of this pilot study, each of the 30 experimental bigrams and each of the 13 filler items were presented just once to each participant. Also note that, unlike in the subsequent experiments in this thesis, the word-by-word presentation of the sentences involved each word immediately replacing the previous word rather than having a white screen between each word. The rationale for including intervening white screens in subsequent experiments is provided in section Chapter 6 (section 6.3.1).

5.3.4 Data analysis

In this pilot study, I follow the data analysis routine outlined in Chapter 4 (section 4.6).

5.4 Results: Experiment 1

5.4.1 Component-based experimental design

5.4.1.1 N400 (300 - 500 ms)

A comparison of the mean amplitude between the 300-500 ms latency range in the two conditions reveals that there is a slightly positive amplitude in the collocational condition ($M = 0.242$, $SD = 4.381$) and a slightly negative amplitude in the non-collocational condition ($M = -0.115$, $SD = 2.859$). The results of the repeated measures ANOVA reveal that this difference is statistically significant, *F*(1, 758) = 5.497, *p* = 0.019. This was unexpected, as Luck (2014a:336) explains that ERP researchers are unlikely to find a significant main effect (see Chapter 4, section 4.6.8).

Figure 5.1 displays the grand average ERP waveforms for one electrode site at each of the nine electrode zones used in the statistical analysis, namely left anterior, mid anterior, and right anterior, left central, mid central, and right central, and left posterior, mid posterior, and right posterior. For clarity, the electrode sites used to represent each zone are circled in figure 5.2. These particular electrode sites were selected because they are used as representative electrode sites by Rhodes and Donaldson (2008:54).

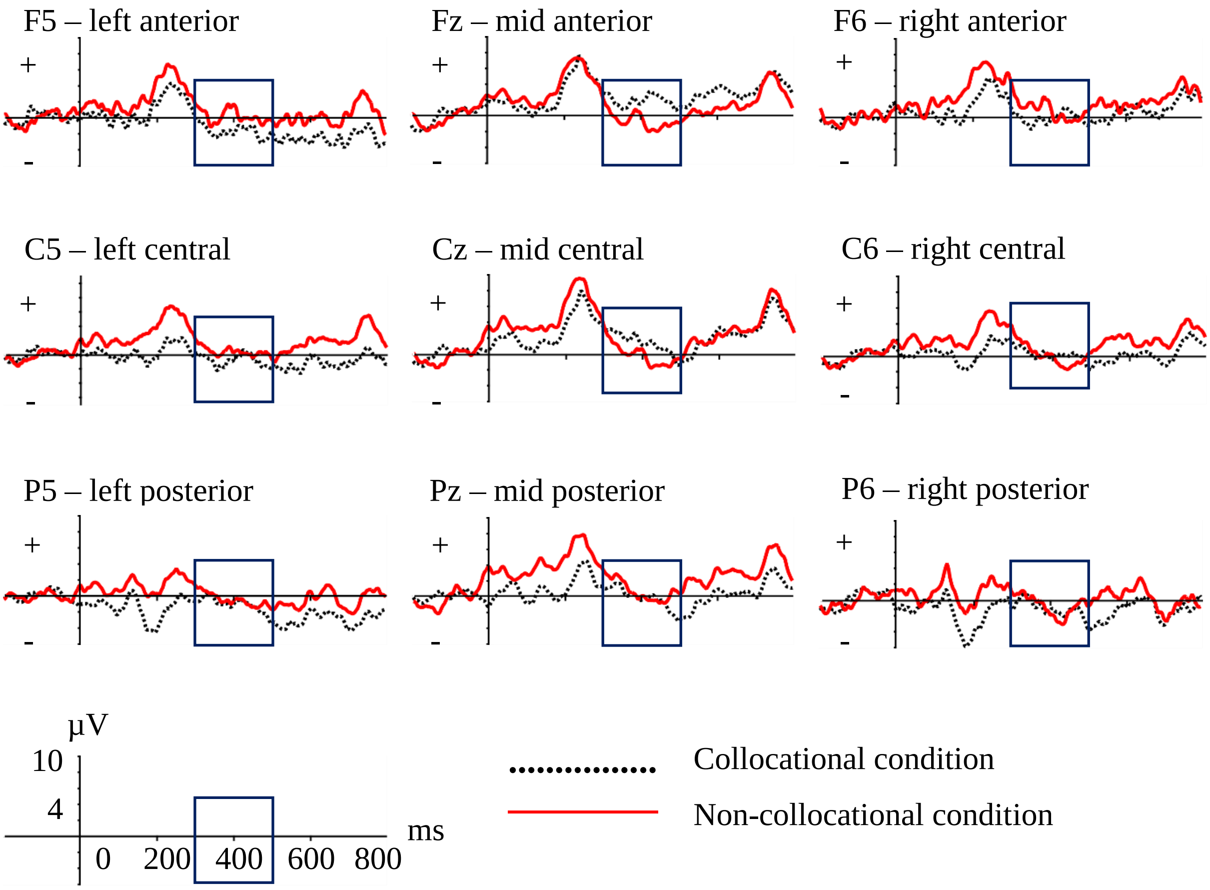


Figure 5.1: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus. The box on each waveform illustrates the typical latency range of the N400 (300-500 ms).

As mentioned in Chapter 3 (section 3.4.2), all waveforms shown in this thesis are plotted with positive voltage upwards.

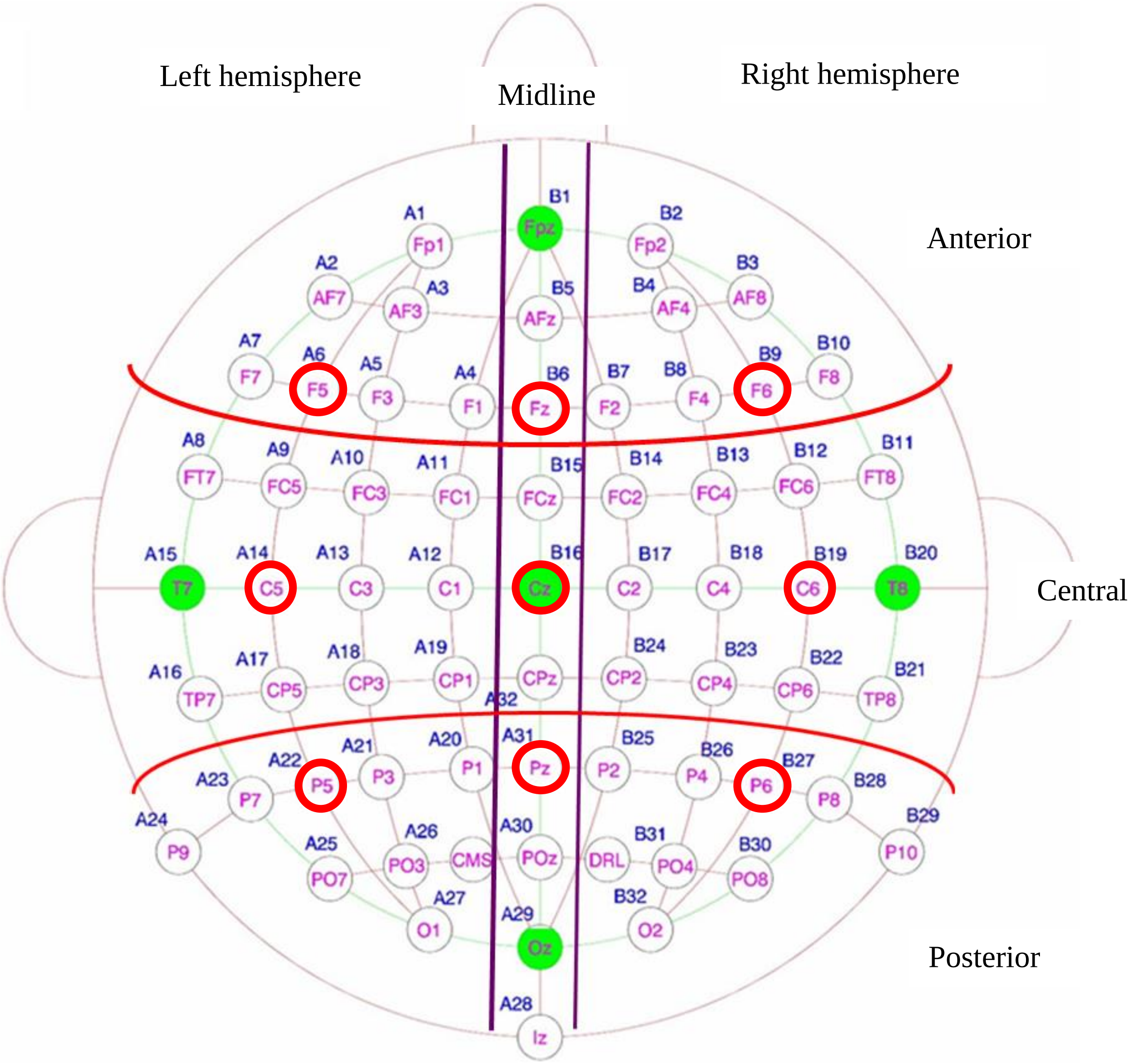


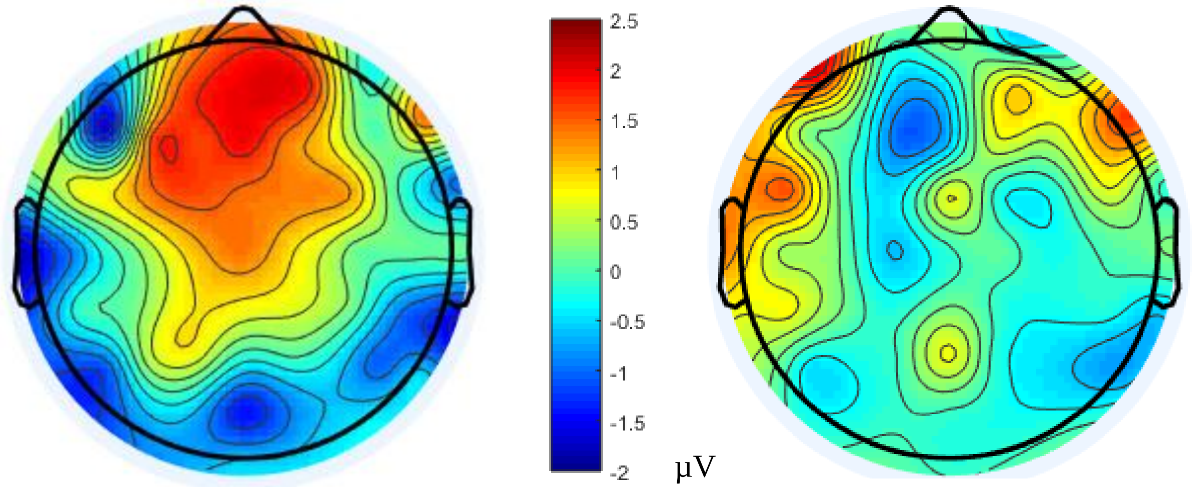
Figure 5.2: BioSemi electrode positions grouped into nine zones for statistical analysis. Purple lines designate left-to-right electrode position while red lines designate anterior-to-posterior electrode position. The electrode sites used to represent each zone are circled in red.

The results of the repeated measures ANOVA reveal that, although there is no interaction between condition and left-to-right electrode position, there is a significant interaction between condition and anterior-to-posterior electrode position, $F(2, 758) = 3.875$, $p < 0.021$. Visual inspection of the waveforms in figure 5.1 reveals that the greater negativity in the non-collocational condition is most notable at the mid central and mid anterior electrode sites. This observation is supported by the results of subsequent repeated measures ANOVAs, carried out separately on anterior, central, and posterior electrode sites. The results of these subsequent ANOVAs are summarised in table 5.5. By convention, *p*-values in tables are presented using the following grouped significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Table 5.5: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 300-500 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Anterior	1.045	5.114	0.164	3	.011*
Central	0.347	4.149	-0.113	2.769	.051
Posterior	-0.586	3.852	-0.345	2.846	.41

The mean amplitude at the 300-500 ms latency range is lower in the non-collocational condition at anterior and central electrode sites, but slightly lower in the collocational condition at posterior electrode sites. The difference between conditions is statistically significant at anterior electrode sites ($F(1, 201) = 6.579$, $p = .011$), and is marginally significant at central electrode sites ($F(1, 321) = 3.838$, $p = .051$). This greater negativity in the non-collocational condition at anterior-central electrode sites is illustrated using topographic scalp maps in figure 5.3.



Collocational condition

Non-collocational condition

Figure 5.3: Topographic scalp maps showing mean amplitude between 300 and 500 ms

Grand average ERP waveforms from each of the mid anterior and mid central electrode sites are shown in figure 5.4. At each of these sites, there is a greater negativity in the non-collocational condition.

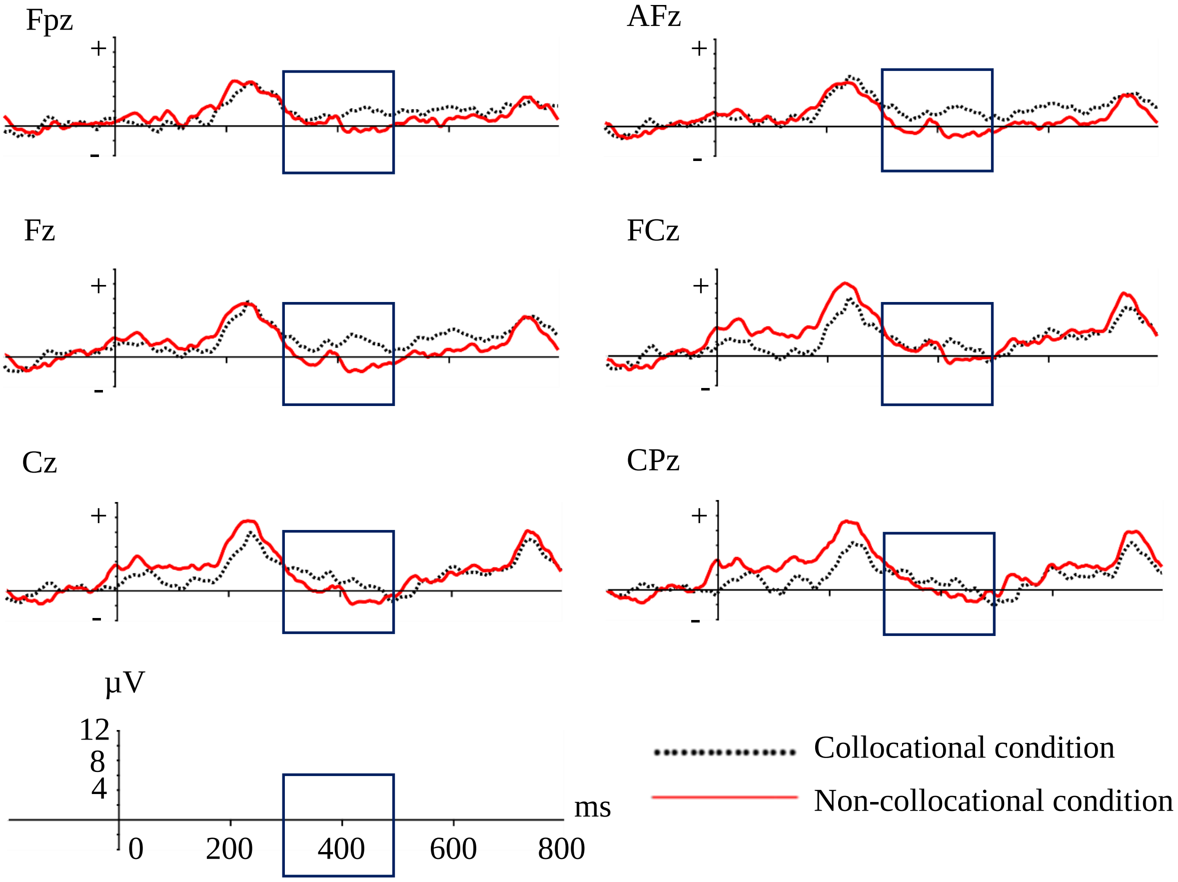


Figure 5.4: Grand average ERPs across mid anterior and mid central electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus.

The fact that there is a significantly greater negativity in condition 2 at the typical latency range for the N400 provides evidence in support of my hypothesis that reading the second word of non-collocational bigrams will elicit an N400 (see Chapter 3, sections 3.4.6 and 3.4.7). However, while the N400 typically has a central-posterior scalp distribution (Swaab et al. 2012:399) (see section 3.4.3), my results show that there is a greater negativity at the typical N400 latency range only at central-anterior electrode sites. This provides evidence against my hypothesis that reading the second word of non-collocational bigrams will elicit an N400. Nevertheless, Kutas and Dale (1997:222) point out that “N400s do differ in latency and scalp distribution, even within presumably similar experimental tasks”. Furthermore, as mentioned in section Chapter 4 (section 4.6.7), there is “potential for extensive individual and group differences in ERP scalp topography", especially for endogenous components such as the N400, so "it may not always be the case that a component of interest will be maximal over the expected scalp sites" (Handy 2005:36). With this in mind, I would argue that the aforementioned results provide sufficient evidence to show that the second word in non-collocational bigrams elicit an N400. I therefore investigate this further in Experiments 2, 3, and 4.

5.4.1.2 P600 (500 - 800 ms)

A comparison of the mean amplitude in the 500-800 ms latency range between the two conditions reveals that there is a greater positivity in the non-collocational condition (*M* = 1.279, *SD* = 3.406) compared to the collocational condition (*M* = 0.24, *SD* = 4.678). The results of the repeated measures ANOVA reveal that this is a statistically significant main effect, *F*(1, 53) = 12.732, *p* = .001. This, again, was an unexpected finding. Grand average ERP waveforms for one electrode site at each of the nine electrode zones used in the statistical analysis are shown in figure 5.5.

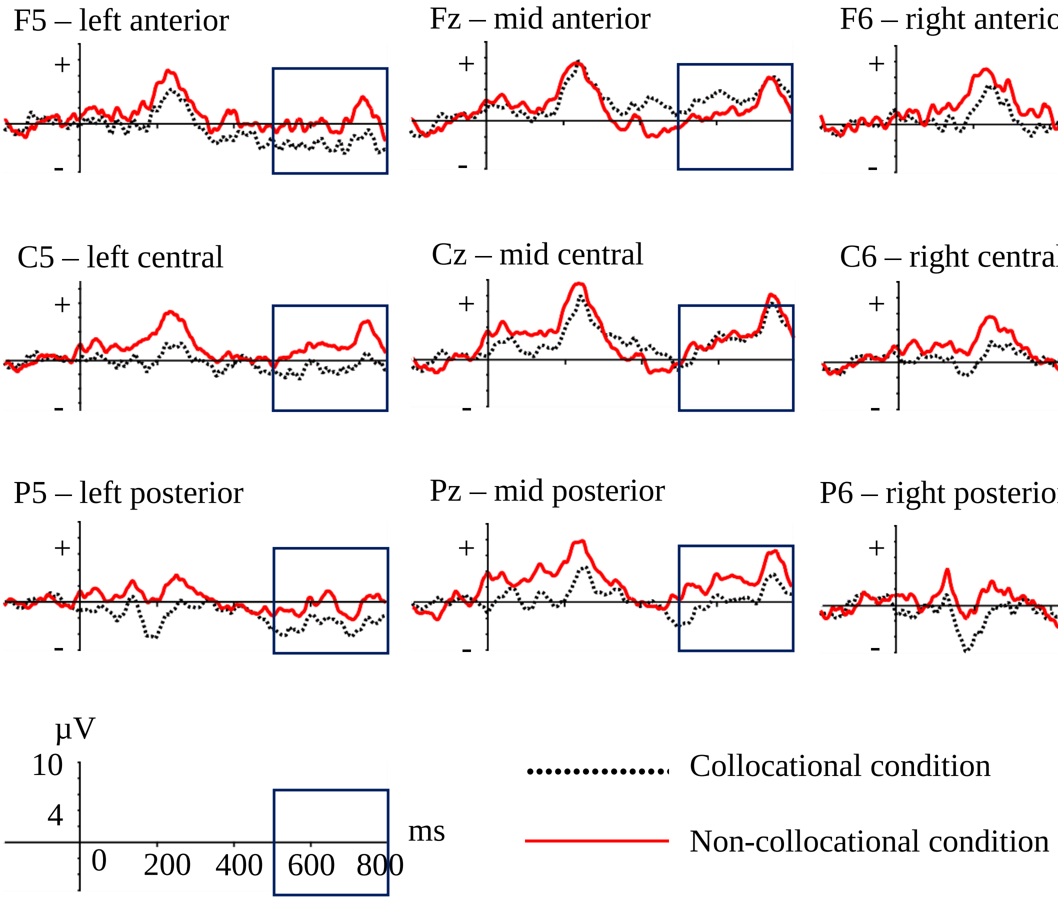


Figure 5.5: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus.

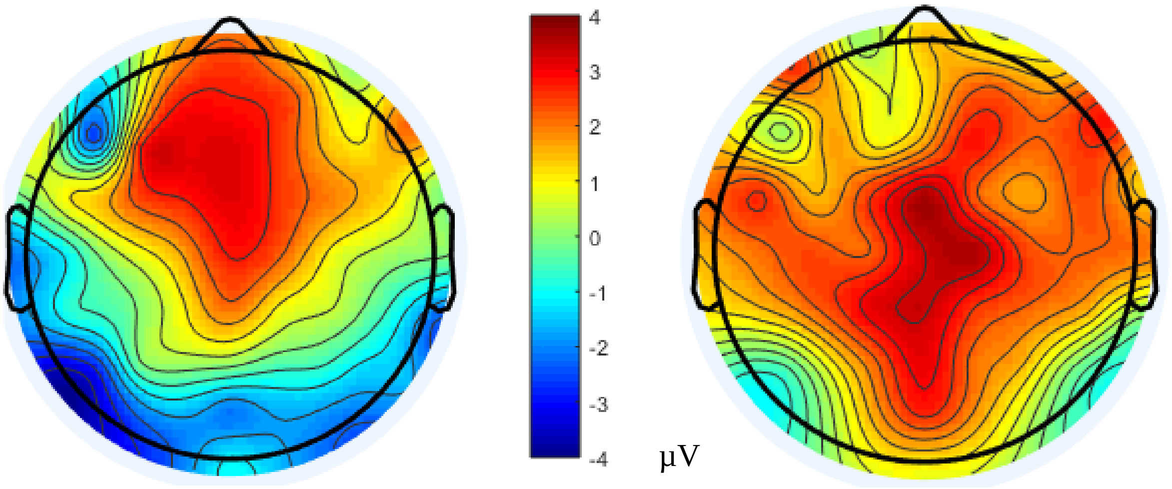
Out of the nine grand average ERP waveforms shown in figure 5.5, seven show a greater positivity in the 500-800 ms latency range in the non-collocational bigram condition compared to the collocational bigram condition. It is only the ERP waveforms representing mid anterior and mid central electrode sites that do not show this pattern.

The results of the repeated measures ANOVA reveal that, as with the ANOVA results for the N400 latency range, there is no interaction between condition and left-to-right electrode position but there is a significant interaction between condition and anterior-to-posterior electrode position, $F(2, 53) = 30.730$, $p < .001$. Follow-up analyses carried out separately on anterior, central, and posterior electrode sites reveal that the amplitude in the 500-800 ms latency range is lower in the non-collocational bigram condition compared to the collocational bigram condition at anterior electrode sites, yet it is significantly higher at central ($F(1, 321) = 14.271$, $p < .001$) and posterior ($F(1, 236) = 53.602$, $p < .001$) electrode sites. A summary of the results from these follow-up analyses is given in table 5.6.

Table 5.6: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 500-800 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Anterior	1.671	6.027	1.416	3.413	.318
Central	0.694	3.886	1.869	3.065	< .001***
Posterior	-1.598	3.677	0.362	3.569	< .001***

The higher amplitude at central and posterior electrode sites in condition 2 is illustrated using topographic scalp maps in figure 5.6.



Collocational condition

Non-collocational condition

Figure 5.6: Topographic scalp maps showing mean amplitude between 500 and 800 ms

Grand average ERP waveforms from each of the mid central and mid anterior electrode sites are shown in figure 5.7. At each displayed site, there is a greater positivity in the non-collocational condition compared to the collocational condition.

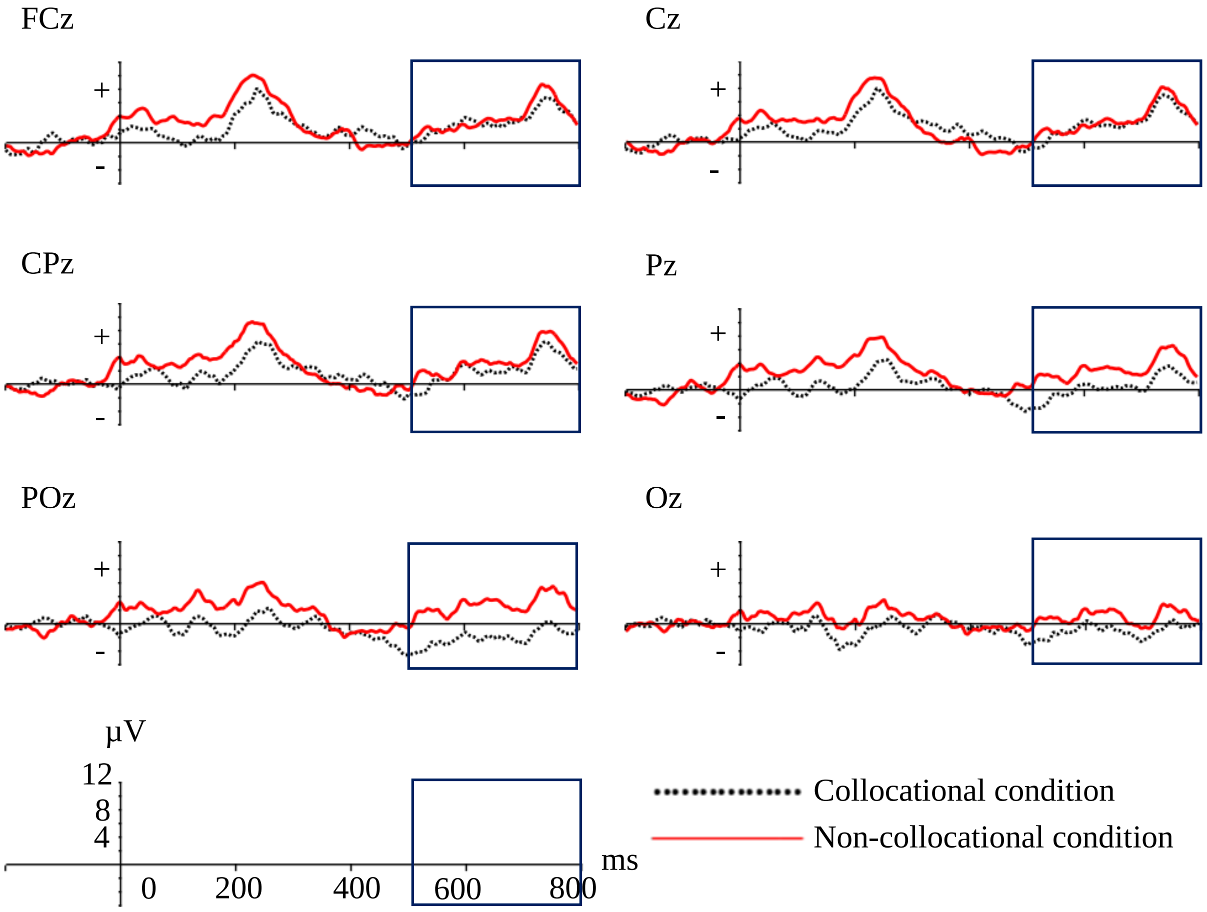


Figure 5.7: Grand average ERPs across mid-central and mid-posterior electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus.

To summarize this sub-section, there is a significantly higher amplitude in the non-collocational condition compared to the collocational condition at the typical latency range for the P600. At this latency range, the amplitude in the non-collocational condition is maximal over central-posterior scalp regions. As mentioned in Chapter 3 (section 3.4.4), Swaab et al. (2012:418) note that the P600 is largest over posterior scalp regions, and that it has no distinct laterality, while others report a central-posterior scalp distribution (e.g. Ingram 2007:323;

Friederici and Männel 2013:186). This therefore provides evidence in support of my hypothesis that reading the second word of a non-collocational bigram will elicit a P600 (see Chapter 3, section 3.4.7). The results of the N400 and P600 analysis are discussed in section 5.4.2.

5.4.1.3 Unexpected results

5.4.1.3.1 Overview

Although I planned to focus on just the N400 and the P600 in this pilot study, visual inspection of the ERP waveforms revealed conspicuous positive peaks which are noticeably bigger than the peaks for the N400 and the P600. The first of these peaks occurs at around 250 ms, so I labelled this the P250; the second unexpected peak occurs at around 750 ms, so I labelled this the P750. Note that, at this stage, the labels P250 and P750 are purely descriptive and do not have any theoretical significance. Thus, I am not implying that the P250 and P750 are distinct from other known components, or the same as known components which share these labels (though I will discuss how the P250 and P750 observed in my data relate to these known components in sections 5.4.4 and 5.4.5). Rather, they represent observed peaks in the waveforms that may or may not reflect specific neural or psychological processes and may or may not be a manifestation of known components. The P250 and P750 identified in this study can be seen in the grand average waveforms shown in figure 5.8. Both of these voltage deflections appear to have a higher amplitude in the non-collocational bigram condition compared to the collocational bigram condition.

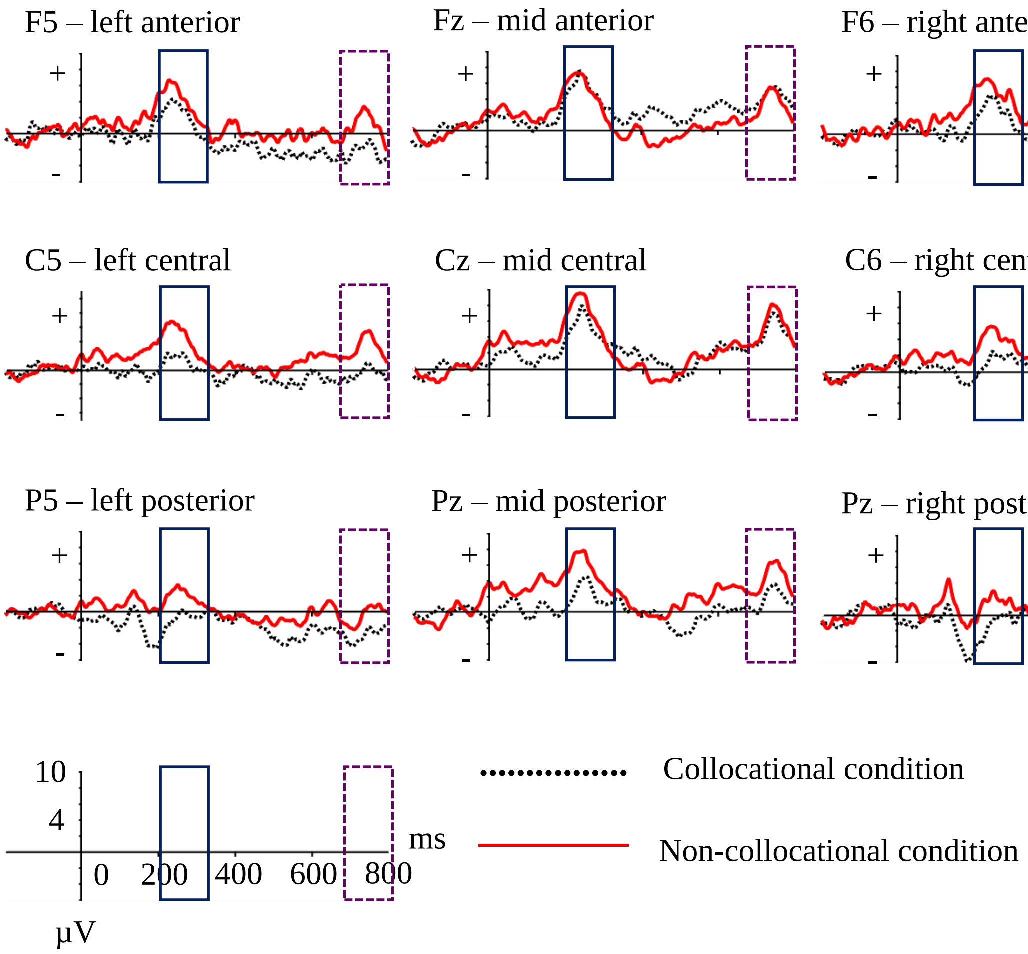


Figure 5.8: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus. The blue box with the solid outline illustrates the latency range of the P250 (200-300 ms); the purple box with the dashed outline illustrates the latency range of the P750 (700-800 ms).

Although I have described and labelled these unexpected voltage deflections in terms of their peak latency, I do not quantify the amplitudes of the P250 and P750 by focusing on their peaks. This is because, as discussed in Chapter 3 (section 3.4.8), it is highly problematic to use peak measures to quantify the amplitudes and latencies of EEG data. Rather, consistent with the analyses for the N400 and P600, I quantified the amplitudes using the *mean amplitude between two fixed latencies* function in ERPLAB. Following Marzi and Viggiano (2007:185), I use a 200-300 ms latency range for the P250; following van Boxtel (1998:88), I use a 700-800 ms latency range for the P750.

ERP studies can be criticised for carrying out this type of post hoc analysis for two main reasons. First, as mentioned in Chapter 4 (section 4.6.7), the components of interest should be determined *before* looking at the ERP waveforms in order to minimise the Type I error rate (Groppe et al. 2011:1711-1712; Keil et al. 2013:7). Second, Luck (2014a:312) advises that a given experiment should focus on a maximum of two components in order to minimise the experimentwise error rate (see section 4.6.7). Nevertheless, since any pilot study is by nature an exploratory study which aims to refine the hypotheses for the main study, I am arguably justified in carrying out these post hoc analyses in order to discover any potentially interesting results that I would have otherwise overlooked. These results would then need to be replicated in subsequent studies in order for them to be considered valid experimental ERP effects.

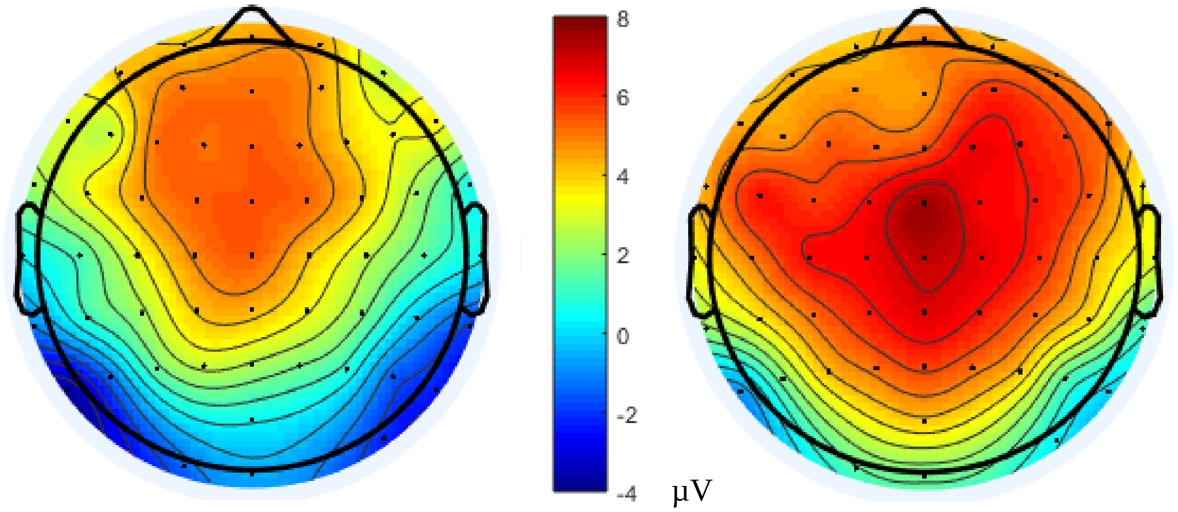
5.4.1.3.2 P250 (200 - 300 ms)

A comparison of the mean amplitude in the two conditions reveals that the positivity in the 200-300 ms latency range is higher in the non-collocational condition (M = 3.917, SD = 3.67) compared to the collocational condition (M = 2.115, SD = 4.689). This difference is statistically significant, $F(1, 1) = 69.425$, $p < .001$. Although there is no interaction between condition and left-to-right electrode position, there is a significant interaction between condition and anterior-to-posterior electrode position, $F(2, 2) = 12.860$, $p < 0.001$. Subsequent repeated measures ANOVAs reveal that the significant interaction occurs at central ($F(1, 1)$ = 37.618, p = < .001) and posterior ($F(1, 1) = 85.756$, $p < 0.001$) electrode sites (see table 5.7).

Table 5.7: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 200-300 ms latency range

Electrode Position	Collocational		Non- collocational		*p*-value
	M	*SD*	*M*	*SD*	
Anterior	4.567	4.264	4.953	2.719	.575
Central	2.851	4.144	4.95	3.333	< .001***
Posterior	-0.963	4.068	1.644	3.792	< .001***

The greater positivity at central and posterior electrode sites in the non-collocational bigram condition is illustrated in figure 5.9.



Collocational condition

Non-collocational condition

Figure 5.9: Topographic scalp maps showing mean amplitude between 200 and 300 ms

Grand average ERP waveforms for each of the mid central and mid posterior electrode sites are shown in figure 5.10. At each of these sites, the positivity in the 200-300 ms latency range is greater in the non-collocational condition compared to the collocational condition.

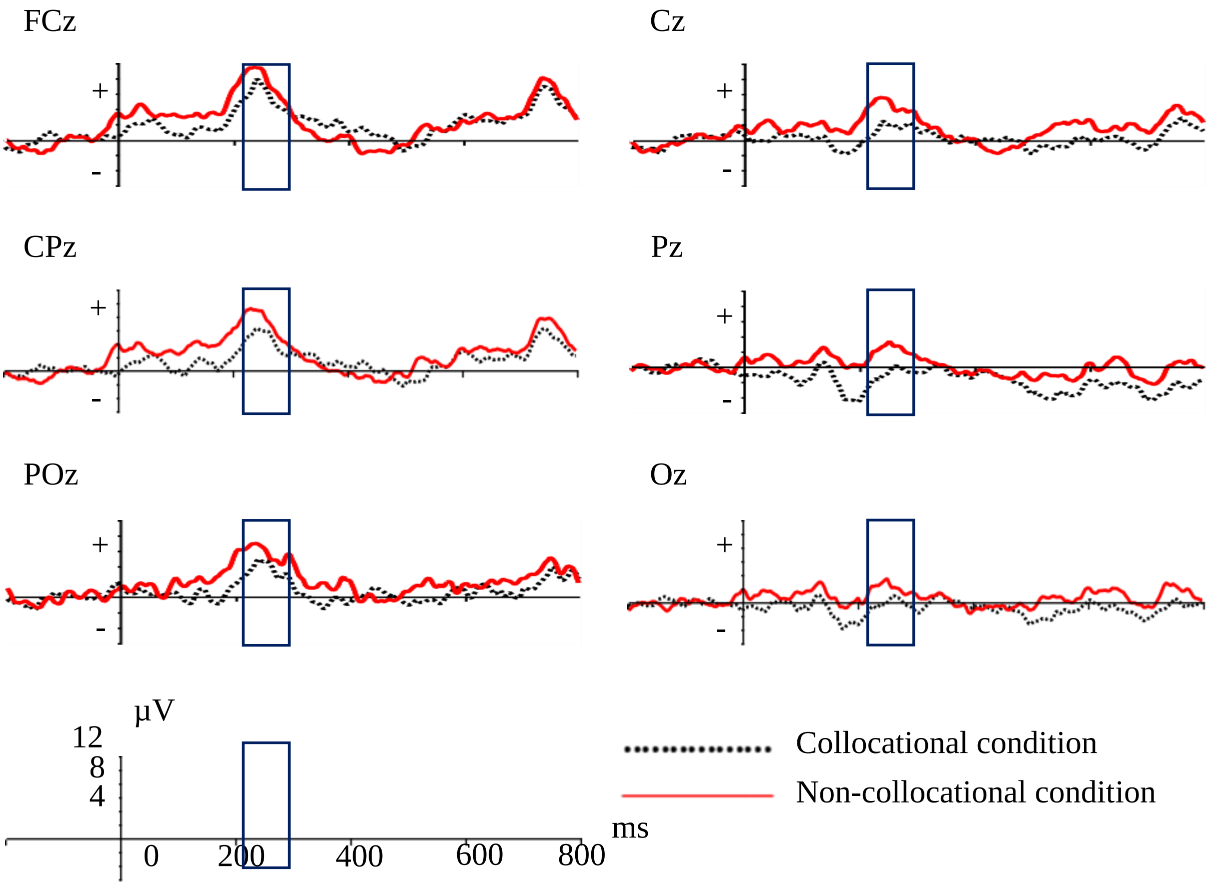


Figure 5.10: Grand average ERPs across mid central and mid posterior electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus.

To summarise, the results of the P250 analysis reveal that the mean amplitude in the 200-300 ms latency range is significantly higher for the non-collocational condition compared to the collocational condition. As with the P600, the P250 has a central-posterior scalp distribution. These results are discussed and critiqued in section 5.5.4.

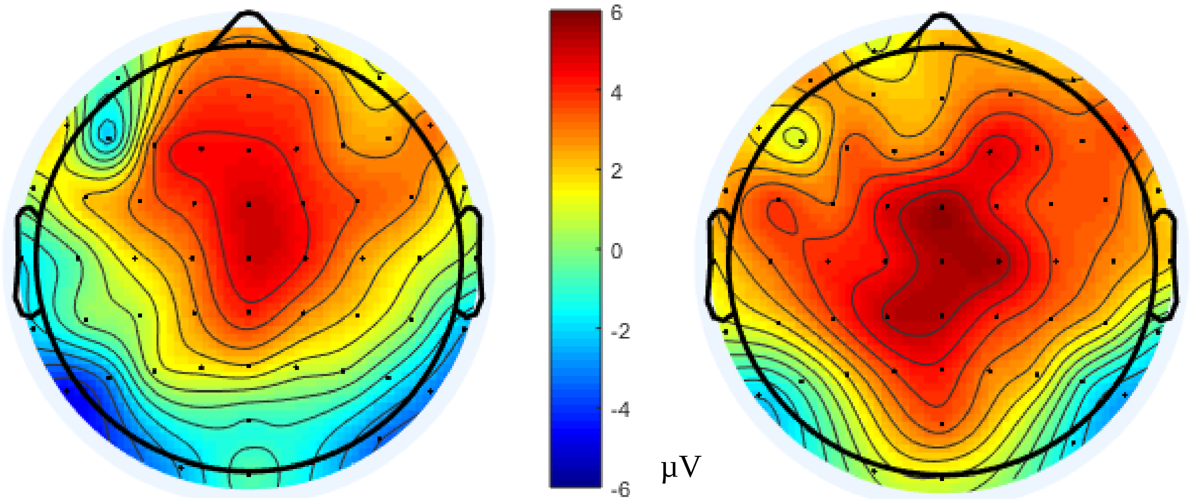
5.4.1.3.3 P750 (700 - 800 ms)

A comparison of the mean amplitude in the two conditions reveals that there is a higher mean amplitude in the 700-800 ms latency range in the non-collocational condition (*M* = 2.519, *SD* = 4.372) compared to the collocational condition (*M* = 0.932, *SD* = 5.764). This difference is statistically significant, *F*(1, 1) = 33.878, *p* < .001. As with the results for the N400, P600, and P250, there is no interaction between condition and left-to-right electrode position, but there is a significant interaction between condition and anterior-to-posterior electrode position, *F*(2, 2) = 4.388, *p* = .013. Table 5.8 shows that, although the mean amplitude is higher in the non-collocational condition at anterior, central, and posterior electrode sites, the difference between the two conditions is significant only at central (*F*(1, 1) = 21.596, *p* < .001) and posterior (*F*(1, 1) = 44.993, *p* < .001) electrode sites.

Table 5.8: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites at the 700-800 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Anterior	2.386	7.297	2.996	4.368	.491
Central	1.819	4.835	3.613	3.675	< .001***
Posterior	-1.500	4.586	0.637	4.642	< .001***

The higher amplitude at central and posterior electrode sites in the non-collocational condition is illustrated in figure 5.11.



Collocational condition

Non-collocational condition

Figure 5.11: Topographic scalp maps showing mean amplitude between 700 and 800 ms

Grand average ERP waveforms for each of the mid central and mid posterior electrode sites are shown in figure 5.12. At each of these sites, the positivity in the 700-800 ms latency range is greater in the non-collocational condition compared to the collocational condition.

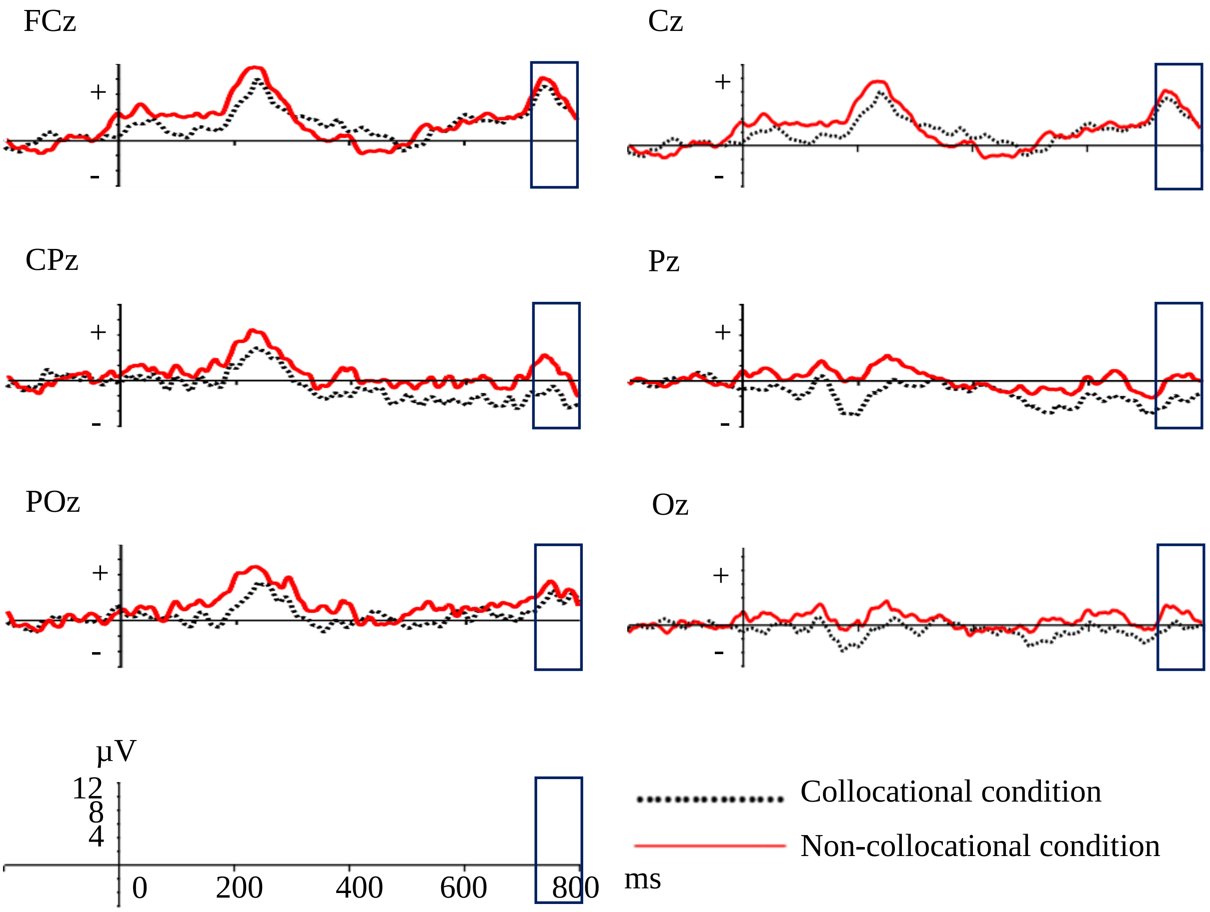


Figure 5.12: Grand average ERPs across mid central and mid posterior electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus.

To summarize this sub-section, the P750 is evident in both conditions yet the amplitude is significantly higher in the non-collocational condition. As with the P600 and P250, the P750 has a central-posterior scalp distribution. These results are discussed and critiqued in section 5.5.5.

5.4.2 Component-independent experimental design

The component-independent dimension of the analysis involves using the fractional peak latency function in ERPLAB in order to measure the mean onset latency from the grand average of each condition. As mentioned in Chapter 4 (section 4.6.7), the fractional peak latency function measures the time point at which the amplitude is half of the value of the peak amplitude in each condition (Luck 2014a:300).

A comparison of the means reveals that the onset latency is greater in the collocational condition (*M* = 364.165, *SD* = 237.152) compared to the non-collocational condition (*M* = 325.556, *SD* = 324.46). Furthermore, while the results of the repeated measures ANOVA reveal a significant main effect (*F*(1, 651) = 4.195, *p* < .05), there are no interactions between condition and left-to-right electrode position or anterior-to-posterior electrode position.

After closer inspection of the onset latency data, however, I became sceptical of those results. There were many instances across both conditions where the 50% peak latency occurred before 150 ms post stimulus. Luck (2014b:9) advises:

> In 99% of cases, you should not see any difference between waveforms prior to approximately 150 ms unless (a) there was a difference in the physical stimuli between conditions, or (b) the study involves shifts of spatial attention prior to stimulus onset.

Neither of these scenarios apply to my data. Therefore, I decided to repeat the onset latency analysis after excluding any cases where the 50% onset latency occurs before 150 ms post stimulus.

A comparison of the means from this second analysis once again reveals that the onset latency is greater in the collocational condition (*M* = 432.055, *SD* = 210.605) compared to the non-collocational condition (*M* = 364.552, *SD* = 227.361). In addition to the significant main effect (*F*(1, 483) = 16.866, *p* < .001), this subsequent analysis also reveals a significant interaction between condition and anterior-to-posterior electrode position, *F*(2, 483) = 4.845, *p* = .009. Follow-up ANOVAs reveal that the effect is maximal over central (*F*(1,215) = 21.854, *p* < .001) and anterior (*F*(1,144) = 6.402, *p* < .012) scalp regions (table 5.9).

Table 5.9: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Anterior	411.937	215.135	347.524	233.53	.012*
Central	442.517	220.07	339.62	220.564	< .001***
Posterior	437.377	187.345	427.073	221.605	.931

To summarize this sub-section, I have shown that the time point at which the amplitude is half of the value of the peak amplitude in the grand average waveforms occurs earlier in the non-collocational condition compared to the collocational condition. There is a significant difference between conditions. Thus, this demonstrates that there is a neurophysiological difference between conditions independent of the identification of specific ERP components. Furthermore, as mentioned in Chapter 4 (section 4.6.7), Rugg and Coles (1995:31) point out that, when interpreting the results from an onset latency measure:

> [i]t is important to note that the ERP difference only places an *upper* bound on the time by which processing is different. It is entirely possible that processing begins to differ at an earlier point in time, but that such a difference is not evident in the ERP.

This implies that the difference in onset latency between both conditions could occur earlier than the results of this pilot study suggest. Moreover, Luck (2014a:130) notes that “the onset latency represents the trials and subjects with the earliest onsets and not necessarily the average onset time”. This suggests that, for some participants, the onsets could have been much later than those reported in this section.

5.5 Discussion: Experiment 1

5.5.1 Summary of aims and results

This pilot study has investigated the neural correlates of collocation by comparing the ERP response to reading the second word of a collocational bigram to the ERP response to reading the second word of a non-collocational bigram. The overall aims of the pilot study were to find out whether or not the ERP response is different across the two conditions, and to refine my hypotheses and methods in preparation for Experiments 2 and 3. A difference between conditions can manifest itself as a difference in onset latency and/or as a difference in the presence or magnitude of specific ERP components. In this pilot study I set out to focus on onset latency as well as the N400 and P600 ERP components. The results of the onset latency analysis reveal that there is in fact a significant difference between the two conditions in terms of the time point at which the amplitude is half of its peak value. Thus, this component-independent analysis confirms my hypothesis that there is a neurophysiological difference in how collocational bigrams and non-collocational bigrams are processed.

The results of the component-based analysis provide further evidence in support of this hypothesis, while also supporting Hypotheses 2 and 3 (that reading the second word of a non-collocational bigram will elicit an N400 and a P600). In section 5.3.1.1, I show that there is a significantly lower amplitude in the non-collocational condition compared to the collocational condition at the typical latency range for the N400. Furthermore, in section 5.3.1.2, I show that there is a significantly higher amplitude in the non-collocational condition compared to the collocational condition at the typical latency range for the P600. The N400 observed in this pilot study has an anterior scalp distribution whereas the P600 has a central-posterior scalp distribution. These results are discussed and problematized in section 5.4.2.

Although I initially planned to focus on just two ERP components, namely the N400 and P600, visual inspection of the ERP waveforms directed my attention to two conspicuous peaks which occurred at around 200-300 ms post stimulus and 700-800 ms post stimulus. It is problematic to carry out this sort of post hoc analysis, for reasons laid out in section 5.3.1.3. Nevertheless, since this pilot study is an exploratory study which aims to refine the methods and hypotheses in preparation for Experiments 2 and 3, I decided to analyse the mean amplitudes within these time windows, which I labelled the P250 and P750. The results reveal that the mean amplitude of both the P250 and the P750 is significantly higher in the non-collocational condition compared to the collocational condition. Furthermore, as with the P600, the P250 and P750 observed in this pilot study have a central-posterior scalp distribution.

In subsequent sections of this discussion I compare the results of this pilot study to the results of previous studies which focus on the same components. I also discuss the implications of the results in terms of the different approaches to collocation and the network model of language processing introduced in chapter 2. Finally, I conclude the discussion by considering some of the limitations of this pilot study (section 5.5.6).

5.5.2 N400 and P600

In section 5.3.1.1, I showed that reading the second word of a non-collocational bigram elicits a significantly greater negativity than reading the second word of a collocational bigram at the typical latency range for the N400. Although the amplitude in this time window was found to be maximal at central-anterior electrode sites, rather than exhibiting the central-posterior distribution typical of the N400 (Swaab et al. 2012:399), I argued that my results provide sufficient evidence in support of the hypothesis that reading the second word of a non-collocational bigram elicits an N400. This is because, as noted by Kutas and Dale (1997:222), the scalp distribution of the N400 can vary even across similar experimental tasks.

It is also worth suggesting at this point that the central-anterior N400 found in this pilot study could actually be better described as the frontal N400 (or FN400) rather than the traditional semantic N400. The FN400 is a functionally distinct ERP component which is thought to be an electrophysiological marker of familiarity and recognition (Bridger et al. 2012:1340). If the results of the subsequent experiments also reveal an N400 which has this frontal (or central-anterior) scalp distribution, it will be worth further investigating this possibility that the putative N400 is actually the functionally distinct FN400.

In section 5.3.1.2, I showed that reading the second word of a non-collocational bigram also elicits a significantly larger P600 than reading the second word of a collocational bigram. Furthermore, the P600 observed in my pilot study exhibits the central-posterior scalp distribution that is typical of the P600s found in previous studies. The P600 is typically characterized as being sensitive to syntactic violations (Osterhout & Holcomb 1992:799) and to structural violations outside the domain of language (Núñez-Peña & Honrubia-Serrano 2004:133, 138; Patel et al.1998:724). As mentioned in Chapter 3 (section 3.4.7), the P600 has also been shown to be sensitive to predictability (Coulson et al. 1998b:34). The results of my pilot study seem to provide further evidence in support of this idea, as the P600 appears to be elicited in situations where a word is unexpected as a result of its low transition probability from the previous word.

However, in retrospect, I now realize that my choice of time windows invalidates the P600 results. I measured the mean amplitude of the P600 in the 500-800 ms latency range, as this is the latency range that is typically used in the study of the P600 (Tanner et al. 2015:1001). Yet this time window overlaps with the 700-800 ms time window that I used to measure the mean amplitude of the P750. Thus, the latency range used to measure the P600 actually encompasses both the P600 and P750. This is problematic because it makes it difficult to discern whether they are independent components, or whether the putative P600 is actually part of the P750, especially since the P600 and P750 share the same central-posterior scalp distribution. The P600 results are problematized further in section 5.5.5.

Looking back at figure 5.5, it is clear that the mean amplitude in the 500-800 ms latency range does not actually peak until around 750 ms. It should be noted that the P600 typically does not have a salient peak (Osterhout & Holcomb 1992:791). Yet, since the amplitude of the voltage deflection is greater at 750 ms, the voltage deflection that I have thus far referred to as the P600 is probably more likely to be the onset of the P750 (a putative component which will also be problematized in section 5.5.5) than an independent component. Thus, contrary to the claim in section 5.4.1.2, where I argued that there is a P600 in my pilot study data, I am now arguing that I do not have sufficient evidence to claim that there is a P600 in my data. In order to find out whether or not there is a P600 that is independent of the putative P750, I would need to exclude the later peak by measuring the mean amplitude within a tighter time window.

Although there is no evidence for a P600 in my pilot study data, the salient peak around 400 ms (e.g. see figure 5.1) suggests that there *is* an N400. As mentioned in Chapter 3 (section 3.4.3), classic studies show that the N400 is sensitive to semantic violations (Kutas & Hillyard 1984). Yet, as with the P600, more recent evidence suggests that the N400 is also sensitive to predictability (Coulson et al. 1998b; Laurent et al. 2006; Davenport & Coulson 2011). The results of my pilot study support this idea by showing that the N400 is sensitive to the transition probabilities between words.

Since my results show that the N400 is sensitive to transition probabilities, they constitute evidence in support of the network model of language processing outlined in Chapter 2 (section 2.8). In this model, words are represented as nodes, and the weighted connections between these nodes are strengthened whenever two words are encountered sequentially. The greater the level of prior exposure to the collocation, and therefore the greater the connection strength between the words, the more likely it is that the second word in the sequence will be mentally activated when the first word of the sequence is encountered.

This network assumption is explicit in Construction Grammar and in Hoey's (2005) theory of Lexical Priming, and to some extent also in Wray's (2002) description of formulaic language. Furthermore, the network assumption exists implicitly in the syntax-based approach to collocation and in Sinclair's (2004) discussion of lexical units. Therefore, by providing evidence in support of the network model of language processing, the results of my pilot study also provide evidence in support of these approaches to the study of collocation.

As discussed in Chapter 2 (section 2.9), Sinclair (1991:114) and Wray (2002:10) both argue for the existence of a separate grammatical system alongside a system which processes language using collocational mechanisms. By contrast, the theories of Lexical Priming, Pattern Grammar, and Construction Grammar all attempt to account for the whole of language without positing the existence of a separate grammatical system. The results of my pilot study do not provide any support for the idea that there is a separate grammatical system. After all, if entirely different components were elicited in response to reading collocational bigrams compared to non-collocational bigrams, this *qualitative* difference would provide evidence to suggest that different processing systems are involved in the processing of collocational bigrams compared to the processing of non-collocational bigrams. However, my data only reveals a *quantitative* difference across conditions, as the same ERP components are present in each condition, but with varying magnitudes.

It is important to note that the possibility of there being a separate grammatical processing system cannot be ruled out, as it could be the case that the brain activity associated with this system was not detected by the electrodes. Similarly, although the data does not provide definitive evidence to suggest that the whole of language can be accounted for using collocational mechanisms, there is also no evidence to suggest that this is not the case.

Finally, it is unsurprising to see that the N400 is elicited in response to both conditions because it is widely known that some N400 activity is elicited in response to encountering any content word (Kumar & Debruille 2004). Furthermore, since I have shown that the N400 is modulated by word predictability, I would expect an N400 to be elicited in both conditions because even bigrams that have a very high transition probability are not entirely predictable. For a bigram to be entirely predictable, and thus presumably not elicit an N400, it would need to have a transition probability of 1. The bigram with the highest transition probability in my pilot study is *nineteenth century*, which has a transition probability of 0.855, followed by *endangered species*, which has a transition probability of 0.683. These are very high transition probabilities but, as mentioned in Chapter 3 (section 3.4.6), even in supposedly fixed multi-word expressions, there are still options as to what word could come at each position in the sequence.

5.5.3 Onset latency

Onset latency is thought to vary as a function of the complexity or salience of the experimental stimuli. For example, Hoehl et al. (2008:14) suggest that a smaller latency is associated with lower stimulus complexity, and therefore "faster and more efficient processing". Webb (2005:606) states that "[t]he latency of a component is thought to reflect the timing of the neuronal activation related to a particular process". Likewise, through triangulation with behavioural measures, it has been shown that a smaller ERP latency is associated with a faster response time (Osman et al. 2000; van der Lubbe et al. 2001:254; Rinkenauer et al. 2004:269).

In this pilot study, I found that the onset latency is smaller in the non-collocational condition compared to the collocational condition. This is highly unexpected. Based on the results of the aforementioned studies, I would expect to find that the onset latency is smaller in the collocational condition. This is because the collocational condition contains more expected, predictable sequences of words, which I assumed would be processed more easily and quickly than the non-collocational bigrams, which contain less expected, less predictable word sequences. It will be interesting to see if these results are replicated in subsequent experiments.

5.5.4 P250

In section 5.4.1.3.2, I showed that a central-posterior P250 is elicited in both conditions, but the amplitude of the P250 is significantly greater in the non-collocational condition compared to the collocational condition. The study of the P250 is not widespread in the ERP literature on the processing of language, and the language processing studies which do focus on the P250 do not seem relevant to the interpretation of my pilot study results. For example, in an ERP study on Chinese morphology, Chung et al. (2010) find that the amplitude of a frontal P250 is significantly greater in response to reading compound words which have the same internal structure compared to reading compound words which have a different internal structure. Nevertheless, although the ERP literature on the processing of language does not provide any insights as to how my pilot study results could be interpreted, interesting comparisons can be made between the P250 found in my pilot study and the P250 found in the ERP literature on face recognition.

Butler et al. (2013) state that there are four distinct stages involved in face recognition that can be detected through ERPs. The first stage involves the initial detection of facial features such as the eyes, nose, and mouth, and this is reflected by the P100. This is then followed by the N170 (mentioned in Chapter 3, section 3.4.2) which reflects the integration of the different facial features into a coherent representation of a face. Then, before the participant realizes which person from their memory matches with the displayed face, which is reflected by the N400, there is a matching stage which involves comparing the newly constructed representation of a face with faces that have already been stored in memory. It is at this matching stage when the P250 is elicited.

One study which demonstrates the elicitation of the P250 during face recognition is carried out by Caharel et al. (2002). In this study, 11 participants are presented with three different types of faces: an unknown face, a famous face, and the participant's own face. The aim of this study was to find out whether or not any of the components elicited during face recognition are modulated by familiarity. The results show that the P250[8] is modulated by familiarity, as the amplitude of the P250 is largest for unknown faces and smallest for the participant's own face (Caharel et al. 2002:1500, 1507). Caharel et al. (2002:1510) provide the following interpretation of their results:

> The smaller amplitude observed for the subject's own face relative to the famous one may correspond to a lesser demand of attentional resources … In a similar vein, the larger amplitude for the unknown face than the familiar ones could reflect a more extensive search in memory.

A comparable interpretation could be applied to the results of my pilot study. It could be the case that the P250 is elicited in response to reading the second word of a collocational bigram because it reflects a memory search for this bigram. Furthermore, the greater amplitude of the P250 in the non-collocational condition could be attributed to the presumably more extensive memory search that would be required for an unfamiliar word pair.

Other face recognition studies provide further evidence in support of this interpretation. In a study carried out by Butler et al. (2013), 20 participants are presented with three types of photographs: photographs of themselves, photographs of their dizygotic (i.e. non-identical) twin, and photographs of an unfamiliar person matched for age and gender. The results show that the amplitude of the P250 is significantly larger when the participants are viewing photographs of an unfamiliar person compared to viewing photographs of themselves or their

---

[8] Note that Caharel et al. (2002) actually refer to this component as the P2 but they measure this using the same time window (200-300 ms) that I use for the P250 in my pilot study. For consistency I therefore refer to their P2 as the P250.

twin. The larger amplitude of the P250 in response to viewing photographs of an unfamiliar person is comparable to the larger amplitude of the P250 in response to reading an unfamiliar bigram in my pilot study.

While Caharel et al. (2002) find that the amplitude of the P250 is different when participants view their own face compared to when they view the face of another person, Butler et al. (2013) find that there is no significant difference in the amplitude of the P250 when a participant views their own face compared to when they view their twin’s face. Butler et al. (2013) attribute this discrepancy between their results and the results of previous studies to the fact that they controlled for exposure: a participant is likely to have had daily exposure throughout their lives to their own face and their twin’s face, but much less exposure to the face of a famous person. This idea that the P250 is modulated by exposure is relevant in the context of my pilot study, because the participants will have had much more exposure to the collocational bigrams compared to the non-collocational bigrams, which do not occur in the BNC1994 at all and can therefore be assumed to be very rare in English generally.

A final point to note is that the P250 found in the study by Butler et al. (2013) exhibits a posterior scalp distribution, and the P250 found in the study by Caharel et al. (2013) exhibits a central-posterior scalp distribution. These results are comparable with the results of my pilot study, which show that the P250 is maximal over central-posterior electrode sites. This suggests that the P250 observed in my pilot study may be the same component as the P250 observed in ERP studies of face recognition.

Alternatively, it could be argued that the putative P250 in this pilot study is actually an early P300[9] or, more specifically, the P3a. This makes sense both functionally, in terms of the

---

[9] Note that the P250 is often considered to be part of the P300 family (see, for example, Verleger 1988), so the discussion of the P250 literature is still relevant even if the component is interpreted as the P300.

cognitive processes that the P3a is typically thought to reflect, and also temporally, in terms of the latency of the component. Temporally, the P3a is a member of the P300 family that peaks earlier than other members; it tends to peak between 250 and 280 ms (Rotschafer & Razak 2014:21), and this is therefore directly in line with the latency of the component under discussion.

Similarly, as mentioned in Chapter 3 (section 3.4.7), the P3a is typically elicited in an auditory oddball paradigm where the participants are *not* explicitly told to attend to the low probability stimulus items. This makes sense in the context of my experiment because the participations are not given any indication of the presence of the non-collocational bigrams, or indeed the focus on the probabilities between words. Interestingly, without initially realising that I had done so, I had essentially created a semi-oddball paradigm because non-collocations are not normative and are therefore surprising. Thus, this suggests that, regardless of whether this ERP component evident in the 200-300 ms latency range is considered to be the P250 or the P3a, the results still provide evidence in support of the idea that the brain is sensitive to the transition probabilities between words. However, it is important to remember that members of P300 family are typically only considered to be sensitive to *task-relevant* probabilities (Donchin & Coles 1988:367), i.e. probabilities that are developed *within* the experimental task rather than probabilities which exist in the real world. As mentioned in section 3.4.7, this suggests that the P300 is unlikely to be sensitive to transition probabilities, as these are based on how the words pattern throughout the whole corpus, rather than in the context of the experimental task.

5.5.5 P750

In section 5.4.1.3.3, I showed that a central-posterior P750 is elicited in both conditions, but the amplitude of the P750 is significantly greater in the non-collocational condition compared to the collocational condition. The P750 is not a widely recognized component in the ERP literature, and the studies that do explicitly focus on the P750 are not entirely comparable as they use a very different latency range to the 700-800 ms range used in this pilot study. For instance, Mai et al. (2012:405) use a latency range of 500-1000 ms when measuring the P750.

The results of my pilot study seem as though they should be comparable with the results of studies which focus on the *late positive component* (LPC) or other “very late” components (Richards 1977:86). However, there is terminological inconsistency in the ERP literature on what constitutes a late positive component as opposed to a very late positive component. As mentioned in Chapter 3 (section 3.4.2), the LPC is another label for the P300 (Polich 2007:2128). Although 300 ms after stimulus onset seems to be a relatively early neural response, this labelling convention reflects the fact that, after the initial studies on the P300 component, it was found that the latency of this component is often more in the range of 600-800 ms (Johnson 1995:146). However, in the same article that describes this discrepancy between the label and the time range, Johnson (1995:138) explicitly refers to the P300 as an “early” component. It would therefore be misleading to label the P750 as a late positive component if this label is conventionally used for the P300, a component which does sometimes reflect an early neural response.

Moreover, Richards (1995:86) states that a component which occurs later than 700 ms after stimulus onset is considered to be a “very late” component. However, studies which focus on the very late positive component differ as to what latency range they use, with some researchers considering latencies as early as 450 ms to be “very late” (e.g. Donald 1983:61). Because of this, studies which focus on the very late positive component are not automatically comparable with my pilot study. I therefore only compare my results with the results of other studies which specifically use a 700-800 ms latency range.

One such study, carried out by Xue et al. (2015), involves 20 Chinese students participating in a modified face-word Stroop task. In this task, there is an incongruent condition where images of happy faces are displayed with a Chinese word meaning ‘fearful’, and images of fearful faces are displayed with a Chinese word meaning ‘happy’. There is also a congruent condition, where the meaning of the adjective is congruent with the emotion portrayed by the image. Participants are instructed to ignore the meaning of the words, and instead focus on identifying the emotions portrayed by the faces. The results of this study show that the amplitude in the 700-800 ms latency range is significantly larger in the incongruent condition compared to the congruent condition. Furthermore, the amplitude in this latency range is maximal over posterior electrode sites.

Xue et al. (2015) identify this as a conflict-sensitive ERP component which, in the context of their study, reflects the identification of facial expressions and the inhibition of the semantics of the interfering word. It could therefore be the case that, in my pilot study, the amplitude of the P750 is largest in the non-collocational condition because it reflects the inhibition of the collocational word(s) that the participants would have been expecting to follow the first word of the bigram, in favour of the non-collocational word which actually did follow the first word of the bigram. This comparison is reinforced by the similar scalp distributions found for the posterior conflict-sensitive ERP component in Xue et al.’s (2015) study and the central-posterior P750 found in my pilot study.

Specifically language-related evidence in support of the idea that this component is conflict-sensitive comes from an ERP study carried out by Kielar et al. (2012). In this study, participants with and without agrammatic aphasia were instructed to read sentences that either did or did not contain an argument structure violation. Specifically, 35 sentences contained a transitive verb with the correct number of arguments (e.g. *Anne visited the doctor and the nurse*), and 35 sentences contained an intransitive verb incorrectly followed by arguments (e.g.

**Anne sneezed the doctor and the nurse*) (Kielar et al. 2012:3323-3324). The results show that reading the sentences which contain an argument structure violation elicits a positive potential in the 700-800 ms latency range for participants without agrammatic aphasia, but not for participants with agrammatic aphasia. This implies that, in normal language processing, there is an increased positivity in the 700-800 ms latency range in response to encountering an argument structure violation. It is interesting to note here that these argument structure violations are essentially identical to violations in colligational patterns, i.e. the co-occurrence of grammatical categories (Firth 1957:13). Thus, it can be said that the increased positivity in the 700-800 ms latency range occurs in response to reading colligational errors.

It is possible that this increased positivity could be attributed to the conflict between what the participant is expecting to follow the verb and what actually follows the verb. Likewise, as mentioned earlier, the larger P750 in the non-collocational condition in my pilot study could be attributed to the conflict between what the participant is expecting to follow the first word of the bigram and what actually follows this word. The similarity between the P750 observed in my study and the positivity in the 700-800 ms latency range observed by Kielar et al. (2012:3326) is heightened by scalp distribution data, as this component exhibits a central-posterior scalp distribution in both studies. Thus, it is plausible to suggest that the P750 is elicited in situations where conflict arises due to an expectation not being met.

An alternative explanation for the presence of this positive ERP component in the 700-800 ms latency range is that it could actually be the P3a of the next word in the sentence. This is plausible because this component occurs roughly 500 ms after the P3a/P250, and each word in the experimental sentences is presented for 500 ms. This suggests that the putative P750 is actually a consequence of the task design, rather than being an interesting experimental effect.

Furthermore, another plausible explanation is that the putative P750 could be caused by saccadic (or microsaccadic) activity. Saccades are defined as “rapid movements of the eyes with velocities as high as 500° per second” (Rayner 1998:363). Although each word in the stimuli was presented in the centre of the computer screen, and the sentences were punctuated with fixation crosses, saccadic activity would still have been present. This is because saccades (or more specifically, microsaccades) are present during periods of fixation (Engbert & Kliegl 2004:431); the eyes are never completely still (Rayner 1998:373). According to Marton and Szirtes (1982:169), the offset of a saccade elicits a P3. Therefore, it is plausible to suggest that the putative P750 is actually a P3 resulting from saccadic activity on the previous word.

Due to the uncertainty in how the P250 and P750 results should be interpreted, I decided not to explore the P250 and the P750 further in subsequent experiments. Instead, I decided to focus on the N400 and P600, as these components are central to the ERP study of language. I also decided to continue focusing on onset latency, as this is a vital measure enabling me to unambiguously detect whether or not there is a neurophysiological difference between conditions, without having to rely solely on the results from specific ERP components.

5.5.6 Limitations of the pilot study

The limitations associated with the ERP technique are discussed in Chapter 3 (section 3.4.8), yet there are some additional limitations which are specific to this pilot study. First, the pilot study is limited by the small number of trials. As mentioned in Chapter 4 (section 4.2.1), the stimuli set for the pilot study contains just 30 bigram pairs (i.e. 15 trials in each condition), but prior work on large language-related ERP components such as the N400 tends to use at least 30 trials in each condition (Luck 2014a:262). Indeed, looking back at the ERP studies of collocation/predictability listed in table 3.4 of Chapter 3 (section 3.4.7), there is one study which uses 20 trials per condition, but all others use at least 30. I will therefore double the number of trials for subsequent experiments (see Chapter 6, section 6.3.3). This will increase the signal-to-noise ratio in the data (following Luck 2005:30) which, in turn, should help to decrease the high attrition rate of the pilot study, where 1 in 4 datasets are unusable due to artifacts.

Second, although I put the stimuli into 4 differently ordered lists in order to minimise any order effects (see Chapter 4, section 4.2.2), the fact that the data for 4 out of the 16 participants could not be included in the analysis means that the data that *is* used in the analysis is not fully counterbalanced. Out of the 4 participants whose data was excluded, 2 read List C, 1 read List B, and 1 read List D. I therefore had proportionately more data for list A, where the collocational items came before the non-collocational items. This could have biased my results because, after reading a collocational bigram, the second word of the matched non-collocational bigram might seem even more unexpected than if the well-formed bigram had not been presented first. Thus, for subsequent experiments, I will continue recruiting participants until I obtain an equal number of usable datasets for each stimuli list.

A third problem with my pilot study is that the latency ranges for the P600 (500-800 ms) and the P750 (700-800 ms) overlap. This makes it difficult to discern whether they are fully independent components or whether the P750 is actually a part of the P600 (or vice versa), especially since the P600 and P750 share the same central-posterior scalp distribution. Moreover, since I have established that the P750 might actually be the P3a of the next word (see section 5.4.5), and that this component in the 700-800 ms latency range is difficult to distinguish from the P600, this suggests that the P600 found in the pilot study might actually be the P3a of the next word. In subsequent experiments, I will exclude the later peak by using a tighter time window to measure the P600 (see Chapter 6, section 6.3.5).

A final problem with the pilot study is the fact that each filler item contained a collocational bigram (see section 5.2). By having a collocational bigram in the filler items as well as the experimental items, this would have increased the probability of encountering a collocational bigram within the task as opposed to a non-collocational bigram. Over the course of the experiment, each participant would have read 28 sentences containing a collocational bigram (15 collocational sentences and 13 fillers), but only 15 sentences containing a non-collocational bigram. Thus, I unintentionally created an oddball paradigm, with the non-collocational items being the oddballs. As mentioned in Chapter 3 (section 3.4.7), the P300 is modulated by task-relevant probability (Donchin & Coles 1988:367). Therefore, since the non-collocational bigrams were comparatively infrequent and therefore low-probability, the presence of the collocational bigrams in the filler items probably contributed to the amplitude of the P300.

Ideally, I want to minimise the amplitude of the P300 as much as possible. This is because having a positive voltage deflection in a similar latency range to the N400 could mask the negative voltage deflection of the N400, making it more difficult to find an experimental effect. This is known to be a common problem in N400 studies; indeed, as Roehm et al. (2007:1260) point out, the existence of a P300 leads to an ‘‘extremely reduced’’ N400 effect. In subsequent experiments, I therefore removed the filler items to ensure that the stimuli consisted of an equal number of sentences with collocational and non-collocational bigrams (see Chapter 6, section 6.3.1).

That said, it is important to note that, even when there are an equal number of sentences with collocational and non-collocational bigrams, there are still many more collocational bigrams compared to non-collocational bigrams due to the fact that all of the other words in the experimental sentences are collocational to some extent (or, at least, not *non*-collocational). This means that the non-collocational bigrams have a low task-defined probability, and will therefore elicit a P300. To solve this problem I would need to present the experimental bigrams in isolation, rather than as part of a sentence. However, this would go against my intentions for this thesis: to investigate the neurophysiological responses to reading collocational and non-collocational bigrams in the most naturalistic reading context that is possible in an ERP experiment (see Chapter 1, section 1.4). Presenting the bigrams in isolation could potentially cause participants to draw on reading strategies that they would not typically use in normal reading. Therefore, in order to maintain the naturalistic reading context, I had to accept that the resulting P300 would dilute any N400 effect found in my experiments.

5.6 Chapter summary and conclusion

In this pilot study, I have investigated the neural correlates of collocation by comparing the ERP response to reading the second word of a collocational bigram to the ERP response to reading the second word of a non-collocational bigram. The aims of the pilot study were (1) to pilot a procedure for determining whether or not there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams, and (2) to refine the hypotheses and methods in preparation for Experiments 2 and 3.

The results of the onset latency analysis reveal that there is in fact a significant difference between the two conditions in terms of the time point at which the amplitude is half of its peak value. Thus, this component-independent analysis confirms my hypothesis that there is a neurophysiological difference in how collocational bigrams and non-collocational bigrams are processed.

The results of the component-based analysis provide further evidence in support of that hypothesis. In Chapter 3 (sections 3.4.6 and 3.4.7), I hypothesized that reading the second word of a non-collocational bigram would elicit an N400 and a P600. Evidence in support of this hypothesis is given in section 5.4.1.1, where I show that there is a significantly lower amplitude in the non-collocational condition compared to the collocational condition at the typical latency range for the N400; and also in section 5.4.1.2, where I show that there is a significantly higher amplitude in the non-collocational condition compared to the collocational condition at the typical latency range for the P600.

Although I initially planned to focus on just the N400 and P600, visual inspection of the ERP waveforms directed my attention to conspicuous peaks which occur around 200-300 ms post stimulus and 700-800 ms post stimulus. I decided to analyse the mean amplitudes within these time windows, which I labelled the P250 and P750. The results reveal that the mean amplitude of the P250 and P750 is significantly higher in the non-collocational condition compared to the collocational condition. Furthermore, as with the P600, the P250 and P750 observed in this pilot study have a central-posterior scalp distribution.

The results of this pilot study therefore provide evidence in support of my hypothesis that there is a neurophysiological difference in how collocational bigrams and non-collocational bigrams are processed. In the following chapter, I focus on Experiment 2, which essentially replicates this pilot study with some minor methodological changes in order to provide stronger evidence in support of the aforementioned hypothesis.

CHAPTER 6: EXPERIMENT 2 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in native speakers of English

6.1 Overview of the chapter

In this chapter I focus on Experiment 2, which investigates the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in native speakers of English. In section 6.3 I detail the aims and hypotheses of Experiment 2. In section 3 I describe the changes that I have made to the methodology in light of the outcomes of the pilot study. In section 4 I then present the results of Experiment 2, first from the component-based experimental design (section 6.4.1) and then from the component-independent design (section 6.4.2). Finally, I discuss the results of Experiment 2 in section 6.5 before providing a chapter summary and conclusion in section 6.6.

6.2 Aims and hypotheses: Experiment 2

The aim of this experiment is to replicate the results from the pilot study in order to provide stronger evidence in support of the idea that there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams. Specifically, I aim to replicate the finding that a larger N400 is elicited in response to reading non-collocational bigrams compared to collocational bigrams. I also aim to replicate the finding that there is a difference in onset latency between collocational and non-collocational bigrams. In addition, I aim to find out whether or not a P600 is elicited in response to reading non-collocational bigrams, as the results regarding this component were inconclusive in the pilot study.

Based on the results of Experiment 1, and on the results from the ERP literature, I make the following hypotheses:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will elicit a P600.

Hypothesis 3: The onset latency will be greater in the collocational condition compared to the non-collocational condition.

6.3 Methodology: Experiment 2

6.3.1 Participants

The experiment was initially carried out on 16 native speakers of English. In the pilot study, 4 out of 16 datasets were excluded from the analysis due to excessive artifacts, leaving just 12 datasets for data analysis (see Chapter 5, section 5.2). However, if a dataset needed to be excluded from the analysis in Experiment 2 (specifically, if over 25% of the trials were rejected due to artifacts), I would recruit an additional participant so that the excluded dataset could be replaced. Overall, I carried out the experiment on 21 participants in order to have 16 usable datasets.

All 21 participants had normal or corrected-to-normal vision and no (history of) neurological disorders. All participants were students at Lancaster University, and were recruited in line with the ethics procedures of the Department of Linguistics and English Language (see appendices 5 and 6 for information sheet and consent form). No participants had taken part in Experiment 1.

6.3.2 Stimuli

The experimental sentences and true/false statements used in this experiment are exactly the same as those used in Experiment 1 (see table 5.3). The same practice sentences were also used (table 5.4). However, whereas 13 filler sentences were used in Experiment 1, no filler sentences were used in Experiment 2. As mentioned in Chapter 5 (section 5.5.6), the fact that each filler contained a collocational bigram probably contributed to the amplitude of the P300, because this component is modulated by task-defined probability (Donchin & Coles 1988:367). Over the course of the experiment, each participant would read 28 sentences containing a collocational bigram (15 collocational sentences and 13 fillers) but only 15 sentences containing a non-collocational bigram. Thus, a P300 would be elicited by reading the non-collocational bigrams because they are infrequent and therefore surprising in the context of this experiment. Ideally, I want to minimise the amplitude of the P300 as much as possible, because having a positive voltage deflection in a similar latency range to the N400 could mask the negative voltage deflection of the N400 (Roehm et al. 2007:1260), making it more difficult to find an experimental effect. Removing the fillers helps to achieve this aim by ensuring that the stimuli consist of an equal number of sentences with collocational and non-collocational bigrams. See appendix 2 for the four counterbalanced lists of experimental stimuli.

6.3.3 Procedure

As mentioned in Chapter 5 (section 5.5.6), a limitation of the pilot study was that there were too few trials. In Experiment 2, I therefore decided to double the number of trials per condition from 15 to 30 in order to increase the signal-to-noise ratio in the data (following Luck 2005:30). This, in turn, should help to decrease the high attrition rate of Experiment 1, where 1 in 4 datasets were unusable due to artifacts. This change is also in line with Swaab et al.'s (2012:428) advice that language-based ERP studies should have a minimum of 25 trials in each condition, as well as Luck's (2005:30) recommendation that ERP studies which focus on large components such as the N400 should have a minimum of 30 trials in each condition.

Ideally, I would have liked to have created 15 entirely new experimental items by extracting 15 more collocational adjective-noun bigrams from the BNC1994. However, due to the difficulty in finding the original 15 bigrams, I felt that it would have been an impossible task to find 15 more collocational bigrams which fit the requirements of the study, i.e. collocational bigrams that fit the fluid conceptualization of collocation outlined in Chapters 1 and 2, that occur in at least 5 different texts, that do not form part of a longer collocation, that can be matched for frequency and length with a semantically plausible non-collocational bigram, and that can be embedded into semantically coherent sentences that allow for an equal contextual constraint in both conditions (see Chapter 4, sections 4.2.2 and 4.2.3, for full specification).

With this in mind, Reid (2016, personal communication) recommended repeating the same stimuli twice, and confirmed that participants would not become habituated to linguistic stimuli after just two presentations (and so the repetition of stimuli would *not* inadvertently cause repetition priming). Therefore, in Experiments 2 and 3, after participants had been presented with the full list of 30 experimental sentences (and corresponding true/false statements), they were then presented with the same list again.Increasing the total number of trials from 30 to 60 inevitably increased the length of the experiment, to approximately 20 minutes. To account for this, I included 7 breaks in the experiment. The positioning of breaks within the structure of the experiment is shown in figure 6.1.

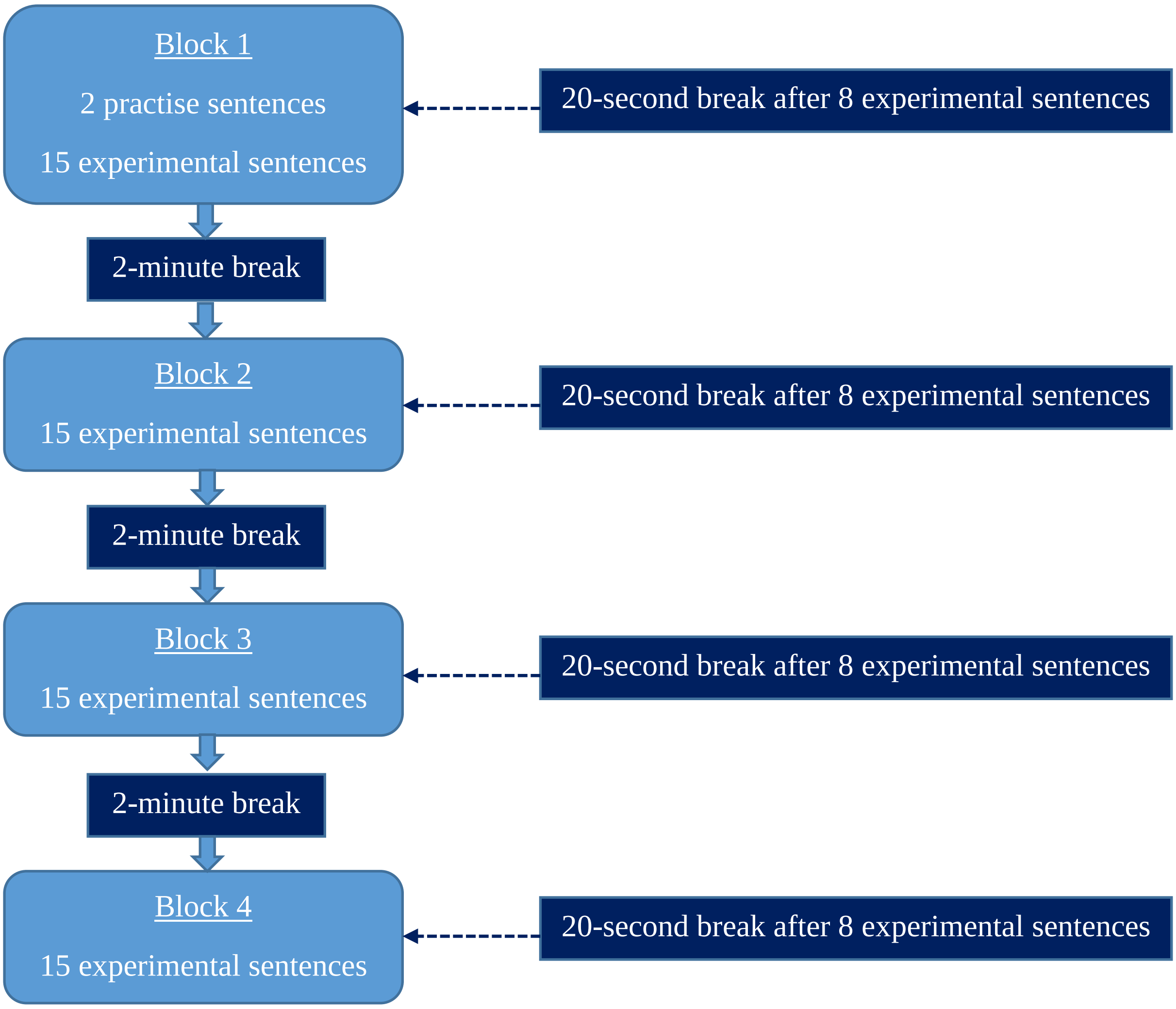


Figure 6.1: Positioning of breaks within and between trial blocks

The structure of the experiment thus complies with Luck's (2014c:2-3) guidelines that each block of trials should last for 5 minutes, with a 2-minute break *between* blocks, as well as a 20-second break *within* each block. The 2-minute breaks were experimenter-controlled, meaning that the participant had to wait for the experimenter to open the next trial block before they could continue with the experiment. The 20-second breaks were automated; the stimulus display counted down from 20 to 1, so that the participants were aware of how many seconds remained before they would need to be ready to continue. At the end of this 20-second break, the participants would be prompted to press the spacebar to continue the experiment. This ensured that the participants were focusing on the stimulus display before being presented with any more sentences.

It is important to note that 20 minutes is very short for a typical ERP experiment. However, this length of time is actually standard for an ERP experiment which focuses on collocation. For instance, the Lau et al. (2016) collocation experiments mentioned in Chapter 3 (sections 3.4.6 and 3.4.7) lasted 20-25 minutes and 10-15 minutes. Similarly, the Siyanova-Chanturia et al. (2017) experiments lasted 20 minutes and 15 minutes. I also think that having longer experimental sessions would have decreased the quality of the data, as I repeatedly observed an increase in alpha activity in Block 4 of my datasets. Alpha activity is associated with tiredness/fatigue and lack of concentration (Tiago-Costa 2016:90).

A final change that I made to the stimulus presentation procedure related to the length of time that each word was presented for. In Experiment 1, each word was presented for 500 ms and was immediately replaced by the next word; in Experiment 2, the stimulus presentation period was still 500 ms per word overall, but the word itself was presented for just 300 ms, and this was followed by a blank white screen presented for 200 ms. This is the more common method of stimulus presentation in language-based ERP studies, because the presence of the interstimulus interval reduces the overlap effects caused by continued processing of the previous word (Swaab et al. 2012:398, 428). In my pilot study, since the P300 peaked at around 250 ms, the positive voltage deflection that peaked around 750 ms was more likely to be the P300 of the next word (presented from the 500 ms point onwards) rather than a late ERP component elicited by the current word. Thus, if the processing of the current word is complete or almost complete by the time the next word is presented, by virtue of having disappeared at the 500 ms point, any positive voltage deflection that occurs in the P600 latency range can more reliably be labelled as the P600 (or as another late ERP component) rather than being attributed to the processing of the following word. This change to the present design should help to clarify the inconclusive P600 results obtained in my pilot study.

A key difference with the experimental setup is that, while in Experiment 1 I placed electrodes above and below *both* eyes, in Experiment 2 (and subsequent experiments) I only placed electrodes above and below the left eye. This is because placing electrodes above and below both eyes is unnecessary for the blink detection procedure described in the following section (Kappenman: personal communication).

Additionally, if I was unable to place an electrode above the eye without having to adjust the positioning of the headcap, I would only place an electrode below the eye, to avoid altering the placement of nearby electrodes. During the data processing stage, I would then specify that the Fp1 channel should be used as the upper vertical electro-oculogram (henceforth VEOG), as this is located directly above the left eye (see figure 4.5).

6.3.4 Data processing

The main global data processing step in Experiment 2 that differs from that of Experiment 1 is the artifact detection procedure. In experiment 1 I only used the *moving-window peak-to-peak amplitude* method of artifact detection. By contrast, in Experiment 2 I carried out 3 separate artifact detection routines on each dataset in order to more accurately detect different types of artifacts. First, I applied the moving-window peak-to-peak amplitude technique on the VEOG channel in order to detect eye blinks. I created this channel specifically for this purpose by subtracting the lower VEOG channel (i.e. the data from the electrode placed below the eye) from the upper VEOG channel (i.e. the data from the electrode placed above the eye), using the *EEG channel operations* function in ERPLAB. Subtracting the lower VEOG from the upper VEOG increases the amplitude of voltage deflections that are caused by eye blinks, and thus makes them easier to detect (Luck 2014a:196). For this procedure, I used a low rejection threshold of 75 µV, a window step of 50 ms, and a moving-window length of 200 ms, as these are the optimal settings for blink detection (Luck et al. 2011:22).

I subsequently carried out the same moving-window peak-to-peak amplitude technique on *all* of the channels, in order to detect other artifacts such as skin potentials and muscle movements. For this, I used a larger moving-window length of 1000 ms and a higher rejection threshold of 200 µV. The benefit of running the moving-window peak-to-peak amplitude routine twice is that it ensures that different types of artifacts are accurately detected for rejection without inadvertently rejecting non-artifactual voltage deflections (Luck et al. 2011:22). If I were to run the routine just once, using a high rejection threshold on all channels, this would not detect blinks in the VEOG channel. Conversely, if I were to use a low rejection threshold on all channels, as I did in Experiment 1, this would detect blinks but not other artifacts such as the slow voltage drifts caused by skin potentials. Running the moving-window routine twice therefore optimizes artifact rejection without unnecessarily rejecting non-artifactual data (Luck et al. 2011:22).

The final method of artifact detection used in Experiment 2 was the *step-like artifacts* function. This function detects horizontal eye movements by identifying their typical step-like shape on the electroencephalogram, i.e. 200 ms at one voltage followed immediately by 200 ms at a different voltage (Luck et al. 2011:24). I used a moving window full width of 400 ms and a window step of 10 ms. I applied this function to the horizontal electro-oculogram (henceforth HEOG) channel, which I manually added by subtracting the left HEOG (i.e. the data from the electrode placed next to the left eye) from the right HEOG (i.e. the data from the electrode placed next to the right eye). By combining this step-like artifacts function with the two applications of the moving-window technique, I was able to increase the signal-to-noise ratio of the data and thereby obtain cleaner data than I did in the pilot study.

While the global data processing steps were largely the same as those carried out in Experiment 1, some of the individual participant datasets in Experiment 2 had atypical artifacts, which meant that those datasets had to be processed slightly differently. One such dataset is that of Participant 14. As can be seen in figure 6.2, channels EXG1 and EXG2 (the mastoid channels) for this participant display the distinctive periodic waveform that is caused by EKG activity (i.e. heart muscle activity).

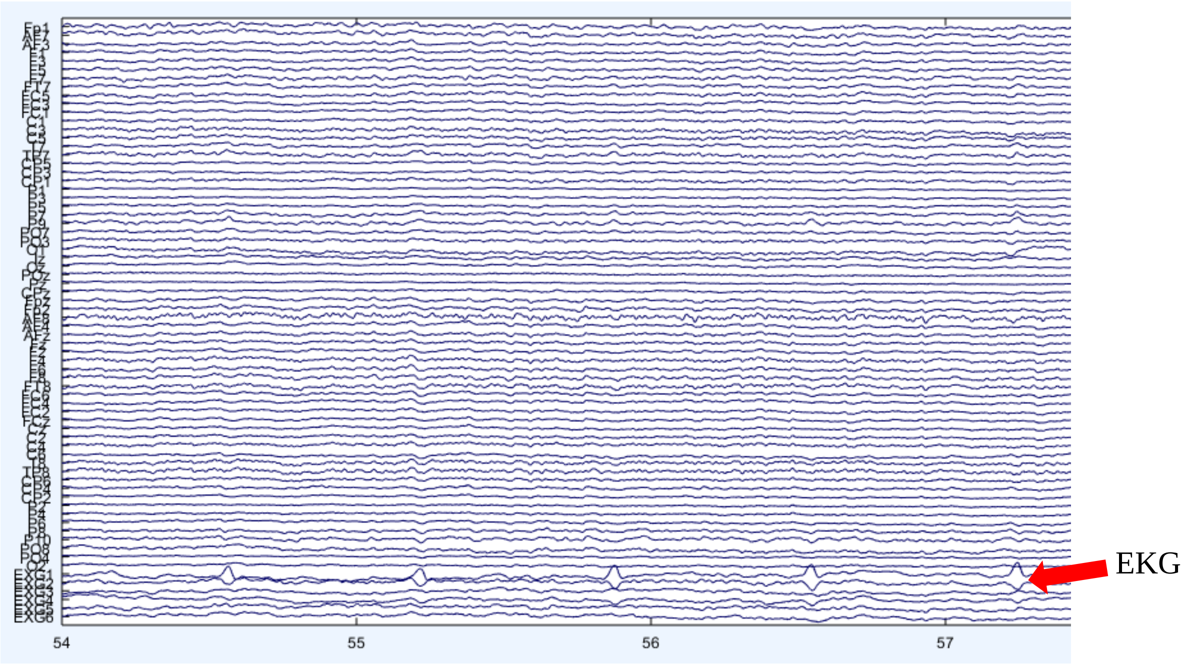


Figure 6.2: EKG activity in channel data from Participant 14

If EKG activity is present in a waveform, it will be present at the mastoid channels as these are located directly above the carotid arteries, where heart rate can be measured. The periodic nature of EKG means that it is unlikely to cause an artifactual difference between experimental conditions, so is not always problematic in EEG experiments. However, it *is* problematic in my experiments since I use the mastoids as the reference. As mentioned in Chapter 4 (section 4.7), any signal that is picked up by the reference is seen in inverted form in all other channels (Luck 2014a:206). Thus, I decided to use a different reference for this dataset in order to prevent the spread of noise in the data. I first re-referenced to the average of

P9 and P10 (see Chapter 4, section 4.6.3), as this is a reference that is increasingly used by leaders in ERP research (Luck 2016: personal communication). However, after noticing that this caused excessive noise to appear in the data, I realised that there was also EKG activity in P9 and to some extent in P10. It is standard practice in ERP research to try a different reference if any given reference causes more noise to appear in the data. Thus, I rereferenced the data again, this time using the average reference (see section 4.6.3). This was sufficient to prevent the unnecessary spread of noise through all the electrode channels.

Another instance where I had to use a reference other than the average of the mastoids was the case of Participant 15, where the mastoid channels EXG1 and EXG2 were particularly noisy. In this case, I *was* able to use the average of P9 and P10 as the reference as the noise was not evident in these channels.

6.3.5 Data analysis

In Experiment 1 I measured the amplitude and latency of four ERP components for the component-based analysis, as well as taking onset latency measurements for the component-independent analysis; in Experiment 2, I take onset latency measurements for the component-independent analysis but I focus on just two ERP components for the component-based analysis, namely the N400 and P600.

In Experiment 1 I used a latency range of 300-500 ms for the N400 and a latency range of 500-800 ms for the P600. These time windows were chosen as they are commonly used in the literature on the N400 and P600, respectively. However, as explained in Chapter 5 (section 5.5.6), the presence of the P300 in the pilot data could have minimised the amplitude of the N400. Similarly, the presence of the peak that occurred at approximately 750 ms in the pilot data made it impossible to state that a P600 existed independently of this later peak. In Experiment 2, I therefore measured the amplitude and latency of the N400 and P600 using tighter time windows. For the N400, I used a latency range of 350-500 ms, as this tighter time window would reduce the impact of the earlier P300; for the P600, I used a time window of 500-650 ms in order to exclude the later peak.

The results of the pilot study showed that the N400 is maximal over anterior-central scalp regions, with no distinct laterality (Chapter 5, section 5.4.1.1). In ERP research, it is customary to use the results of a pilot study to narrow down the electrode sites of interest so that these can become the focus of subsequent studies (Luck 2014a:122). Therefore, rather than including all of the electrode zones used in the pilot study, I focused the statistical analysis on the anterior and central electrode zones. This constitutes 44 channels across 6 statistical zones, namely left frontal, mid frontal, and right frontal, and left central, mid central, and right central. This focused analysis is presented in the following sub-section. However, for reasons outlined at the end of section 6.4.1.1.1, I then repeated the Experiment 2 N400 analysis using all 9 electrode zones, as in the original pilot study. This latter analysis is presented in section 6.4.1.1.2.

For the P600 analysis, I again used all 9 electrode zones. This is because the results of the pilot study were inconclusive regarding the P600; I decided that, due to the large latency range used, what appeared to be a potential P600 was actually the P300 of the next word. In contrast, since the onset latency results from the pilot study were unproblematic, I opted to use these results as the basis of the subsequent experiments. Thus, since the onset latency effect was maximal over central and anterior scalp regions, these are the electrode sites that I used in this and subsequent experiments. As with the initial N400 analysis described above, this constitutes 44 channels across 6 statistical zones, namely left frontal, mid frontal, and right frontal, and left central, mid central, and right central. I also excluded any instances where the 50% onset latency occurs before 150 ms post stimulus (see Chapter 5, section 5.4.2, for an explanation).

6.4 Results: Experiment 2

6.4.1 Component-based experimental design

6.4.1.1 N400 (350 - 500 ms)

6.4.1.1.1 Initial N400 analysis (anterior central electrode sites)

A comparison of the mean amplitude between the 350-500 ms latency range in the two conditions reveals that there is a slightly positive amplitude in the collocational condition (*M* = 0.179, *SD* = 1.98) and a slightly negative amplitude in the non-collocational condition (*M* = -0.404, *SD* = 2.268). The results of the repeated measures ANOVA reveal that this difference is statistically significant, *F*(1, 698) = 20.533, *p* < .001.

Figure 6.3 displays the grand average ERP waveforms for one electrode site at each of the six electrode zones used in the statistical analysis. These particular electrode sites were selected because I had previously used them as six of the nine representative electrode sites in Experiment 1 (and originally by Rhodes & Donaldson 2008:54).

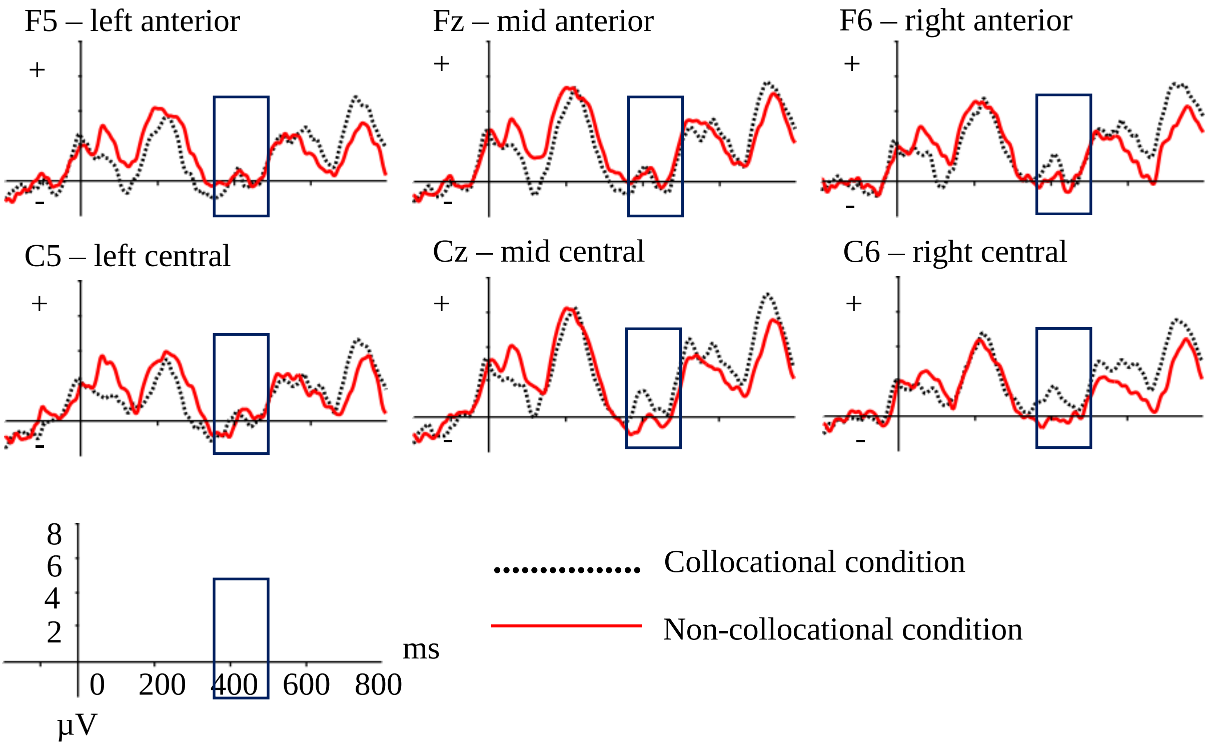


Figure 6.3: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus

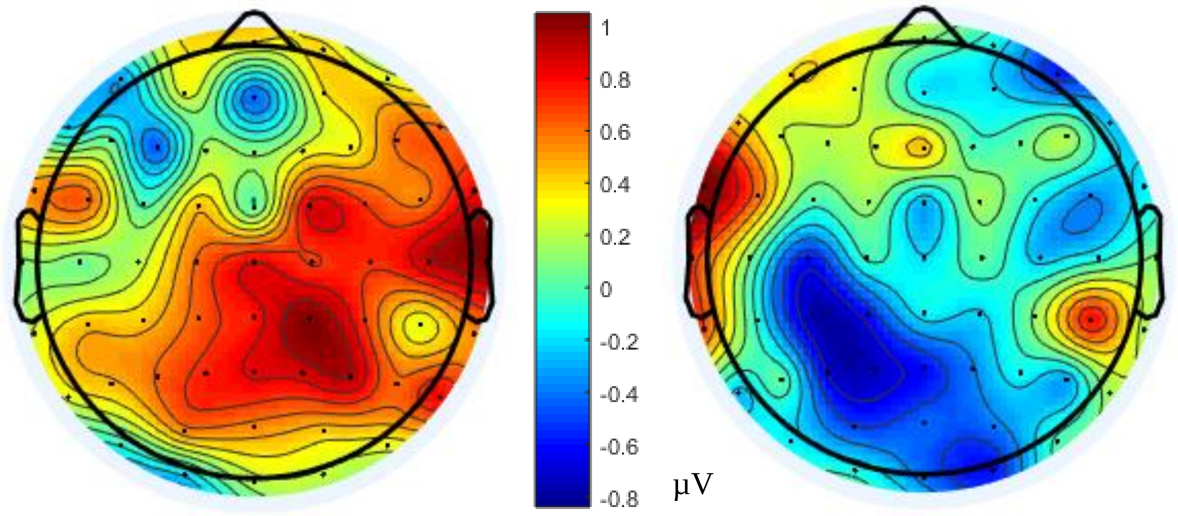
It is clear from figure 6.3 that the amplitude in the 350-500 ms latency range is lower in the non-collocational condition compared to the collocational condition at electrode sites F6, Cz, and C6. However, the difference between conditions is less clear at F5, C5, and Fz. This suggests that there is an effect of laterality, with the N400 effect being maximal over the right hemisphere. Indeed, this is supported by the results of subsequent repeated measures ANOVAs, which show that, while there is no interaction between condition and anterior-to-central electrode position, there is a significant interaction between condition and left-to-right electrode position, $F(2, 698) = 4.067$, $p = .018$.

Specifically, as shown in table 6.1, the mean amplitude is lower in the non-collocational condition at the three different electrode zones, yet the difference between the means is not significant at electrode sites located in the left hemisphere; the difference is significant at midline electrode sites ($F(1, 92) = 5.012$, $p = .028$), and highly significant at electrode sites over the right hemisphere ($F(1, 302) = 23.07$, $p < .001$). This is interesting because there was no evidence for a laterality effect in Experiment 1, and the N400 is traditionally thought to be a non-lateralized component (Swaab et al. 2012:399).

Table 6.1: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites at the 350-500 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Left hemisphere	-0.052	1.884	-0.28	2.134	.315
Right hemisphere	0.379	1.934	-0.501	2.415	< .001***
Midline	0.23	2.342	-0.536	2.201	.028*

The greater negativity in the non-collocational condition at right hemisphere and midline electrode sites is evident in the topographic scalp maps in figure 6.4.



Collocational condition

Non-collocational condition

Figure 6.4: Topographic scalp maps showing mean amplitude between 350 and 500 ms

However, these scalp maps also reveal that the greater negativity in the non-collocational condition is not fully restricted to the midline and the right hemisphere as the ANOVA results suggest. Moreover, the greater negativity is also not restricted to anterior and central scalp sites, as was implied by the results of Experiment 1. Rather, looking at figure 6.4, it is clear that the greater negativity in the non-collocational condition extends to posterior scalp sites as well as, to some extent, left hemisphere scalp sites. The posterior negativity in particular could not have been detected by the analysis procedure used thus far in Experiment 2, which focuses solely on anterior and central scalp sites. Therefore, I decided to repeat the analysis using all scalp sites.

For this new analysis, I used the same procedure that I used in Experiment 1. This involved conducting a repeated measures ANOVA using all nine electrode zones, with three factors (experimental condition, anterior-to-posterior electrode position, and left-to-right electrode position), and then conducting subsequent repeated measures ANOVAs to locate the electrode zones where the effect is maximal. The results of this new N400 analysis are provided in the following sub-section.

6.4.1.1.2 New N400 analysis (all electrode sites)

A comparison of the mean amplitude between the 350-500 ms latency range of the two conditions reveals that there is a slightly positive amplitude in the collocational condition (*M* = 0.129, *SD* = 1.857) and a slightly negative amplitude in the non-collocational condition (*M* = -0.416, *SD* = 2.142). The results of the repeated measures ANOVA reveal that this difference is not statistically significant. Nevertheless, as mentioned in Chapter 4 (section 4.6.8), it is expected that there will be no main effect in ERP studies that analyse data from all electrode sites. As Luck (2014a:336) explains, "[b]ecause the difference between conditions is likely to be large at a subset of the sites and small or even opposite at others, you probably won't see a significant main effect of condition".

Figure 6.5 displays the grand average ERP waveforms for one electrode site at each of the nine electrode zones used in the statistical analysis.

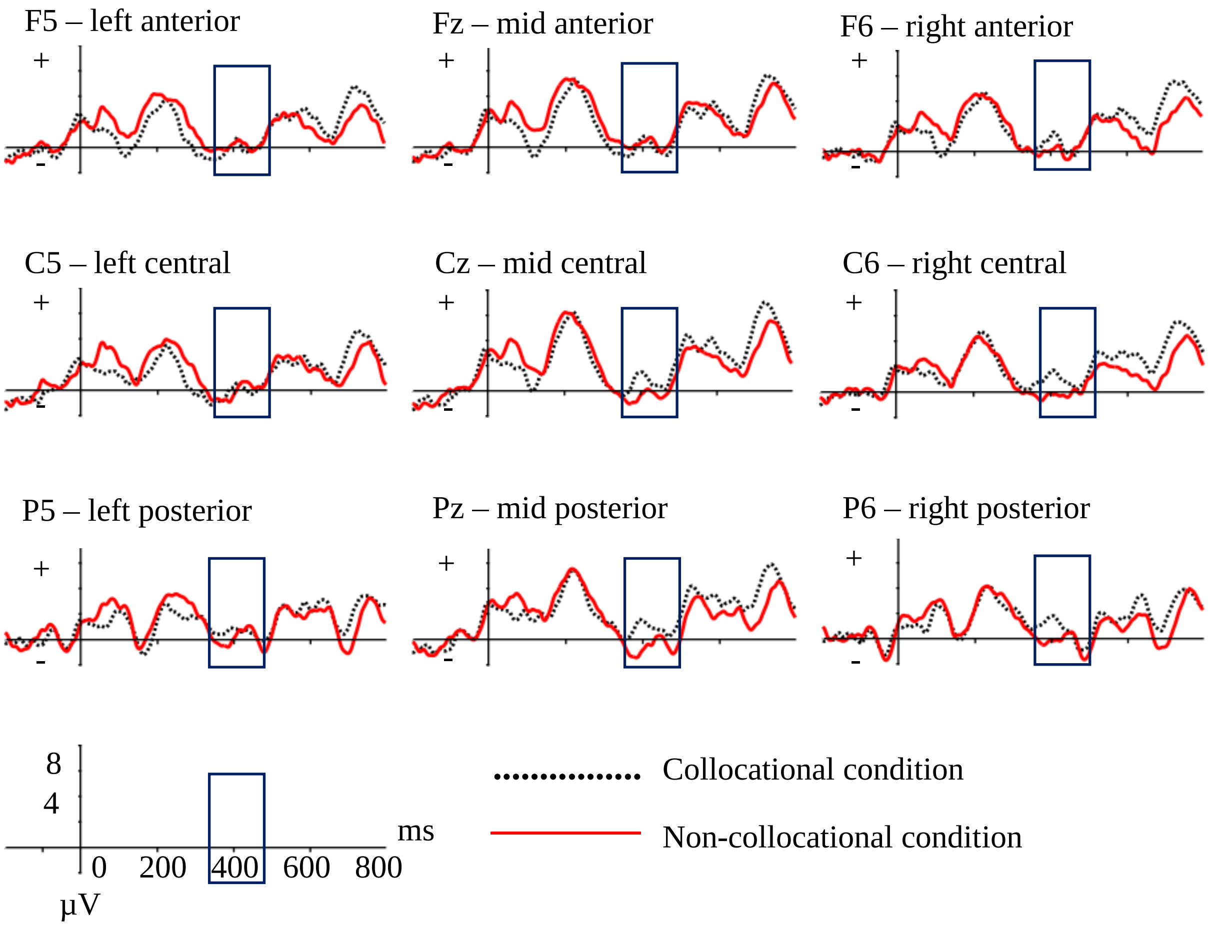


Figure 6.5: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus

It is clear from these waveforms that there is a greater negativity in the N400 time window in the non-collocational condition, particularly at central and posterior electrode sites. This result supports my decision to re-do the Experiment 2 N400 analysis, as the previous analysis conceals the posterior effect. That said, the laterality results appear to be in line with those of the previous analysis, as any potential N400 effect is less clear at left hemisphere electrode sites compared to midline or right hemisphere electrode sites.

Indeed, the results of the repeated measures ANOVA reveal that, although there is no interaction between condition and anterior-to-posterior electrode position, suggesting that the effect is significant at anterior, central, *and* posterior sites, there *is* a significant interaction between condition and right-to-left electrode position: $F(1, 1020) = 5.603$, $p = .018$. The results of subsequent repeated measures ANOVAs, carried out separately on left, right, and midline electrode sites, are summarised in table 6.2.

Table 6.2: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites in the 350-500 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Left hemisphere	-0.835	1.797	-0.328	2.047	.124
Right hemisphere	0.321	1.829	-0.48	2.261	< .001***
Midline	0.171	2.082	-0.427	2.1	.021*

The mean amplitude is lower in the non-collocational condition compared to the collocational condition at right hemisphere and midline electrode sites. This difference is significant at midline sites ($F(1, 147) = 5.449$, $p = .021$), and is highly significant at right hemisphere sites ($F(1, 429) = 29.781$, $p < .001$).

The results of this analysis show that there is in fact a greater negativity in the N400 latency range in the non-collocational condition compared to the collocational condition. The N400 effect in this experiment is maximal at right hemisphere and midline electrode sites at all levels of anteriority.

6.4.1.3 P600 (500 - 650 ms)

In the 500-650 ms latency range, there is a greater positivity in the collocational condition (*M* = 1.703; *SD* = 3.434) compared to the non-collocational condition (*M* = 1.583; *SD* = 3.213). This immediately suggests that there is no P600 in the data. Moreover, the difference between conditions is not statistically significant: $F(1, 1020) = 2.189$, $p = .139$. Waveforms demonstrating the greater positivity in the collocational condition can be seen in figure 6.6.

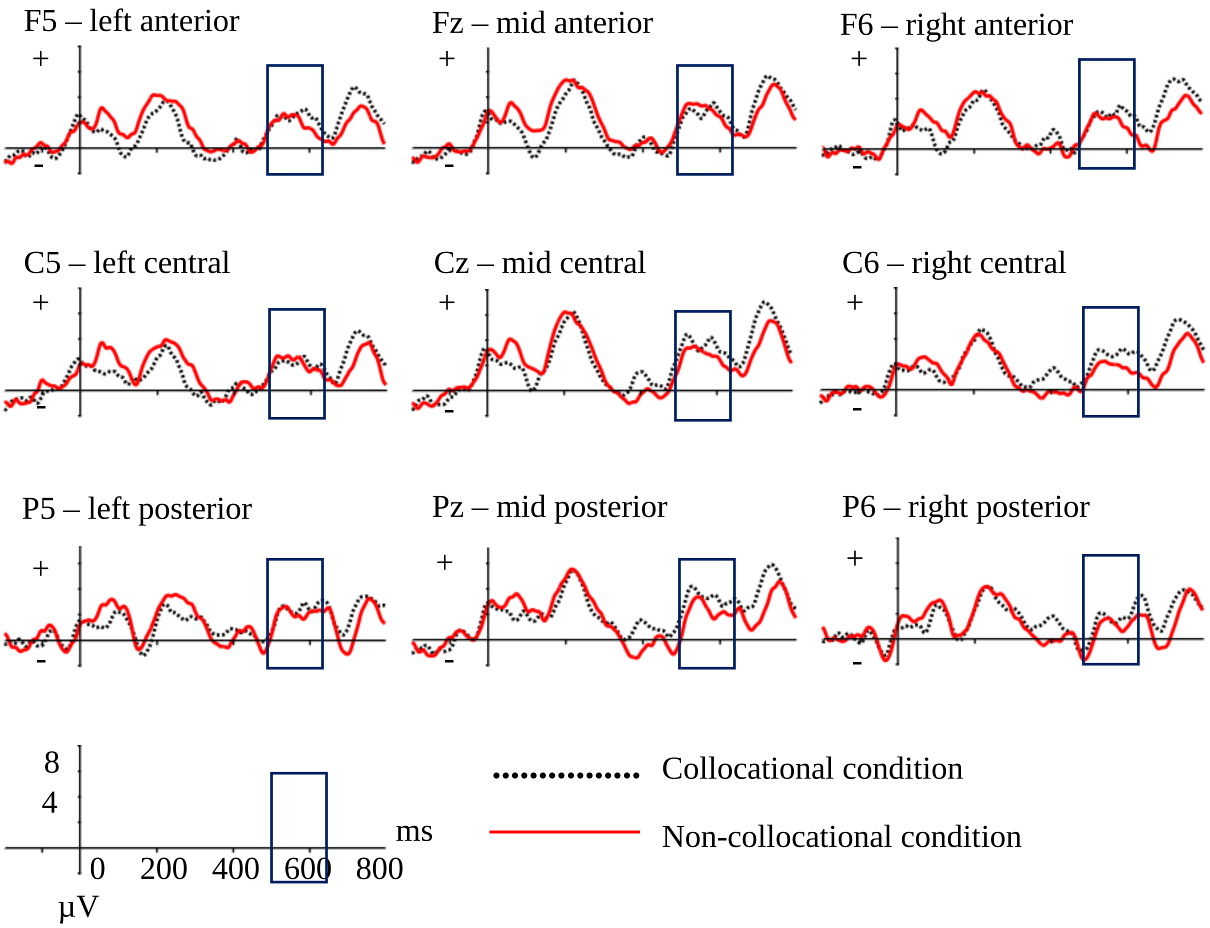


Figure 6.6: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus

While there is no interaction between condition and anterior-to-posterior electrode position, there *is* a significant interaction between condition and left-to-right electrode position: $F(1, 1020) = 4.309$, $p = .038$. Specifically, the results of subsequent repeated measures ANOVAs (see table 6.3) reveal that the effect is significant at right hemisphere electrode sites: $F(1, 429) = 7.445$, $p = .007$.

Table 6.3: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites in the 500-650 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Left hemisphere	1.466	3.376	1.675	3.274	.222
Right hemisphere	1.832	3.396	1.394	3.071	.007**
Midline	2.046	3.657	1.851	3.488	.427

However, it is important to observe here that the mean at right hemisphere (and indeed midline) electrode sites is actually higher in the collocational condition. This provides evidence against the hypothesis made in section 6.2, that a P600 would be elicited in the non-collocational condition.

6.4.2 Component-independent experimental design

As hypothesized in section 6.2, the onset latency is indeed greater in the collocational condition (*M* = 478.954; *SD* = 217.126) compared to the non-collocational condition (*M* = 409.744; *SD* = 230.18). While there is no significant interaction between condition and anterior-to-central electrode position, or condition and left-to-right electrode position, there *is* a significant main effect: *F*(1, 688) = 7.798, *p* = .005. This shows that, in Experiment 2, as in Experiment 1, there *is* evidence for a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams, regardless of the identification of specific ERP components.

6.5 Discussion: Experiment 2

6.5.1 Summary of aims and results

The aim of Experiment 2 was to replicate the results from the pilot study in order to provide stronger evidence in support of the idea that there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams.

To summarise the results of the component-based analysis for Experiment 2, it appears that there *is* an N400 effect in the data but that there is no P600 effect; or, at least, if there is a P600 effect, it is not in line with what I hypothesized. Evidence for the N400 effect comes from the fact that there is a greater negativity in the 350-500 ms latency range in the non-collocational condition compared to the collocational condition. This is in line with Hypothesis 1, i.e. reading the second word of non-collocational bigram will elicit an N400. The evidence against the existence of a P600 in the non-collocational condition comes from the fact that, rather than there being a greater positivity in the non-collocational condition, as I would expect if a P600 was elicited by reading the second word of a non-collocational bigram, there is actually a greater positivity in the collocational condition. Thus, I found the opposite to what I hypothesized; it could be said that there is actually a P600 in the collocational condition. This is discussed further in section 6.5.3 below.

The results for the component-independent analysis for Experiment 2 show that there *is* a significant difference between conditions, with the onset latency being greater in the collocational condition compared to the non-collocational condition. This means that the time point at which the amplitude is half of the value of the peak amplitude in the grand average waveforms occurs earlier in the non-collocational condition compared to the collocational condition. This confirms Hypothesis 3 and, importantly, demonstrates that there is a neurophysiological difference between conditions independent of the identification of specific ERP components; thus, this finding constitutes further evidence in support of the overarching hypothesis of this thesis.

The Experiment 2 results replicate the results of experiment 1 in that they again demonstrate the hypothesized N400 effect (despite there being a different scalp distribution in each case) as well as showing the same pattern of difference in onset latency. They also go beyond the results of Experiment 1 by providing a clearer picture of the status of the P600 in the context of collocation. However, the results of Experiment 2 still leave many issues unexplained, as I will now discuss.

6.5.3 Issues raised by the results

A key issue raised by the results of Experiment 2 is that the N400 found in the data does not share the same scalp distribution as the classical semantic N400. The classical N400 has a central-posterior scalp distribution with no distinct laterality (Swaab et al. 2012:399), yet the N400 found in the present study is maximal at right hemisphere and midline electrode sites at all levels of anteriority. This is not necessarily a problem. After all, as mentioned in Chapter 5 (section 5.4.1.1), Kutas and Dale (1997:222) point out that “N400s do differ in latency and scalp distribution, even within presumably similar experimental tasks”. Furthermore, Handy (2005:36) notes that there is “potential for extensive individual and group differences in ERP scalp topography”, especially for later components such as the N400, and so “it may not always be the case that a component of interest will be maximal over the expected scalp sites”.

That said, the fact that the scalp topography of the N400 found in the present study is different from that of the classical N400 does raise the question of whether or not what we are dealing with here is actually the N400 component that is widely discussed in the ERP literature. It *could* be the case that what we are dealing with is an entirely different component that just happens to share the same latency range as the classical N400. This is an issue that will be explored further in the following chapters in light of the results of the subsequent experiments.

An additional question is raised by the results of Experiment 2, namely: if there is an N400 effect in the absence of a P600 effect, does this mean that the N400 and P600 are in fact entirely separate ERP components reflecting entirely separate processing mechanisms (assuming for now that the N400 and P600 apparent in my experiments have any connection to semantic and semantic processing, respectively)? Recall from Chapter 3 (section 3.3.2.4) that there is growing evidence to suggest that the N400 and P600 are strongly related, in that the cognitive processes which they reflect are overlapping and interactive. Indeed, some studies have shown that the P600 is actually sensitive to semantic errors in the absence of any syntactic error (Kuperberg 2007:23), while other studies have shown that the N400 may be sensitive to morphosyntactic errors in the absence of any semantic error (e.g. Severens et al. 2008; Nieuwland et al. 2013). Based on those findings, Swaab et al. (2012) argue that "the interaction between semantic and syntactic processes in the brain may be more dynamic than was previously suggested". This constitutes important evidence in support of theories of language processing which do not posit the existence of separate lexical and grammatical systems. However, findings such as mine which demonstrate one of these components working independently of the other may suggest that they do not actually interact all that much, contrary to Swaab et al.'s (2012) interpretation. This issue will be discussed in depth in Chapter 9 (section 9.3.1).

A final issue that is raised by the results of Experiment 2 concerns the potential existence of a P600 in the collocational condition. As mentioned in the previous section, the P600 results were in opposition to what I hypothesized, with a higher mean amplitude existing in the collocational condition compared to the non-collocation condition at right hemisphere and midline electrode sites. Not only that, but the difference at right hemisphere electrode sites proved to be significant at the $p < 0.01$ level. This is interesting in two ways. First, it is interesting because it raises the question as to why there appears to be a P600 effect in the collocational condition; after all, since the collocational condition is the control condition representing conventionalized language use, there is no reason to believe that it *would* be elicited in response to reading collocational stimuli. Second, the apparent presence of a P600 in the collocational condition is interesting because *it still demonstrates a difference between* *conditions*, regardless of the direction of the effect or how well it fits my hypothesis. This provides further evidence in support of the overarching hypothesis that there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocation bigrams.

6.6 Chapter summary and conclusion

In this chapter, I have outlined the methodology and presented the results of Experiment 2, which compares the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in native speakers of English. This experiment contributes to the overall aim of the thesis by further investigating the overarching question of whether or not there in a neurophysiological difference in the processing of collocations compared to non-collocations. While Experiment 1 focused on four ERP components in addition to onset latency, Experiment 2 focuses more narrowly on two ERP components, namely the N400 and the P600, as well as onset latency.

The N400 and onset latency results mirror those of Experiment 1, showing that an N400 *does* appear to be elicited by reading the second word of a non-collocational bigram, and that the onset latency is smaller in the non-collocational condition compared to the collocational condition. Moreover, while the Experiment 1 results were inconclusive with regard to the P600, the Experiment 2 results provide no evidence of a P600 in the non-collocational condition.

In the following chapter, I present the methodology and results of Experiment 3, which compares the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in *non*-native speakers of English. Note that, when I refer to the results of Experiment 2 in subsequent chapters, I am referring to the new N400 and P600 analysed that use all electrode sites, rather than the initial analyses that use a subset of electrode sites based on the results of Experiment 1.

CHAPTER 7: EXPERIMENT 3 – A comparison of the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in non-native speakers of English

7.1 Overview of the chapter

In this chapter I focus on Experiment 3, which investigates the processing of collocational adjective-noun bigrams and non-collocational adjective-noun bigrams in non-native speakers of English. In conducting this experiment, I carried out exactly the same experimental procedure and used the same data analysis techniques as in Experiment 2 but, this time, with non-native speakers of English (specifically native speakers of Mandarin Chinese) rather than native speakers of English. Since the methodology is the same as in the previous chapter, and thus need not be described in detail here, this chapter is shorter than the other experiment chapters.

In section 7.2 I outline the aims and hypotheses of Experiment 3, with particular emphasis on how these have been shaped by the results of prior studies. I then very briefly outline the methodology in section 7.3 before providing the results of the component-based analysis in section 7.4.1, and the component-independent analysis in section 7.4.2. In section 7.5 I then discuss the results, before summarizing the chapter in section 7.6.

7.2 Aims and hypotheses: Experiment 3

The aims of Experiment 3 are to find out whether or not there is a neurophysiological difference in the way that collocational bigrams and non-collocational bigrams are processed by non-native speakers of English, and to see how the ERP results for the non-native speaker group compare to those of the native speaker group in Experiment 2

Based on the results of the self-paced reading experiment conducted by Hughes and Hardie (forthcoming), which reveal that non-native speakers of English are actually *more* sensitive to the transition probabilities between words than native speakers, I hypothesize that there *will* be a difference between conditions and, moreover, that this difference will be bigger than that demonstrated by the native English speakers. The logic behind this is that the learners have probably not encountered the non-collocational bigrams before or, if they have, they are unlikely to have encountered them frequently enough for them to become entrenched. By contrast, the native English speakers are likely to have encountered the non-collocational bigrams at some point, making them somewhat entrenched, or they will at least have greater flexibility in their use of (non-)collocational patterns than non-native speakers. Therefore, in a sense, this would make the non-native speakers more sensitive to the transition probabilities between words than the native speakers, leading to the non-native speakers exhibiting a proportionately larger ERP response. I thus predict that the difference between the native and non-native speakers will be quantitative rather than qualitative, with the non-native speakers demonstrating the same brain responses as the native speakers, but on a larger scale. In line with the results of Experiment 2, I thereby make the following hypotheses:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will *not* elicit a P600.

Hypothesis 3: The onset latency will be greater in the collocational condition compared to the non-collocational condition.

Hypothesis 4: The ERP responses will be larger than those demonstrated by the native English speakers in Experiment 2.

7.3 Methodology: Experiment 3

7.3.1 Participants

A total of 22 native speakers of Mandarin Chinese were recruited in order to obtain 16 usable datasets. This is a particularly high participant attrition rate, but with one methodological change I was able to reduce this attrition rate for Experiment 4 (see Chapter 8, section 8.3.2).

It was important for all of the participants in Experiment 3 to speak the same native language because, according to Granger (1998:158), the collocations that exist in the first language “inevitably play a role” in the acquisition of collocations in the second language. Similarly, as part of his theory of Lexical Priming, Hoey (2005:183) states that any primings that are acquired in the second language are “necessarily superimposed on the primings of the first language”. As a result, different findings could potentially be obtained from speakers of different languages. This could make it more difficult to interpret the results.

As well as it being important for all participants to speak the same native language, it was also fairly important for all participants to have a similar proficiency level. This is because different second/foreign language proficiency levels are associated with distinct patterns of EEG activity (e.g. Reiterer et al. 2005, 2009, 2011), even for specific components including the N400 and P600 (e.g. Hampton Wray & Weber-Fox 2013). Moreover, it was important for the participants to have a mid-to-high proficiency level, as this would minimise the chance of them having any problems with reading and understanding the experimental stimuli.

Ideally, I would have systematically tested the English proficiency levels of the participants by asking them to complete a language proficiency test such as DIALANG (available at http://dialangweb.lancaster.ac.uk/). In Hughes and Hardie (forthcoming), we used this online proficiency test, along with a language exposure questionnaire and self-reports of IELTS (International English Language Test System) scores (or equivalent) to assess the proficiency levels of participants. However, the DIALANG test was time-consuming, and not all participants were able to recall their IELTS scores. Moreover, while it was important for the proficiency information to be highly accurate in Hughes and Hardie (forthcoming), as language proficiency is central to one of the research questions in this study, language proficiency was not central to the aims of the present study; instead, it was an additional variable that could potentially influence the results. I did not want to further increase the already lengthy duration (approximately 1 hour overall) of the experimental sessions purely to account for a variable that is not central to the aims of the study. Therefore, following Reiterer et al. (2009), the only way that I attempted to control for variability in proficiency level was by selecting participants from a single cohort, namely the MA students in the Department of Linguistics and English Language, Lancaster University.

At present, the level of English required for postgraduate study in this department is an IELTS score of at least 6.5 (or equivalent). A score between 6 and 7 indicates that an individual is a “competent” user of the English language (IELTS Partners 2018). Therefore, although it is likely that there is still large variability in proficiency level, with some participants having an IELTS score of up to 8 or even 9, I can at least be confident in stating that the participants are unlikely to have had any problems with reading and understanding the experimental stimuli. The study of participants who fit into more narrowly defined proficiency bands is reserved for future ERP studies of collocational processing.

As with all experiments conducted for this thesis, all participants in Experiment 3 had normal or corrected-to-normal vision, and no (history of) neurological disorders. Participants were recruited in line with the ethics procedures of the Department of Linguistics and English Language, Lancaster University (see appendices 4 and 6 for information sheet and consent form).

7.3.2 Stimuli, procedure, and data analysis

The stimuli and procedure used are identical to those used in Experiment 2 (see Chapter 6, sections 6.3.2 and 6.3.2; for the four counterbalanced lists of experimental stimuli, see appendix 2). The data processing and analysis is also the same as in Experiment 2 (see Chapter 6, section 6.4.3 and 6.4.5). Specifically, I measured the amplitude of the N400 in the 350-500 ms latency range using all electrode sites. I then measured the amplitude of the P600 in the 500-650 ms latency range, again using all electrode sites. Finally, in line with Experiment 2, I measured the onset latency using the fractional peak latency function, focusing on anterior and central electrode sites.

7.4 Results: Experiment 3

7.4.1 Component-based experimental design

7.4.1.1 N400 (350 - 500 ms)

A comparison of the mean amplitude between the 350-500 ms latency range of the two conditions reveals that there is a positive amplitude in the collocational condition (*M* = 1.924, *SD* = 2.2) and a slightly negative amplitude in the non-collocational condition (*M* = -0.486, *SD* = 2.263). This is in line with the results of Experiments 1 and 2. Moreover, while the difference between conditions is not significant in Experiment 2, it is highly significant in Experiment 3: *F*(1, 892) = 51.602, *p* < .001.

Figure 7.1 displays the grand average ERP waveforms for one electrode site at each of the nine electrode zones used in the statistical analysis.

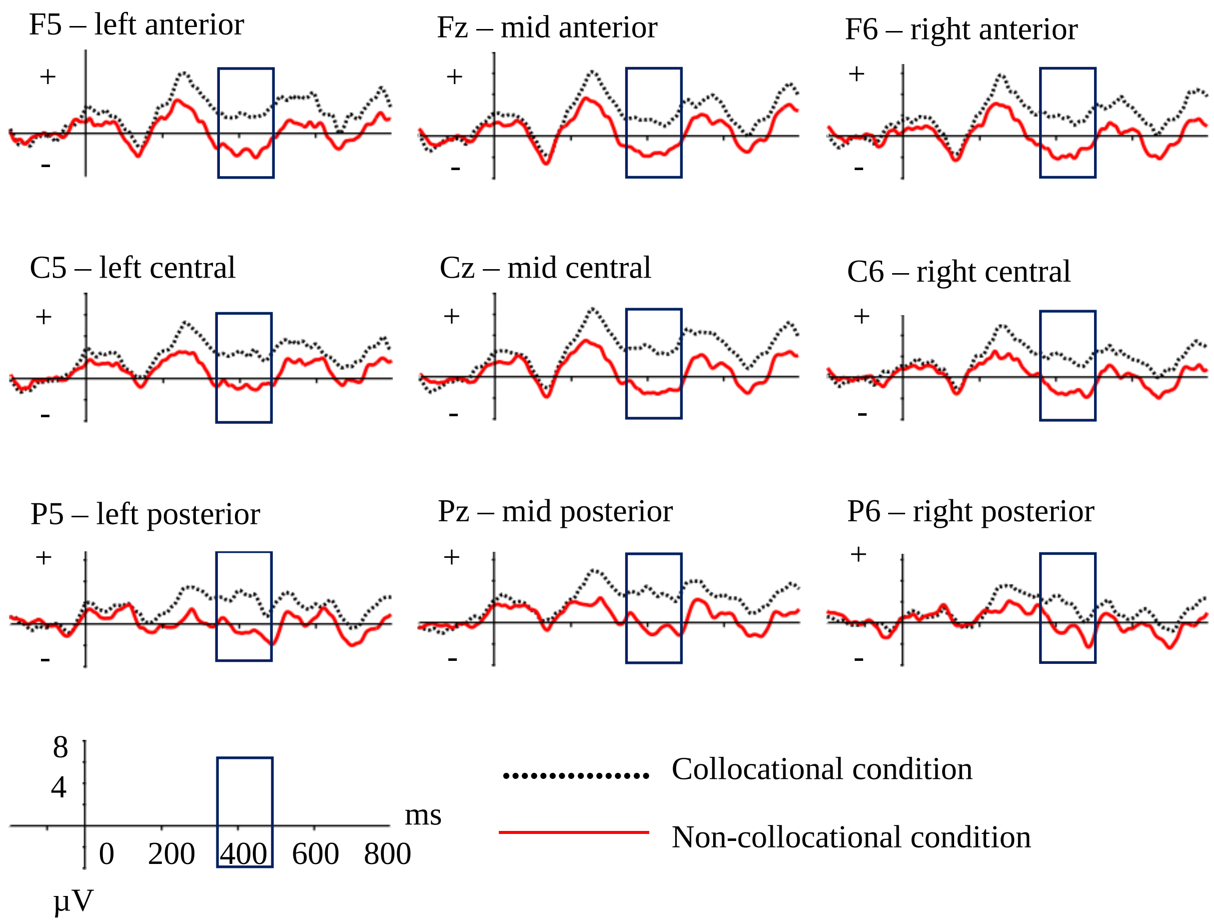


Figure 7.1: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus

The N400 effect shown in these waveforms is exceptionally large in comparison to the N400s present in Experiments 1 and 2 of this thesis. This suggests that, as hypothesized, the non-native speakers of English are more sensitive to the transition probabilities between words than the native speakers.

It is important to note, however, that the actual mean amplitude in the N400 latency range for the non-collocational condition is very similar in Experiments 2 (*M* = -0.416) and 3 (*M* = -0.486). Thus, the key difference between the experimental results is not the difference in the actual negative value in the N400 time window but, rather, the size of the difference between the conditions. The range between conditions in the N400 time window in Experiment

2 is 0.545, compared to 2.41 in Experiment 3. It is this difference which supports the hypothesis that the non-native speakers of English are likely to be more sensitive to the transition probabilities between words than the native speakers.

The clear difference between conditions in Experiment 3 can be seen in figure 7.2.

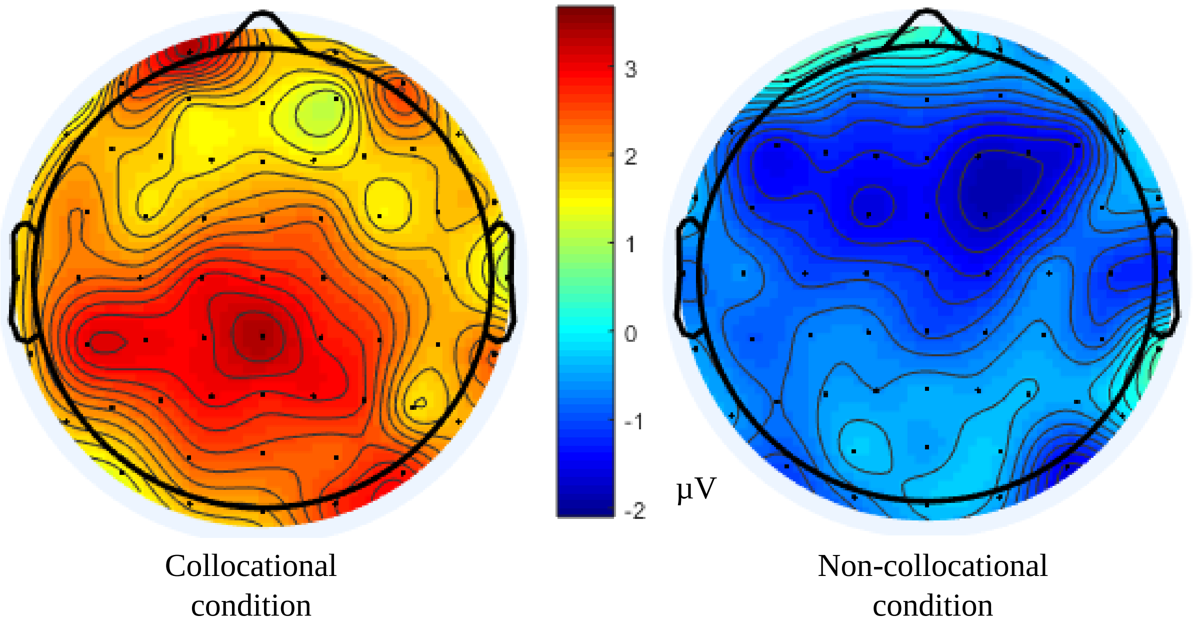


Figure 7.2: Topographic scalp maps showing mean amplitude between 350 and 500 ms

These topographic scalp maps demonstrate that the effect is evident at all scalp regions. Indeed, the results of subsequent repeated measures ANOVAs (see table 7.1) reveal that the N400 effect is significant at all levels of anteriority, with highly significant results at central and posterior electrode sites.

Table 7.1: Summary of post hoc ANOVA results carried out at left, right, and midline electrode sites at the 350-500 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Anterior	1.814	2.54	-0.162	2.54	< .01**
Central	2.068	1.947	-0.524	2.113	< .001***
Posterior	1.868	2.215	-0.425	1.912	< .001***

In summary, the results of this analysis show that, as in Experiments 1 and 2, there is a greater negativity in the N400 latency range in the non-collocational condition compared to the collocational condition. This greater negativity in the non-collocational condition is evident at all scalp regions. Moreover, the difference between conditions is much larger in Experiment 3 compared to Experiment 2, suggesting that non-native speakers of English have a greater sensitivity to the transition probabilities between words than native speakers.

7.4.1.2 P600 (500 - 650 ms)

In line with the results of Experiment 2, the mean amplitude in the 500-650 ms latency range is higher in the collocational condition (*M* = 2.291, *SD* = 2.271) compared to the non-collocational condition (*M* = 1.046, *SD* = 3.159). This again shows that there is no P600, at least not in the non-collocational condition as was initially hypothesized (section 6.2). Interestingly, though, unlike in Experiment 2, this *does* represent a significant difference: *F*(1, 892) = 8.418, *p* = .004.

The greater positivity in the collocational condition can be seen at all electrode sites displayed in figure 7.3.

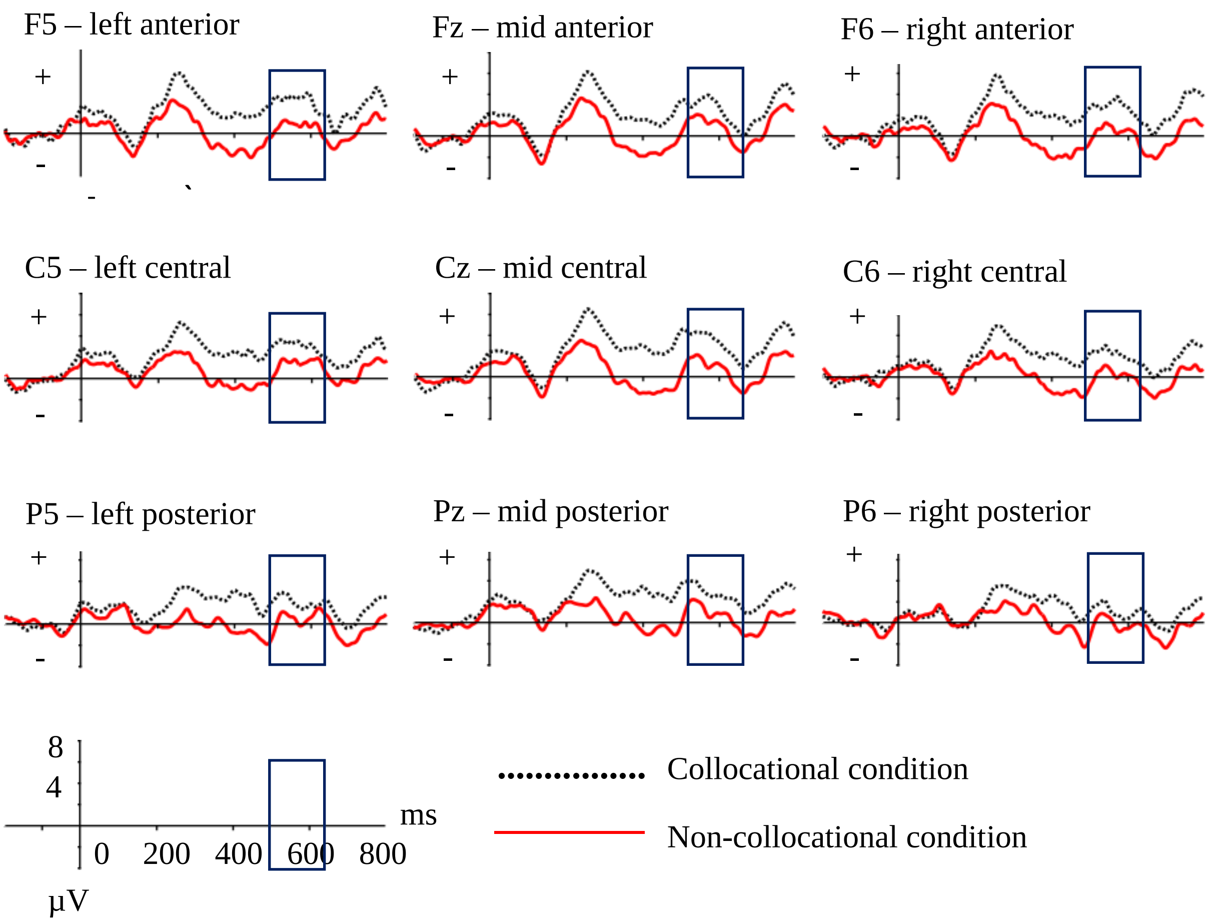
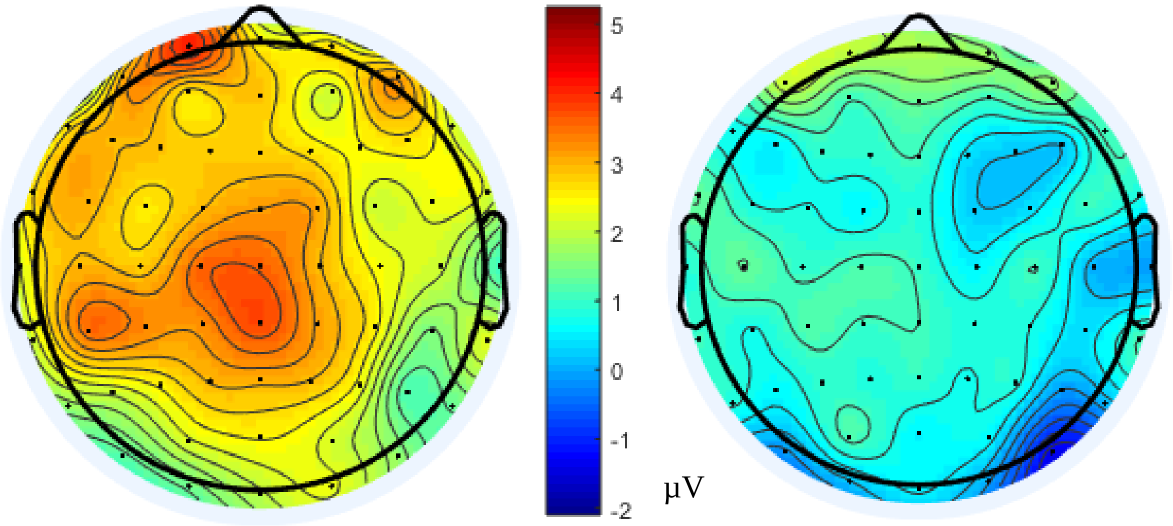


Figure 7.3: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus

The results of the repeated measures ANOVA reveal no significant interactions between condition and central-posterior electrode position, or condition and left-to-right electrodeposition. Indeed, looking at the topographic scalp maps in figure 7.4, the greater positivity in the collocational condition in the 500-650 ms latency range is apparent at all scalp regions



Collocational condition

Non-collocational condition

Figure 7.4: Topographic scalp maps showing mean amplitude between 500 and 650 ms

The issues raised by these P600 results will be discussed in section 7.5.2.

7.4.2 Component-independent experimental design

In contrast to the onset latency results of Experiments 1 and 2, the onset latency in Experiment 3 is slightly greater in the *non-collocational* condition (*M* = 395.386, *SD* = 204.667) compared to the collocational condition (*M* = 393.113, *SD* = 215.107). Moreover, the difference between conditions is not statistically significant: *F*(1, 600) = 0.376, *p* = .54. This actually provides evidence *against* the overarching hypothesis of this thesis, suggesting that the brain does *not* process collocations differently from how it processes non-collocations. This result will be discussed further in section 7.5.2.

7.5 Discussion: Experiment 3

7.5.1 Summary of aims and results

The aim of Experiment 3 was find out whether or not there is a neurophysiological difference in the way that collocational bigrams and non-collocational bigrams are processed by non-native speakers of English. The component-based experimental results confirm that this neurophysiological difference does exist. However, this is not supported by the results of the component-independent design.

The results of the N400 analysis reveal that there *is* an N400 effect in the non-native speaker data, and that this N400 is considerably larger than that found in the native speaker data in Experiment 2. This confirms hypothesis 1, which states that reading the second word of a non-collocational bigram will elicit an N400 for the non-native speakers. It also provides evidence in support of hypothesis 4, which states that the ERP responses will be larger than those demonstrated by the native speakers of English in Experiment 2.

Since this difference between participant groups is a difference in the size of the same component, rather than there being a difference in the nature of the components that are elicited, it seems that the difference between native and non-native speakers is quantitative rather than qualitative (though this claim will be problematized in section 7.5.2). Moreover, since the N400 is larger for the non-native speakers compared to the native speakers, it seems that the non-native speakers are actually more sensitive to the transition probabilities between words than the native speakers. This is in line with the results of Hughes and Hardie (forthcoming).

As mentioned in section 7.2, this finding could be attributed to the non-native speakers never having encountered the non-collocational bigrams before, making it even more unexpected and surprising when they do encounter the non-collocational bigrams in the experimental setting. By contrast, the native speakers are more likely to have encountered the non-collocational bigrams at least occasionally in their linguistic experience, and they are likely to have greater flexibility in their use of rare, novel, or unexpected combinations of items. Thus, encountering a non-collocational bigram will be less surprising and unexpected for them than it is for the non-native speakers.

The results of the P600 analysis reveal that the mean amplitude in the 500-650 ms latency range is higher in the collocational condition compared to the non-collocational condition. This is in line with the results of Experiment 2, providing further evidence to show that reading the second word of a non-collocational bigram does not elicit a P600. This therefore provides confirmation of hypothesis 2.

The results of the P600 analysis also support hypothesis 4, which states that the ERP responses demonstrated by the non-native speakers will be larger than those demonstrated by the native English speakers in Experiment 2. Evidence in support of this hypothesis comes from the fact that, although there appears to be no P600, the difference in amplitude between conditions is still larger for the non-native speaker group (range: 1.245) compared to the native speaker group (range: 0.120). Thus, even if we cannot specify a particular component that is being modified in the 500-650 ms latency range, we can say that the difference between conditions is associated with a larger amplitude for the non-native speakers compared to the native speakers. This can be interpreted as reflecting a greater cognitive load for the non-native speakers.

For the component-independent design, the results refute hypothesis 3, which states that the onset latency will be greater in the collocational condition compared to the non-collocational condition, as the onset latency is actually greater for the *non-collocational* bigrams. However, the difference is only marginal and it is not statistically significant. Moreover, the onset latency results weigh against hypothesis 4, which states that the ERP responses demonstrated by the non-native speakers will be larger than those demonstrated by the native English speakers. The range between the onset latency of the collocational condition and the onset latency of the non-collocational condition is actually much smaller for the non-native speakers (2.272 ms) compared to the native speakers (69.21 ms). This suggests that it is actually the native speakers who are the most sensitive to the transition probabilities between words. This will be discussed in depth in Chapter 9 (section 9.4.3).

7.5.2 Issues raised by the results

The N400 results of Experiment 3 are largely unproblematic, as they fall in line with those of Experiments 1 and 2; all of these experiments demonstrate an enlarged N400 in the non-collocational condition. However, the scalp distribution of this purported N400 is not the same in each experiment. The N400 in Experiment 1 is maximal over anterior scalp sites (though recall from Chapter 5, section 5.5.6, that the Experiment 1 results are not entirely reliable, as the N400 is diluted by the P300 effect). In Experiment 2, the N400 is maximal over right hemisphere and midline electrode sites; and, in Experiment 3, the N400 effect is significant at *all* electrode zones.

As well as being different across experiments, these scalp distribution results are not the same as those found in classical N400 studies, where the N400 has a central-posterior scalp distribution with no distinct laterality (Swaab et al. 2012:399). As mentioned in sections 5.4.1.1 and 6.5.3, this is not necessarily a problem, as "N400s do differ in latency and scalp distribution, even within presumably similar experimental tasks" (Kutas & Dale 1997:222). However, it does warrant further investigation. It could be the case that it is not actually the same component that is present across each of these experiments. Rather, they could be different components that just happen to occur within the N400 latency range. This would mean that there is a *qualitative* difference in language processing across participant groups, rather than a quantitative difference as was previously suggested (see sections 7.2 and 7.5.1). Future studies with non-native speakers would be needed in order to see whether or not this reported scalp distribution is consistently present.

The key issue raised by the results of P600 analysis is that the mean amplitude in the 500-650 ms latency range is significantly larger in the collocational condition compared to the non-collocational condition. It is not clear why this is the case. It is very unexpected for there to be a potential component resembling a P600 in the collocational condition, as this suggests that it is the *collocational* condition which is associated with greater processing demands, even though this is actually the control condition that should be associated with lesser processing demands. In this context, it should be emphasized that, even though the mean amplitude is larger in the 500-650 ms range for the collocational condition compared to the non-collocational condition, I do not claim that reading the second word of a collocational bigram elicits a P600. ERP findings always need to be replicated in order to confirm that a particular ERP effect exists. As Luck (2012:312) notes, "the best approach is often to analyse multiple components but take the results seriously only for a few *a priori* comparisons (and rely on replication for assessing the reliability of the other significant effects)". Thus, since a significant effect in the 500-650 ms latency range has only been found in Experiment 3, further experiments would need to be conducted to confirm the presence of a P600.

Two key issues are raised by the results of the component-independent design. First, the onset latency is greater for the non-collocational condition compared to the collocational condition (albeit only marginally and insignificantly). This is in opposition to the results of Experiments 1 and 2, and it is unclear why the results would be so different for the non-native speakers. Second, the range between the onset latency of the collocational condition and the onset latency of the non-collocational condition is much smaller for the non-native speakers compared to the native speakers, suggesting that it is actually the native speakers who are more sensitive to the transition probabilities between words.

One possible explanation for these unexpected results is that reading non-collocational bigrams engages different brain regions for the non-native speakers (insofar as the location of electrode sites has any relation to the underlying brain activity). The decision to carry out the onset latency analysis on anterior and central electrode zones is based on the fact that the effect was maximal over these electrode zones in Experiment 1. However, the participants in Experiment 1 were all native speakers of English. It could be the case that, for non-native speakers, I would need to look at other (or all) electrode zones in order to get a full picture of the onset latency effect. Furthermore, as mentioned earlier, any significant effects need to be replicated in order to be considered conclusive findings in ERP research.

7.6 Chapter summary and conclusion

In this chapter, I have reported and discussed the results of Experiment 3, which aimed to find out whether or not there is a neurophysiological difference in the way that collocational bigrams and non-collocational bigrams are processed by non-native speakers of English. The results of the component-based analysis confirm that this is indeed the case. For the non-native speakers, reading the second word of a non-collocational bigram elicits an N400 that is larger than that elicited by the native speakers in Experiment 2. Moreover, the results of Experiment 3 show that, although reading the second word of a non-collocational bigram does not elicit a P600, there is a significant difference between conditions in the 500-650 ms latency range. This provides further evidence in support of the idea that there is a neurophysiological difference in the way that collocational bigrams and non-collocational bigrams are processed. However, my hypotheses were not supported by the results of the component-independent analysis, which reveal no significant difference between conditions. In the following chapter, I focus on Experiment 4, which replicates Experiment 2 and also investigates the psychological validity of different measures of collocation strength.

CHAPTER 8: EXPERIMENT 4 – An investigation into the psychological validity of different measures of collocation strength

8.1 Overview of the chapter

In this chapter I present the final experiment of this thesis. There are two parts to this experiment. In Part 1, I provide a replication of Experiment 2 using a new set of stimuli. In Part 2, I investigate the correlation between the strength of the bigrams and the amplitude of the ERP response. I state the aims and hypotheses of the present experiment in section 8.2, before outlining the methodology in section 8.3. I then present the results in section 8.4 for Part 1 (section 8.4.1) and Part 2 (section 8.4.1) of the experiment. Within section 8.4.1, there are two sub-sections for the component-based design (section 8.4.1.1) and the component-independent design (section 8.4.1.2). In section 8.5 I discuss the results of both parts of Experiment 4, first by providing a summary, and then by outlining the issues raised by the results. Finally, I conclude the chapter in section 8.6.

8.2 Aims and hypotheses: Experiment 4

There are three aims of Experiment 4, namely:

Aim 1: To replicate Experiment 2 in order to strengthen the confidence of its conclusions.

Aim 2: To investigate the strength of the correlation between the transition probability of a bigram and the amplitude of the ERP response.

Aim 3: To find out which measure of collocation strength most closely correlates with the amplitude of the ERP response, and thus may be seen as having the most psychological validity.

Having set out the aims of Experiment 4, I will now state the hypotheses that are associated with these aims:

Hypothesis 1: Reading the second word of a non-collocational bigram will elicit an N400.

Hypothesis 2: Reading the second word of a non-collocational bigram will *not* elicit a P600.

Hypothesis 3: The onset latency will be greater for the collocational condition compared to the non-collocational condition.

Hypothesis 4: There *is* a correlation between the transition probability of a bigram and the amplitude of the ERP response.

Hypotheses 1, 2, and 3 are based on the results of Experiment 2. When formulating these hypotheses, I did not take into account the results of Experiment 3, as the participant group in the present experiment consists of native speakers of English rather than non-native speakers. Hypothesis 4 is based on the assumption that, since the results of the previous experiments have confirmed that there *is* a neurophysiological difference in the processing of collocations and non-collocations, a correlation will be apparent when collocation strength is treated as a continuous variable rather than as a dichotomy. Finally, due to conflicting results in the literature (see Chapter 2, section 2.11), I do not make any hypotheses regarding which measures of collocation strength seem to have the most or least psychological validity.

8.3 Methodology: Experiment 4

8.3.1 Participants

A total of 17 participants, all native speakers of English, were needed in order to obtain 16 usable datasets. This represents a considerable improvement on Experiment 3, where 22 participants were required in order to arrive at 16 datasets that could be used in the analysis (see Chapter 7, section 7.3).

As with the previous experiments, all participants were students at Lancaster University and had normal or corrected-to-normal vision, and no (history of) neurological disorders. Participants were recruited in line with the ethics procedures of the Department of Linguistics and English Language (see appendices 4 and 6 for information sheet and consent form). None of the participants used in Experiment 4 had taken part in Experiments 1, 2, or 3.

8.3.2 Stimuli

The same stimuli were used for Part 1 and Part 2 of the experiment. The set of stimuli was necessarily different from that used for Experiments 1, 2, and 3 because, in this experiment, collocation strength is treated as a continuous variable rather than there being a dichotomy between collocations and non-collocations. In order to conduct the correlational analysis, I needed to have collocational bigrams which spanned a range of different levels of transition probability. Specifically, each bigram falls in a different transition probability band ranging from $0<b<0.1$ (where *b* stands for bigram transition probability) for the weakest collocation, to $0.8\leq b<0.9$ for the strongest collocation.

In Experiment 4 I used a total of nine bigram pairs, five of which were also part of the previous stimuli set. The remaining bigrams from the previous stimuli set were excluded from Experiment 4 either because (a)there was more than one bigram falling within the same transition probability band, or (b) because the adjective and noun in the collocational bigram were more semantically related than the adjective and noun in the matched non-collocational bigram. I chose to have just one bigram pair in each transition probability band because, for the highest bands, it was impossible to find more than one suitable bigram pair. Moreover, it was important to ensure that the words in the collocational bigrams were less semantically related than the words in the matched non-collocational bigrams, as this allows experimental effects to be attributed to collocationality rather than semantic relatedness. I did not take semantic relatedness into account in the previous experiments in this thesis. Therefore, if the results of Experiment 2 are not replicated in Experiment 4, this suggests that the results of previous experiments might actually reflect semantic processing as opposed to collocational processing.

For Experiment 4, I measured the semantic relatedness of the adjective and noun in each bigram pair using WordNet::Similarity 2.07 (Pederson et al. 2005-2008): an open source Perl module based on the WordNet 3.0 lexical database (Princeton University 2010). Within WordNet::Similarity 2.07, the semantic distance between a pair of words can be quantified using 11 different metrics. However, most of these metrics actually measure semantic *similarity*, which captures just one semantic relationship such as hyponymy (car~vehicle); by contrast, semantic *relatedness* is a much broader notion capturing a wider range of relationships such as near-synonymy (*true~correct*), antonymy (*love~hate*), and meronymy (*wheel~car*) (Resnik 1995:448; Budanitsky & Hirst 2001:2; Michelizzi 2005:15, 21). Moreover, while semantic similarity can only be calculated for two words from the same word class, semantic relatedness can be calculated for two words from different word classes (Lintean & Rus 2012:245; Jurafsky & Martin 2017:16). It was therefore necessary for me to use a semantic relatedness metric in order to quantify the semantic relationships between adjectives and nouns, and to capture a broad range of semantic relations. Specifically, I used the Hirst and St. Onge (henceforth HSO) metric, which quantifies the semantic relatedness of a pair of words from 0 (least related) to 16 (most related). I excluded the bigram pairs from the Experiment 1/2/3 stimuli set that have a higher HSO score in the collocational condition compared to the non-collocational condition. This was the case for the bigram pairs *head teacher/character* and *chief executives/definitions*. The use of the HSO metric is problematized in Chapter 9 (section 9.5).

The four new collocational bigrams were extracted from the written BNC1994, using the same method that I used to select the bigrams from the previous stimuli set (see Chapter 4, section 4.2.1). For each new collocational bigram, I also manufactured a non-collocational bigram which was closely matched for frequency and length. As mentioned in Chapter 4 (section 4.2.1), the non-collocational bigrams do not occur in the BNC1994 at all. This is based on the assumption that absence of a particular bigram from the BNC1994 can serve as a proxy for the identification of non-collocational bigrams. Since the non-collocational bigrams do not occur in the BNC1994, for the purposes of this thesis, I state that the non-collocational bigrams have a transition probability of zero. However, as the non-collocational bigrams are, by definition, semantically plausible word pairs, they cannot actually have a transition probability of zero; rather, they have a transition probability that is lower than can be measured in the written BNC1994.

The bigrams used in Experiment 4 are shown in table 8.1, with the collocational bigrams listed in descending order of transition probability. The transition probability of each collocational bigram is given in brackets.

Table 8.1: Bigrams for Experiment 4

TP band	Collocational bigrams (transition probability)	Non-collocational bigrams
0.8≤b<0.9	nineteenth century (0.855)	nineteenth position
0.7≤b<0.8	prime minister (0.796)	prime period
0.6≤b<0.7	foreseeable future (0.678)	foreseeable weeks
0.5≤b<0.6	integral part (0.509)	integral thought
0.4≤b<0.5	twenty-four hours (0.429)	twenty-four patients
0.3≤b<0.4	disposable income (0.353)	disposable property
0.2≤b<0.3	minimum wage (0.246)	minimum prize
0.1≤b<0.2	vast majority (0.182)	vast opportunity
0<b<0.1	crucial point (0.017)	crucial night

Full details of the statistical properties of the bigrams can be seen in table 8.2.

Table 8.2: Statistical properties of Experiment 4 bigrams extracted from written BNC1994

TP band	Adjective (X)	Noun (Y)	Frequency of X	Frequency of Y	Frequency of X-then-Y	Transition Probability	Number of letters in Y	Number of syllables in Y	In no. of texts
0.8≤b<0.9	Nineteenth	century	3057	19025	2614	0.855	7	3	569
0		position	3057	21010	0	0	8	3	0
0.7≤b<0.8	Prime	minister	11634	23935	9264	0.79628	8	3	9264
0		period	11634	24108	0	0	6	3	0
0.6≤b<0.7	foreseeable	future	410	12933	278	0.678	6	2	234
0		weeks	410	13018	0	0	5	1	0
0.5≤b<0.6	Integral	Part	1182	53414	602	0.5093	4	1	602
0		thought	1182	53567	0	0	7	1	0
0.4≤b<0.5	twenty-four	hours	842	16351	361	0.4287	5	1	361
0		patients	842	16906	0	0	8	2	0
0.3≤b<0.4	disposable	income	399	11146	141	0.35338	6	2	141
0		property	399	12023	0	0	8	3	0
0.2≤b<0.3	minimum	wage	1413	2811	347	0.2456	4	1	109
0		prize	1413	2853	0	0	5	1	0
0.1≤b<0.2	vast	majority	4398	9296	803	0.1826	8	4	484
0		opportunity	4398	9314	0	0	11	5	0
0<b<0.1	Crucial	point	4267	29662	71	0.0166	5	1	65
0		night	4267	30076	0	0	5	1	0

The experimental sentences used in Experiment 4 are shown in table 8.3.

Table 8.3: Experimental sentences for Experiment 4

Experimental sentences - bigrams in **bold**; T/F statements indented and in *italics*
The **nineteenth century** was a time of religious revival and controversy.
There was no religious controversy in the nineteenth century.
The **nineteenth position** in the competition’s scoring system is second to last.
The competition has a scoring system.
The **prime minister** was seen by the public to be working against political trends.
The public saw that the prime minister was working against political trends.
The **prime period** for international expansion by transnational corporations was between 1963 and 1972.
International expansion by transnational corporations mostly took place in the 1980s.
In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.
In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.
An **integral part** of delegation is the setting of targets for those given responsibility for tasks.
Setting targets is an integral part of delegation.
An **integral thought** came into his mind but he was unable to articulate it.
He was able to clearly articulate the integral thought.
Twenty-four hours have indeed passed but the end of a day does not necessarily mean sleep.
Twenty-six hours have passed.
Twenty-four patients were recruited for the study into childhood asthma.
The study focused on childhood obesity.
Disposable income will be severely reduced following the recent tax increases.
The recent tax increases will reduce disposable income.
Disposable property continues to pile up in the streets yet the council refuses to collect it.
There is disposable property piling up in the streets.
The **minimum wage** is designed to help people in low pay service industries.
The minimum wage is designed to help people in high pay service industries.
The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.
The minimum prize included £500 plus four nights in a luxury hotel.
The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
Most people do not lose weight on 1,500 calories per day.
The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
The competition was in the March issue of the Clothes Show Magazine.
It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.
A very important point was raised in parliament.
It was a **crucial night** for basketball in this country and there were many disappointed fans.
Basketball fans were very happy.

Of the eight experimental sentences that were new in this experiment, three were taken word-for-word from the BNC1994 concordance lines, and the remaining sentences were edited from concordance lines, as described in Chapter 4 (section 4.5.3). The same two practice sentences were used in this experiment as in the previous experiments. Moreover, following Experiments 2 and 3, no fillers were used in Experiment 4. See appendix 3 for the four counterbalanced lists of experimental stimuli.

8.3.3 Procedure

As with the previous stimuli set (see Chapter 4, section 4.5.3), the sentences were placed into 4 differently ordered lists to control for the order effects associated with repetition priming. The 18 sentences were presented twice to each participant, so each participant was exposed to 36 sentences along with the corresponding true/false statements. As shown in figure 8.1, the 36 sentences were divided into 3 blocks of 12, with a 20-second break after every 6 sentences. This complies with Luck's (2014c:2-3) guidelines that each block of trials should have a duration of 5 minutes, with a 2-minute break *between* blocks, as well as a 20-second break *within* each block. It was also advantageous to have just 3 experimental blocks, rather than the 4 that I used in the previous experiments, as I had found that there tended to be more alpha activity in block 4. As mentioned in Chapter 6 (section 6.3.3), alpha activity is associated with tiredness/fatigue and lack of concentration (Tiago-Costa 2016:90).

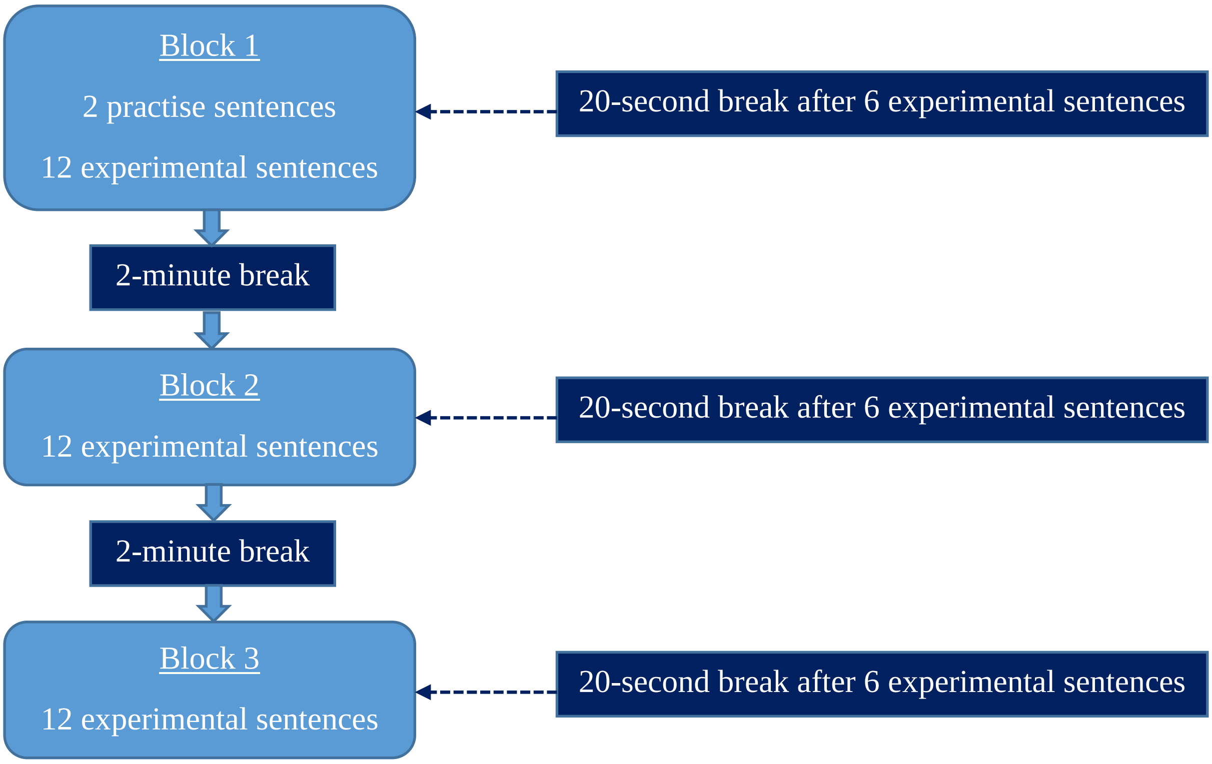


Figure 8.1: Positioning of breaks within and between trial blocks

A final point to note in relation to the stimuli presentation procedure is that, for Experiment 4, I increased the duration of the fixation cross from 1000 ms to 2000 ms. This is the longest length of time that is typically used for fixation periods in ERP studies of language (Swaab et al. 2012:2). The purpose of this change was to reduce the number of blinks occurring immediately after the fixation cross, and therefore reduce the number of datasets lost to excessive artifacts. When running Experiment 3, which had a particularly high participant attrition rate (see Chapter 7, section 7.3), I observed that the excessive artifacts were partly caused by the way in which the participants would often blink immediately after pressing the ‘T’ or ‘F’ key in response to the true/false statements.

The instructions presented to the participants explicitly stated that participants should try to avoid blinking, except during breaks and when reading and responding to the true/false statements (see Chapter 4, section 4.4, figure 4). I verbally reiterated this point, and I showed each participant the effect that blinking has on the waveforms, and I also explained the negative impact that this has on the data. It seemed, then, that the participants were simultaneously trying to blink and press the response key in order to avoid blinking when reading the word-by-word sentences. However, in many cases, the blink occurred slightly after the participant had pressed the response key. In other cases, where the blink and the response *did* occur (almost) simultaneously, the portion of the waveform immediately following the blink and the response would be contaminated by *blink offsets*, i.e. the recovery period immediately following the blink, where the eyelids are opening and the eyes are stabilising (Luck: personal communication). This caused a systematic increase in the number of blink artifacts occurring at the beginning of sentences, which is where the experimental bigrams were located. Similarly, participants would often not regain focus sufficiently quickly after the 20-second break, which again caused excessive artifacts sentence-initially. Thus, by doubling the length of the fixation period, I was able to reduce the amount of artifacts and thereby avoid the high participant attrition rate of Experiment 3.

8.3.4 Data analysis

The measurements taken in Experiment 4 Part 1 were exactly the same as those taken in Experiments 2 and 3, namely: mean amplitude between 350-500 ms at all electrode sites, mean amplitude between 500-650 ms at all electrode sites, and onset latency (50% peak latency) at anterior and central electrode sites. The Part 2 correlational analysis was carried out using only the N400 data. I thought it was best to correlate the transition probability of the collocational bigrams with the amplitude of the N400 (I will discuss how I arrived at a single amplitude value for each bigram shortly), because the N400 results are the most consistent across experiments. In Experiments 1, 2, and 3, there is evidence to suggest that an N400 is elicited in response to reading the second word of the non-collocational bigrams; across all of these experiments, the amplitude in the N400 latency range is lower in the non-collocational condition compared to the collocational condition.

By contrast, the results of the P600 analysis are not as consistent across experiments. Experiment 1 provides some evidence for the elicitation of a P600 in response to reading the second word of a non-collocational bigram, as the amplitude in the P600 latency range is higher in the non-collocational condition compared to the collocational condition. However, these results are seriously undermined, as the latency range used for the P600 in Experiment 1 encompasses a later peak, which is likely to reflect the P3a of the next word (see Chapter 5, section 5.4.5). There is no evidence of a P600 in Experiments 2 and 3 (at least not in response to reading the non-collocational bigrams), as the amplitude in the P600 latency range is higher in the *collocational* condition. Therefore, since a P600 was not elicited in response to reading a non-collocational bigram in two out of three experiments, there is no reason to believe that there would be any correlation between the transition probability of a bigram and the amplitude of the ERP response.

Similarly, while the results of the onset latency analyses reveal that the onset latency is greater in the collocational condition compared to the non-collocational condition in Experiments 1 and 2, the opposite is the case for Experiment 3. The onset latency results are therefore less consistent than the N400 results, making me less able to confidently draw conclusions about how the transition probability of a bigram modulates the onset latency. For Experiment 4 Part 2, it therefore made sense to correlate the transition probability of the collocational bigrams with the amplitude of the N400, rather than attempting to correlate transition probability with P600 amplitude or onset latency as well.

Although the N400 results are consistent across experiments in the sense that the results of all three experiments reveal a lower amplitude in the N400 latency range in the non-collocational condition compared to the collocational condition, the scalp topography of the

N400 varies greatly across experiments. In Experiment 1, the N400 has an anterior-central scalp distribution; in Experiment 2, the N400 is maximal at midline and right hemisphere electrode sites, at all levels of anteriority; in Experiment 3, the N400 is significant at all electrode sites, but (like the traditional semantic N400) is most significant at central-posterior electrode sites. Since the scalp topography of the N400 varies so greatly across experiments, I am unable to draw strong conclusions about where on the scalp the N400 effect is most likely to be detected. For this reason, I decided to conduct the correlational analysis using all electrode sites.

In order to conduct the correlation analysis for Experiment 4 Part 2, I needed to compute an independent amplitude value for each bigram pair that would allow me to correlate the difference in transition probability of the bigram pairs with the amplitude of the ERP response for those bigram pairs[10]. There existed no ready-made method for doing this in ERP research, as this is an entirely novel investigation in the field. Therefore, I devised a new process for computing an independent amplitude value that could be correlated with measures of collocation strength. This process involves four distinct steps, which I will now describe.

First, using the ERP data from all participants, I made a *difference wave* for each bigram pair. In ERP research, a difference wave is a single waveform that is computed by subtracting the waveform for one condition from the waveform for the other condition (Luck 2012:14). In this case, I subtracted the waveform for the collocational bigram from the waveform for the matched non-collocational bigram. The result is a single wave representing the difference in ERP response between conditions (see figure 8.7 for examples).

---

[10] Since the transition probability of the non-collocational bigram pairs was effectively zero, the difference in transition probability of the bigram pairs is equal to the transition probability of the collocational bigram in each pair.

After computing the difference waves for each bigram pair, I took the mean amplitude measurement in the 350-500 ms latency range from each difference wave. I then extracted the N400 values for the nine representative electrode sites used throughout this thesis. Lastly, I calculated the mean of the amplitude values from these nine electrode sites, leaving me with an independent amplitude value for each bigram pair.

Having confirmed the normal distribution of these values, I then conducted a Pearson correlation and created a scatterplot to analyze the correlation. Additional Pearson correlations were later conducted to look for relationships between the amplitude and the other statistical measures of collocation strength that are commonly used in corpus linguistic research (Hoffmann et al. 2008), namely mutual information, MI3, z-score, t-score, log-likelihood, and Dice coefficient, as well as the raw frequency. The final stage of the analysis was to rank the different statistical measures in terms of the Pearson's *r* of their correlation with the ERP response, and thus (by proxy) what may be seen as their psychological validity.

8.4 Results: Experiment 4

8.4.1 Part 1: Replication of Experiment 2

8.4.1.1 Component-based experimental design

8.4.1.1.1 N400 (350 - 500 ms)

The results of Experiment 4 mirror the results of Experiment 2 in that the mean amplitude in the 350-500 ms latency range is lower in the non-collocational condition (*M* = 0.146, *SD* = 2.999) compared to the collocational condition (*M* = 1.132, SD = 3.353), and that there is no significant main effect. However, while the results of Experiment 2 reveal a significant interaction between condition and laterality but no interaction between condition and anteriority, the results of Experiment 4 reveal the opposite: there is no interaction between condition and right-to-left electrode position but there *is* a significant interaction between condition and anterior-to-central electrode position: *F*(2, 820) = 7.28, *p* = .001. This will be discussed further in section 8.5.2.

The grand average ERP waveforms presented in figure 8.2 show that there is a greater negativity in the N400 time window in the non-collocational condition at central and posterior electrode sites, but there is no clear evidence for an N400 effect at anterior electrode sites.

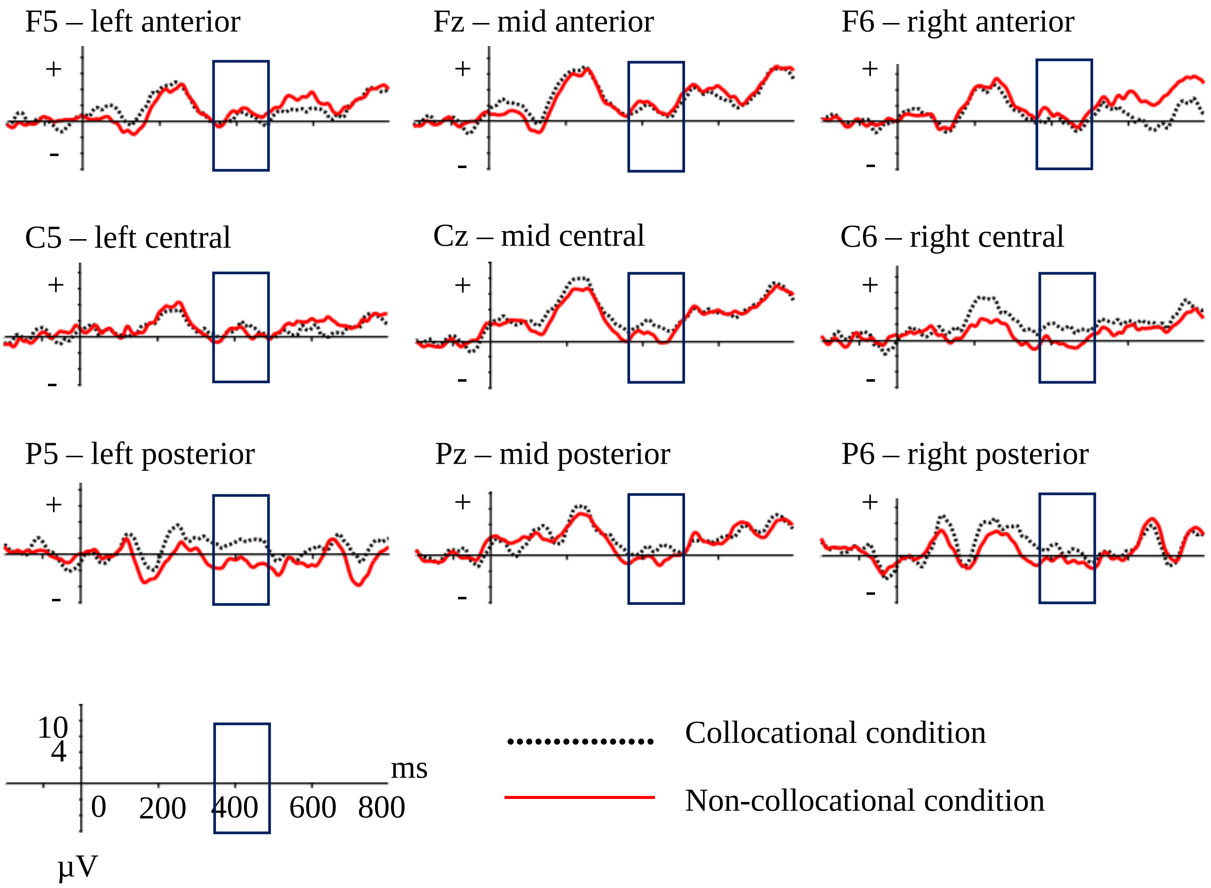


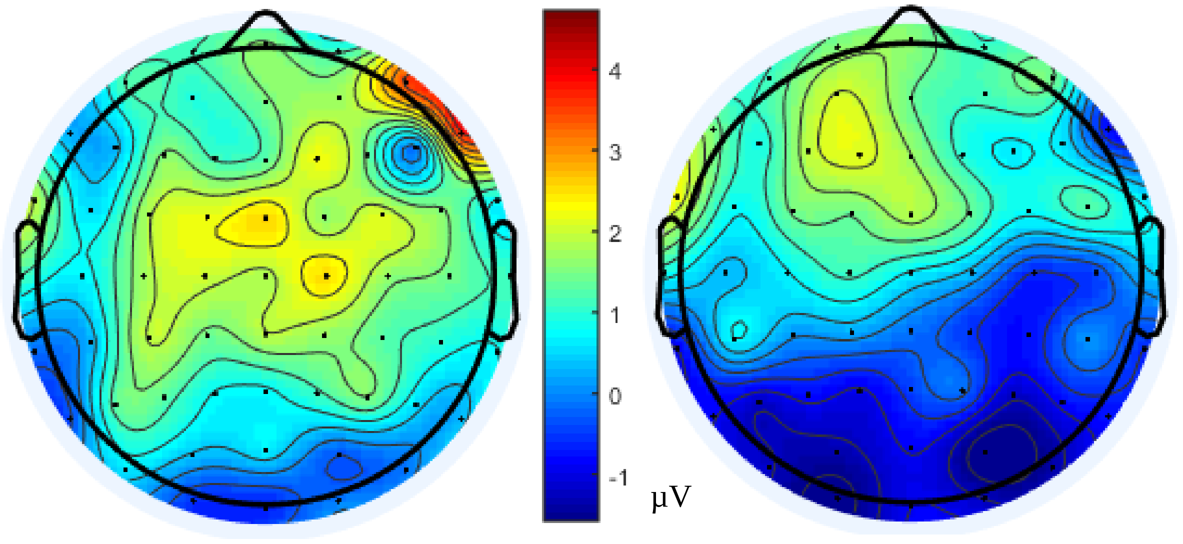
Figure 8.2: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus.

Indeed, as is shown in table 8.4, the results of subsequent repeated measures ANOVAs carried out separately at anterior, central, and posterior electrode sites reveal that there is no significant effect at anterior electrode sites, but the effect is highly statistically significant at central (*F*(1, 346) = 32.529, *p* < .001) and posterior (*F*(1, 257) = 47.072, *p* < .001) electrode sites.

Table 8.4: Summary of post hoc ANOVA results carried out at anterior, central, and posterior electrode sites in the 350-500 ms latency range

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Anterior	1.393	3.84	1.144	3.165	.384
Central	1.419	2.788	0.383	2.968	< .001***
Posterior	0.524	3.535	-1.019	2.482	< .001***

In contrast to the results of Experiment 2, which suggest an N400 effect in the absence of the typical scalp distribution of the N400, the results of Experiment 4 reveal an N400 effect with a central-posterior scalp distribution (as shown in figure 8.3), which is the same as that of the classical semantic N400 (Swaab et al. 2012:399). These results will be discussed in section 8.5.2.



Collocational condition

Non-collocational condition

Figure 8.3: Topographic scalp maps showing mean amplitude between 300 and 500 ms

8.4.1.1.2 P600 (500 - 650 ms)

The results of the P600 analysis reveal that the mean amplitude in the 500-650 ms latency range is higher in the non-collocational condition (*M* = 0.754, *SD* = 7.29) compared to the collocational condition (*M* = -0.119, *SD* = 9.2). These findings are in opposition to those found in Experiments 2 and 3, and they suggest that a P600 might actually be elicited by reading the second word of a non-collocational bigram. However, the difference between conditions is not statistically significant: *F*(1, 756) = 6.76, *p* = .992. There is also no significant interaction between condition and central-posterior electrode position, or condition and left-to-right electrode position.

Looking at figure 8.4, the difference between conditions appears to be marginal at most of the representative electrode sites. Moreover, where the difference *does* appear to be clear, that is, at site F6, it is notable that the difference is most apparent in the latter half of the 500-650 ms time window. This suggests that, as in Experiment 1 (see Chapter 5, section 5.4.6), the putative P600 might actually be the P3a of the next word. Even so, the difference between conditions at site F5 cannot be attributed to the P300 effect, as the difference is apparent throughout the whole 500-650 ms time window, and there is a clear trough in the waveform between this and the later peak for the next word's P3a.

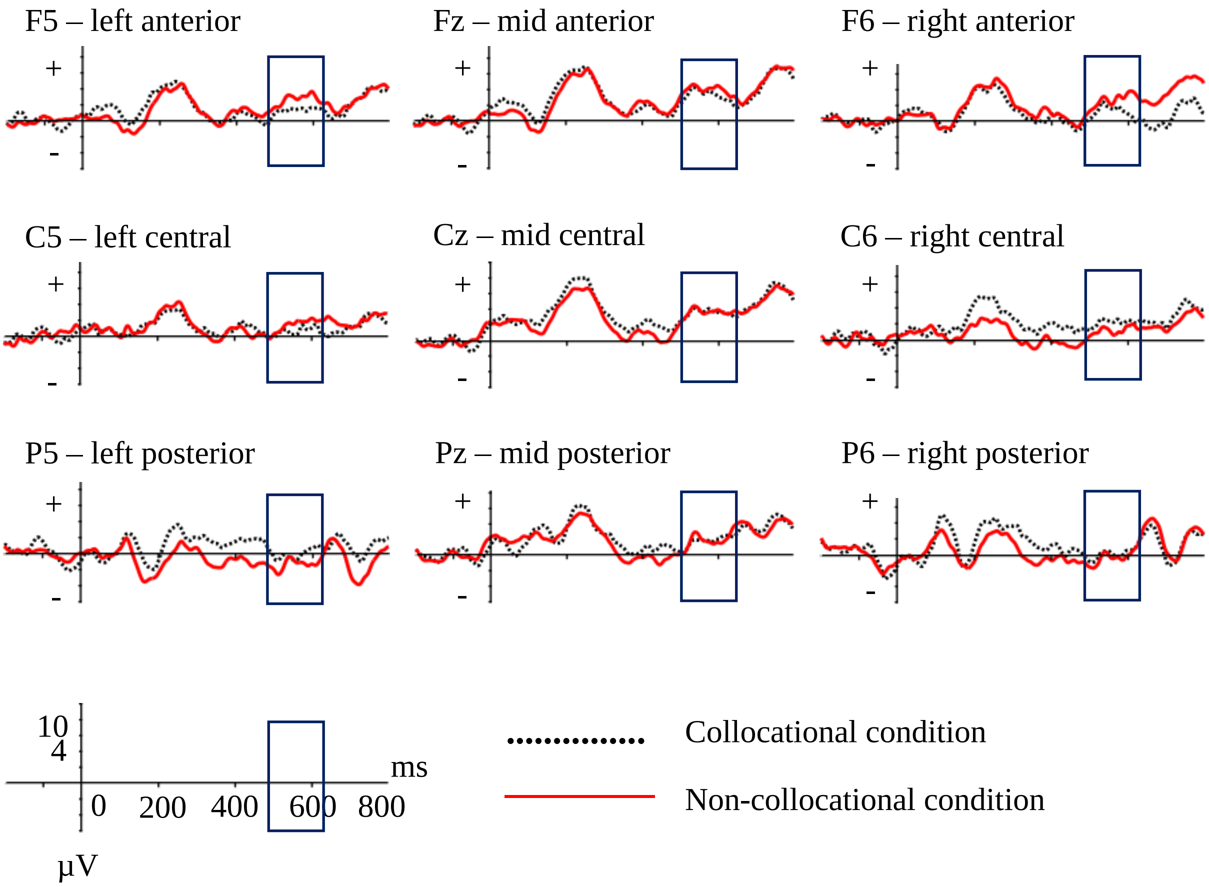
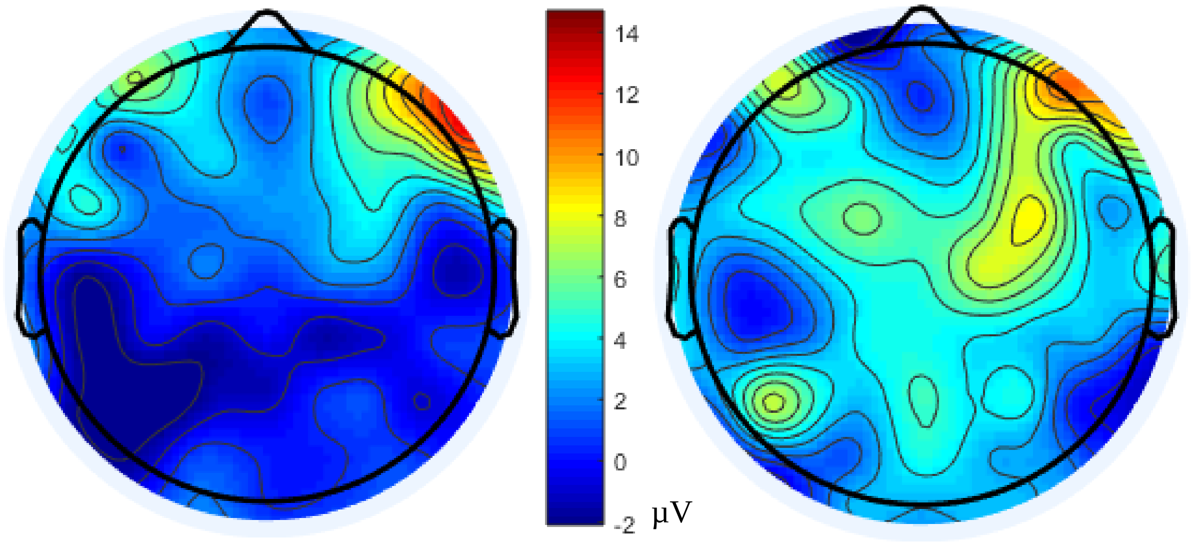


Figure 8.4: Grand average ERPs across a representative sample of electrode sites, from -200 ms pre-stimulus to 800 ms post-stimulus

Although the results of the P600 analysis are not significant, the topographic scalp maps in figure 8.5 clearly show that there is a greater positivity in the non-collocational condition at central and posterior electrode sites. This scalp distribution is in line with the scalp distribution of the classical P600 (Ingram 2007:323; Friederici & Männel 2013:186). These results will be discussed further in section 8.5.



Collocational condition

Non-collocational condition

Figure 8.5: Topographic scalp maps showing mean amplitude between 500 and 650 ms

8.4.1.2 Component-independent experimental design

As with the N400 results, the onset latency results mirror those of Experiment 2, as the onset latency is greater in the collocational condition (*M* = 448.077, *SD* = 246.448) compared to the non-collocational condition (*M* = 439.573, *SD* = 228.678). This constitutes a significant difference: *F*(1, 534) = 7.716, *p* = .006.

As well as there being a significant main effect, there is also a significant interaction between condition and left-to-right electrode position: *F*(1, 534) = 8.067, *p* = 0.005. Table 8.5 shows that the effect is significant at left hemisphere and right hemisphere electrodes sites, but not at midline electrode sites. However, the result for the right hemisphere is significant *in the opposite direction*, with the onset latency being greater in the non-collocational condition compared to the collocation condition.

Table 8.5: Summary of post hoc ANOVA results carried out at anterior and central electrode sites

Electrode Position	Collocational		Non-collocational		*p*-value
	M	*SD*	*M*	*SD*	
Left hemisphere	456.26	253.259	414.013	226.632	.023*
Right hemisphere	400.889	239.938	444.256	230.263	.004**
Midline	458.654	249.534	466.112	232.303	.787

Since this onset latency analysis was carried out using the anterior and central electrode zones, the results show that the onset latency effect occurs at left anterior-central electrode sites. The onset latency results will be discussed further in Chapter 9 (section 9.2). I now move on to presenting the results of the Part 2 analysis.

8.4.2 Part 2: Correlational analysis

The results of the Part 2 analysis reveal that there is a strong negative correlation (though not a significant correlation) between the transition probability of the collocational bigrams and the difference in amplitude between the two conditions: $r = -0.621$, $p = 0.074$. This correlation can be seen on the scatterplot in figure 8.6.

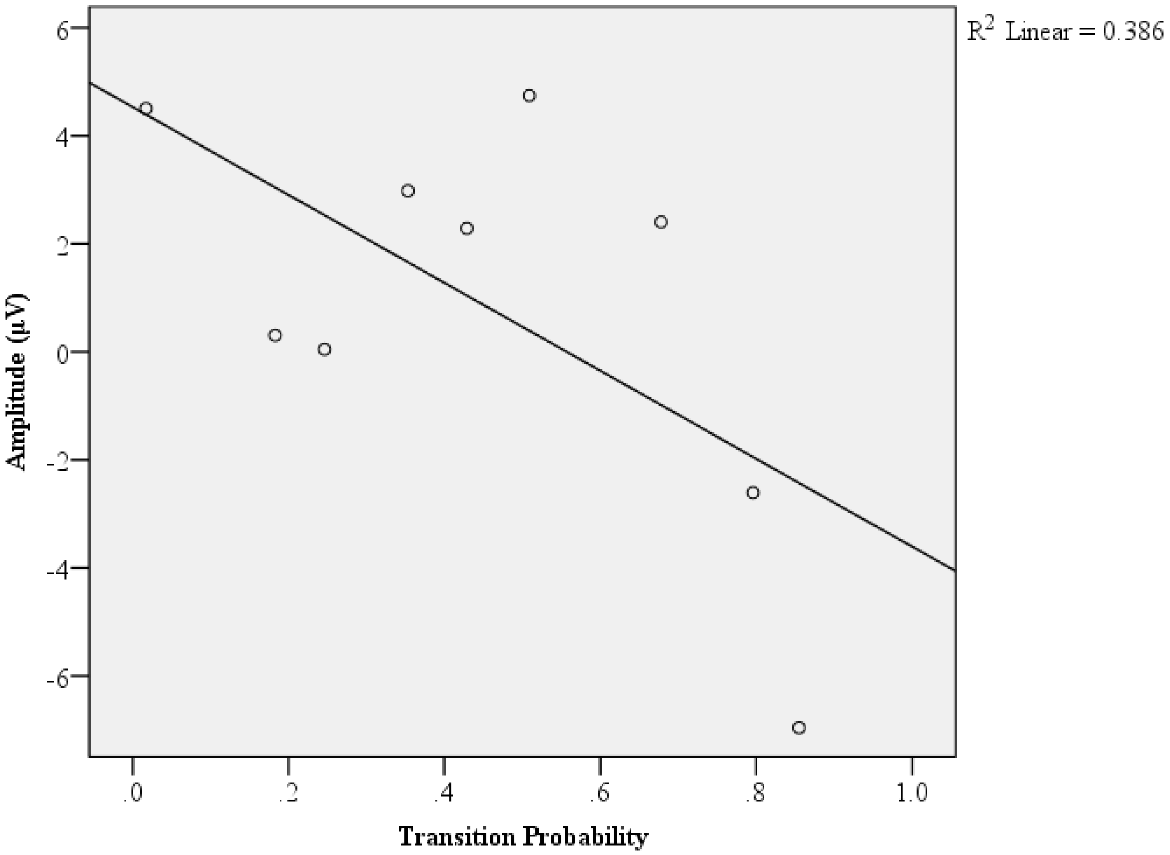


Figure 8.6: Scatterplot showing the relationship between amplitude and transition probability

The correlation is negative, with a lower amplitude value for the collocational bigrams with the higher transition probabilities, because the amplitude value represents the amplitude of the N400 (as measured from a difference wave, showing the difference in amplitude between the two conditions). The lower the amplitude, the bigger the difference between conditions in the N400 latency range. In turn, the bigger the difference between conditions, the larger the N400, and therefore the bigger the experimental effect.

I am, therefore, interpreting the amplitude of the N400 as proportional to the size of the cognitive load associated with reading the second word of a non-collocational bigram. When a collocational bigram has a high transition probability, there is a stronger expectation for what word will follow the first word. Therefore, breaking this strong expectation will result in a greater increase in cognitive load than breaking a weaker expectation associated with a collocational bigram that has a lower transition probability. Thus, the results of this correlational analysis are in line with Hypothesis 4: there *is* a correlation between the transition probability of a bigram and the amplitude of the ERP response. When a participant reads a non-collocational bigram that is matched with a bigram with a high transition probability, this is associated with a lower amplitude, and therefore a larger N400, and a large increase in cognitive load; by contrast, when a participant reads a non-collocational bigram that is matched with a bigram with a relatively lower transition probability, this is associated with a higher amplitude, and therefore a smaller or non-existent N400, and a comparatively smaller (or non-existent) increase in cognitive load.

Figure 8.6 shows that the value of R^2 is 0.386, indicating that transition probabilities account for 38.6% of the variance in amplitude. This is a reasonably high value, especially when we consider that corpus-derived transition probability is only a proxy for psychological transition probability, and that even psychological transition probabilities are different for every individual (see sections 2.4.2 and section 4.2.1).

For reference, figure 8.7 shows a difference wave for each bigram pair, taken at the most central scalp electrode site (Cz). It is notable that the largest N400 is evident in the difference wave for the bigram pair *nineteenth century/position* – the pair with the collocational bigram that has the highest transition probability in this study.

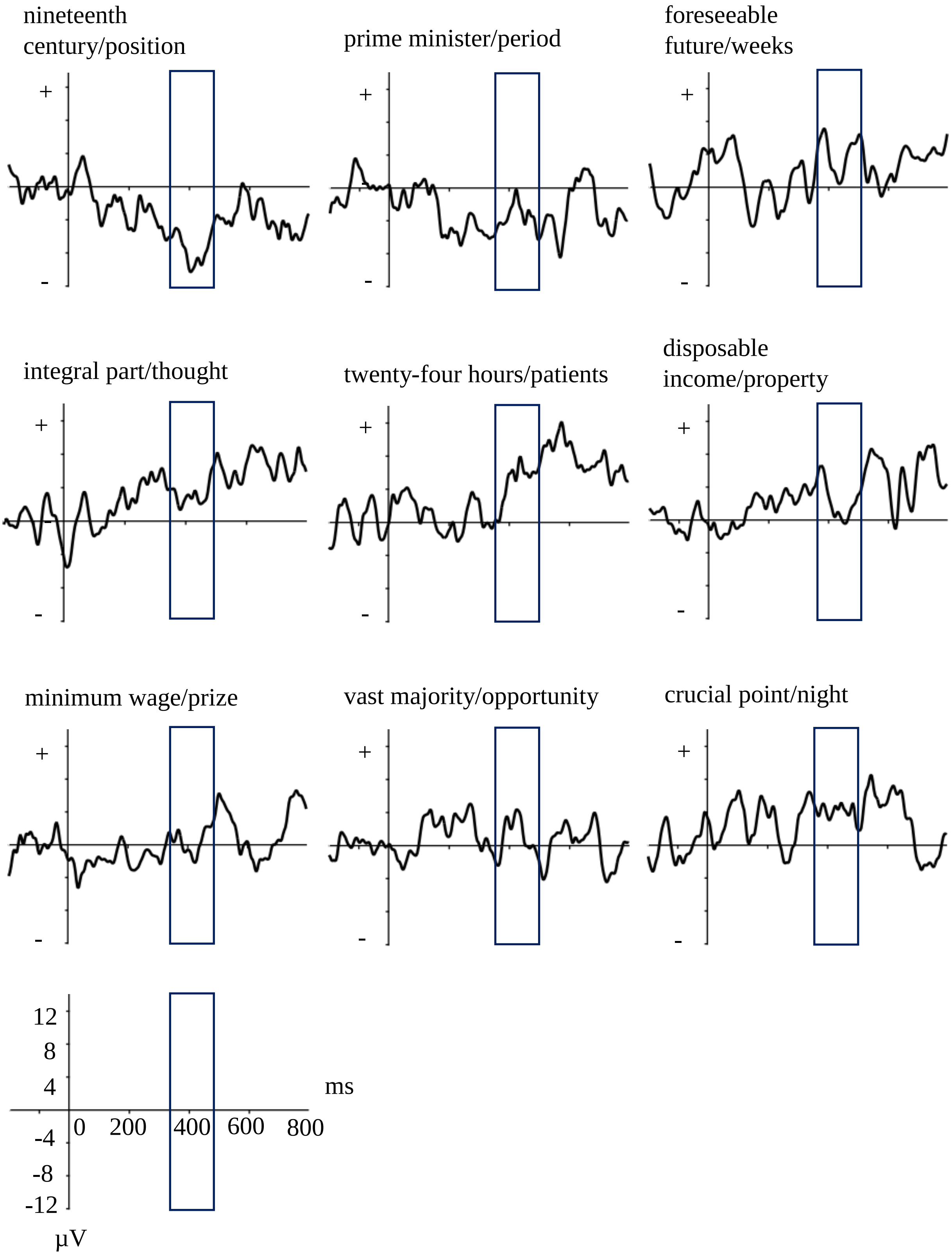


Figure 8.7: Difference waves at Cz for each bigram pair, from -200 ms pre-stimulus to 800 ms post-stimulus. The box on each waveform shows the N400 latency range.

Although there is a strong correlation between transition probability and ERP amplitude, with transition probability accounting for 38.6% of the variance in amplitude, the results of subsequent Pearson correlations show that transition probability is not the statistical measure of collocation strength that is *most* strongly correlated with the amplitude of the ERP response. Table 8.6 ranks the statistical measures of collocation strength in order of strength of correlation (Pearson's r), with the strongest in the top row and the weakest in the bottom row; table 8.7 shows the amplitude value of each bigram pair, along with the association measure values of each collocational bigram.

Table 8.6: Association measures ranked by strength of correlation

Association measure	Pearson's *r*	*p*-value
1. Z-score	-0.773	0.014*
2. MI3	-0.772	0.015*
3. Dice coefficient	-0.712	0.031*
4. T-score	-0.679	0.044*
5. Frequency	-0.636	0.065
6. Transition probability	-0.621	0.074
7. MI	-0.575	0.105
8. Log-likelihood	-0.566	0.112

Table 8.7: Amplitude and association scores for each collocational bigram

Collocational bigram	Amplitude value (of bigram pair)	TP	Mutual information	MI3	Z-score	T-score	Log-likelihood	Dice coefficient	Frequency
nineteenth century	-6.962	0.855	12.136	34.84	3428.27	51.116	42642.448	0.237	569
prime minister	-2.606	0.796	12.831	38.091	5620.4	96.201	147416.41	0.529	773
foreseeable future	2.402	0.678	11.663	27.901	947.547	16.668	4199.455	0.026	234
integral part	4.745	0.509	10.005	28.472	785.109	24.512	7531.789	0.024	434
twenty-four hours	2.286	0.429	11.374	28.365	977.024	18.992	5163.947	0.042	222
disposable income	2.98	0.353	11.632	25.911	666.426	11.87	2050.59	0.024	80
minimum wage	0.044	0.246	12.063	28.94	1216.32	18.624	5196.976	0.122	109
vast majority	0.305	0.183	10.941	30.24	1255.18	28.323	10802.384	0.117	484
crucial point	4.506	0.017	5.679	17.978	58.708	8.262	421.022	0.004	65

While there is a strong correlation between amplitude and *all* of the statistical measures tested in this study, the strongest correlation exists between amplitude and z-score (see figure 8.8), closely followed by amplitude and MI3, and amplitude and Dice coefficient. Meanwhile, the weakest correlation exists between amplitude and log-likelihood (figure 8.9).

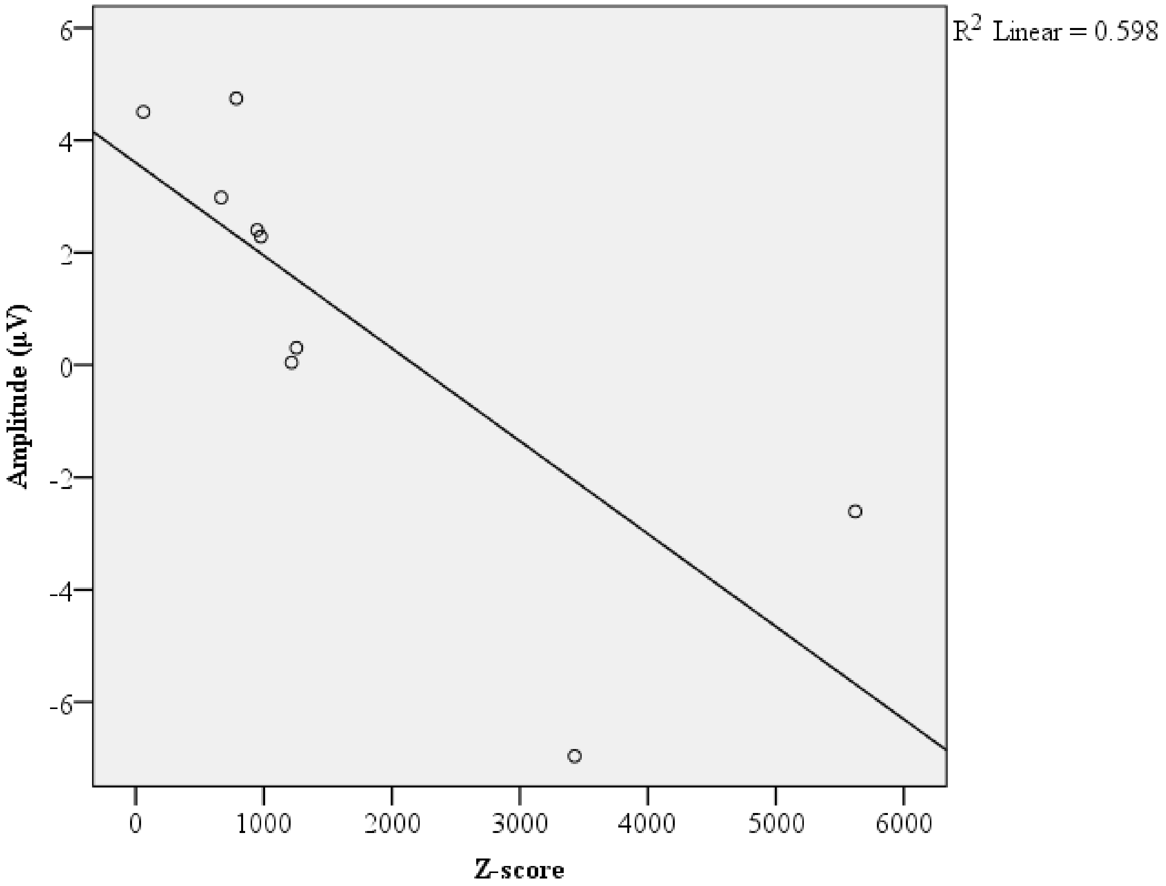


Figure 8.8: Scatterplot showing the relationship between amplitude and Z-score

At 0.598, the value of R² is particularly high in the case of the relationship between amplitude and Z-score. This indicates that the strength of the association as measured via the Z-score measure of association strength accounts for 59.8% of the variations in amplitude. Meanwhile, at 0.320, the value of R² is still reasonably high in the relationship between amplitude and log-likelihood, even though the results of this experiment suggest that log-likelihood is the statistical measure of collocation strength that is the least psychologically valid. An R² value of 0.320 indicates that log-likelihood accounts for 32% of the variance in amplitude.

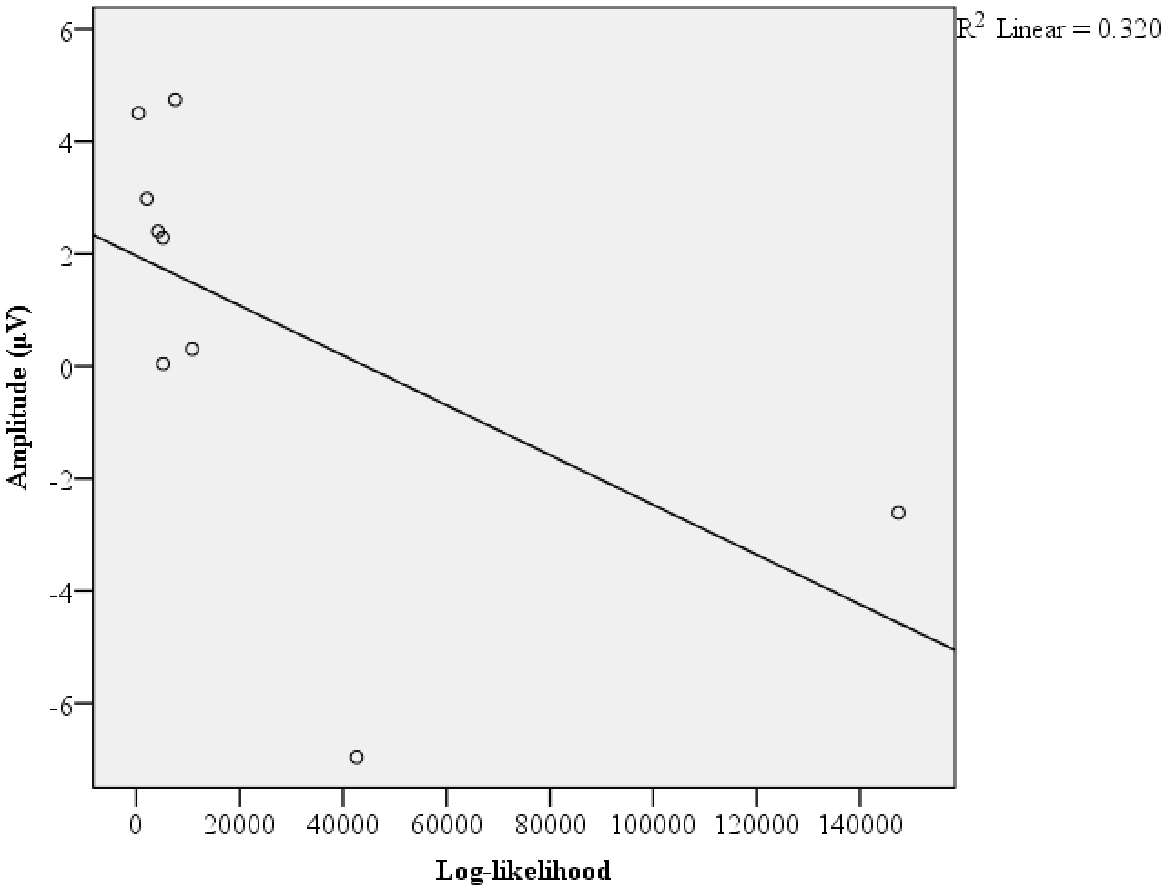


Figure 8.9: Scatterplot showing the relationship between amplitude and log-likelihood

The scatterplots for the remaining five association measures are shown in figures 8.10 through to 8.14.

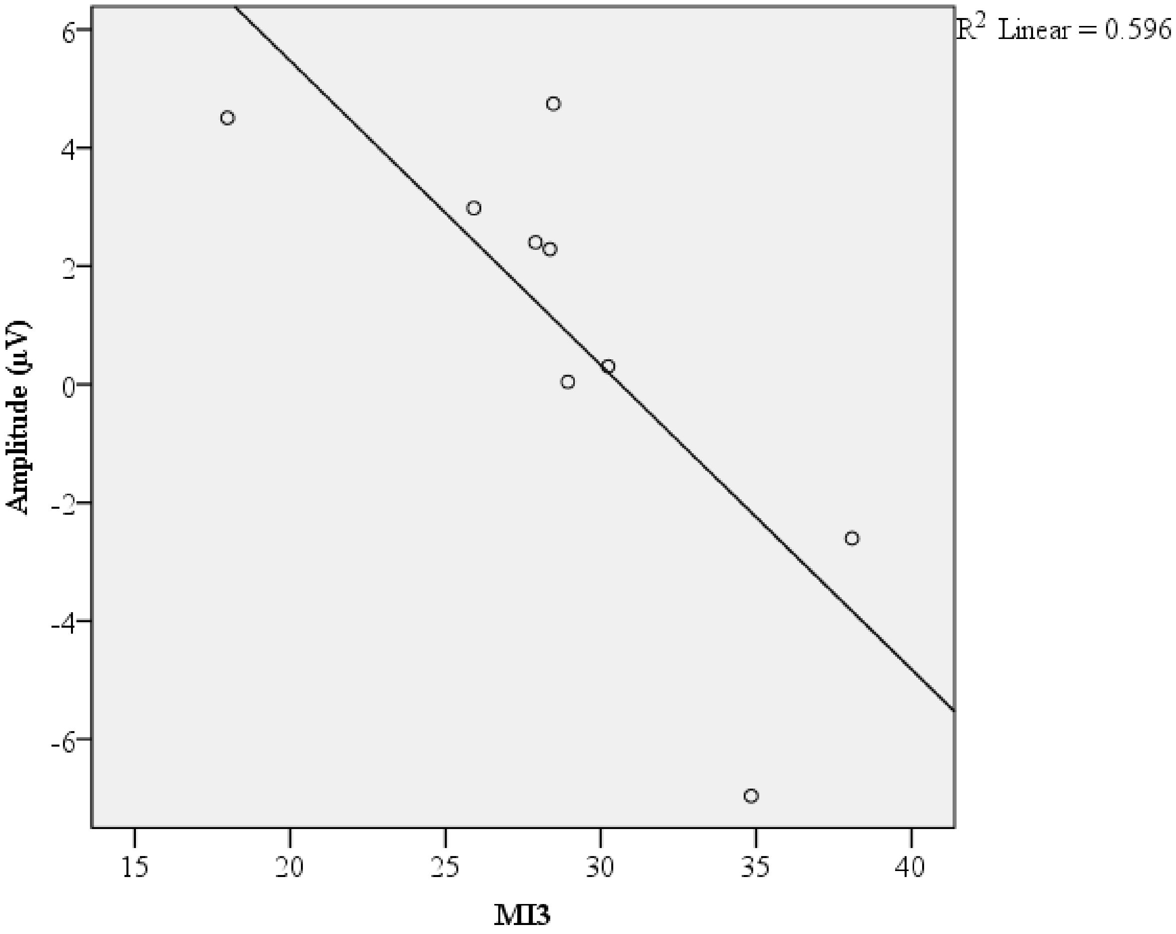


Figure 8.10: Scatterplot showing the relationship between amplitude and MI3

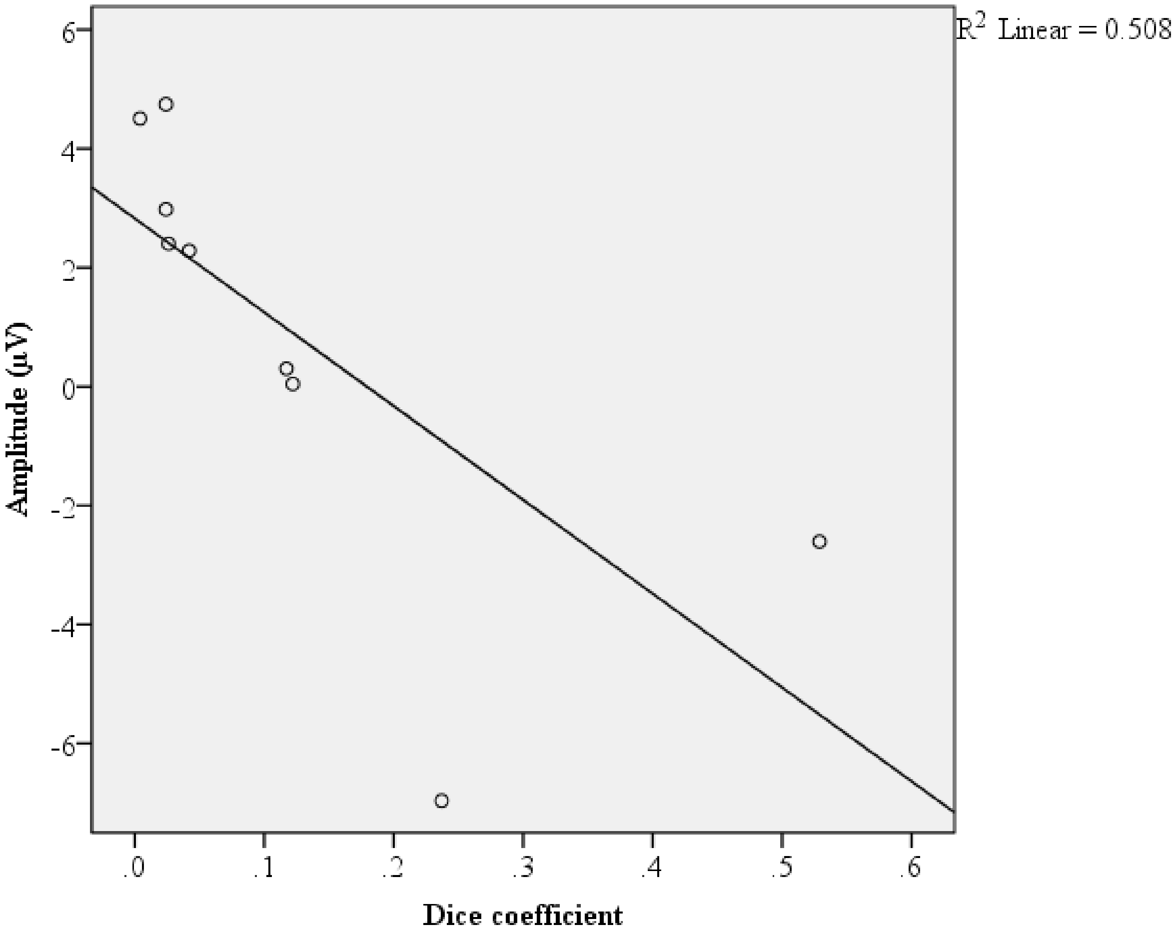


Figure 8.11: Scatterplot showing the relationship between amplitude and Dice coefficient

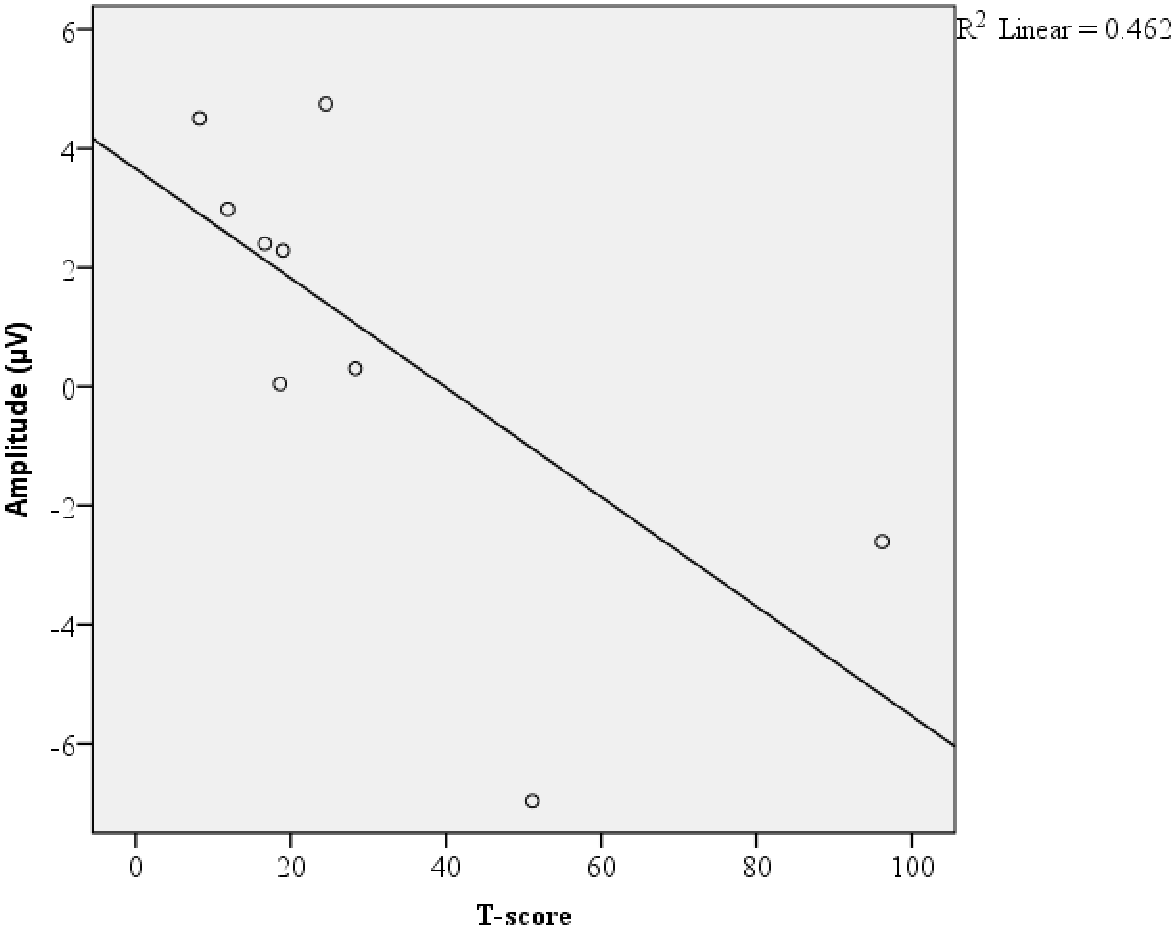


Figure 8.12: Scatterplot showing the relationship between amplitude and T-score

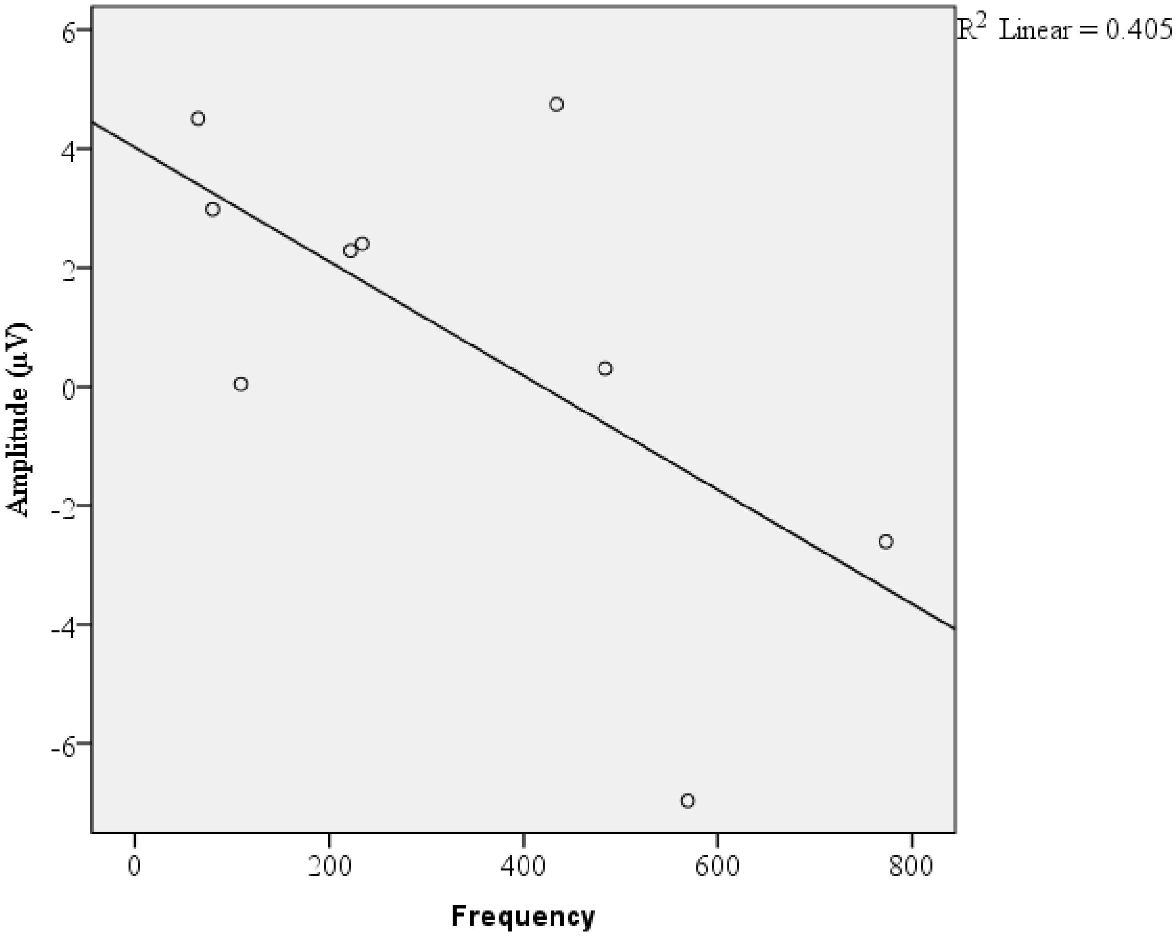


Figure 8.13: Scatterplot showing the relationship between amplitude and frequency

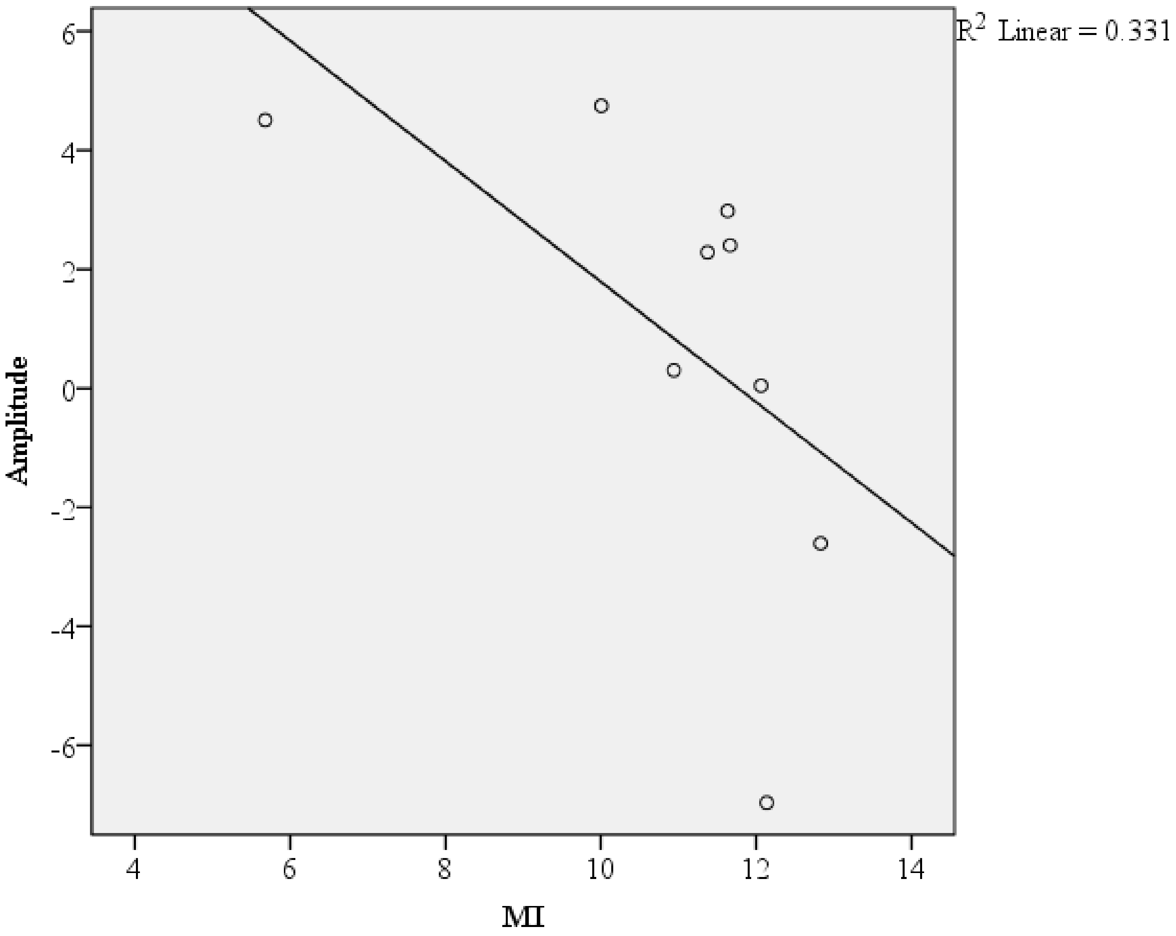


Figure 8.14: Scatterplot showing the relationship between amplitude and MI

Looking back at table 8.6, it is interesting to see that the four statistical measures which are *most* strongly correlated with amplitude (and indeed also significantly correlated) are hybrid measures, meaning that they measure collocation strength by combining two different statistical dimensions (see Chapter 2, section 2.11). In contrast, the four statistical measures which are more weakly correlated with amplitude (albeit still strongly correlated) are those that measure collocation strength using a single parameter: either statistical significance (log-likelihood), effect size (mutual information, transition probability), or raw frequency. This suggests that the hybrid measures are those with the highest level of psychological validity. The implications of this will be discussed in section 8.5.2.

8.5 Discussion: Experiment 4

8.5.1 Summary of aims and results

The aims of Experiment 4 were (1) to replicate Experiment 2 in order to strengthen the confidence of its conclusions, (2) to investigate the strength of the correlation between the transition probability of a bigram and the amplitude of the ERP response, and (3) to find out which measure of collocation strength can be seen as being the most psychologically valid.

The results mirror those of Experiment 2 by showing that the onset latency is greater in the collocational condition, and that reading the second word of a non-collocational bigram elicits an N400. This therefore confirms Hypotheses 1 and 3, and suggests that the results of the previous experiments do actually reflect collocational processing as opposed to semantic processing (see Chapter 9, section 9.5 for associated limitations)However, the finding that reading the second word of a non-collocational bigram elicits a potential P600 is in opposition to the results of Experiment 2, which show that the potential P600 is actually elicited in the collocational condition. This provides evidence against Hypothesis 2. That said, it is important to remember that the P600 results of Experiment 4 are not significant, and that they may be influenced by the P3a of the following word (see section 8.4.1.1.2).

The results of Experiment 4 Part 2 confirm Hypothesis 4 by showing that there *is* a correlation between the transition probability of a bigram and the amplitude of the ERP response. Yet, although there is a strong correlation between transition probability and ERP amplitude, with transition probability accounting for 38.6% of the variance in amplitude, the results of subsequent Pearson correlations show that transition probability is not the statistical measure of collocation strength that is *most* strongly correlated with the amplitude of the ERP response. Rather, the strongest correlation exists between amplitude and z-score, closely followed by amplitude and MI3, and amplitude and Dice coefficient. Meanwhile, the weakest correlation exists between amplitude and log-likelihood.

8.5.2 Issues raised by the results

Two key questions need to be addressed in relation to Experiment 4 Part 1. Since the results reveal an N400 and a potential P600, both of which have the classical central-posterior scalp distributions, why does the N400 found in previous experiments not share this classical scalp distribution, and why do the results of previous experiments show no clear evidence for a P600? The answer could be related to the transition probabilities of the experimental bigrams. Compared to the stimuli set used in Experiments 1, 2, and 3, the stimuli set used in Experiment 4 contains more collocational bigrams with an exceptionally high transition probability. The mean transition probability of the bigrams in Experiments 1/2/3 is 0.231; the mean transition probability of the bigrams in Experiment 4 is almost double that, at 0.452. The higher the transition probability, the stronger the collocation, and thus the stronger the expectation that word Y will follow word X. In turn, the stronger the expectation, the greater the increase in cognitive load when that expectation is violated. It could be the case, then, that the N400 and P600 are only elicited ‘fully’ (i.e. in a way that resembles the classical components) when the cognitive load in the non-collocational condition is exceptionally high.

This explanation is supported by a comparison of the magnitude of the N400 effects in Experiments 2 and 4. The range between conditions in the N400 time window in Experiment 2 is 0.545, compared to 0.986 in Experiment 4. Thus, it could be argued that the higher the transition probability of the collocational bigrams, the greater the amplitude of the N400 effect. This explanation is also in-fitting with the results of the Lau et al. (2016) study because, as with the results of Experiment 4, the results of the Lau et al. (2016) study reveal that a central-posterior N400 is elicited in response to reading collocational bigrams with an exceptionally high transition probability (0.5 and above). However, this explanation is not supported by the onset latency results, as the range between the onset latency of the collocational condition and the onset latency of the non-collocational condition is actually much smaller in Experiment 4 (8.504) compared to Experiment 2 (69.21).

There are two issues raised by the results of Experiment 4 Part 2. First, the finding that mutual information and log-likelihood are the statistical measures of collocation strength that are least strongly correlated with amplitude (and therefore the cognitive load caused by breaking the collocational expectation set up by the first word of a bigram) has important implications for corpus linguistics. Specifically, since mutual information and log-likelihood are two of the most widely used measures in corpus linguistics (Gries 2014a:37), and over a third of corpus studies use collocation analysis (Gilquin & Gries 2009), this suggests that the many studies which have utilised these measures of association strength have not actually used the measure that has the most psycholinguistic validity (according to the results of the present study). Nevertheless, even if this is the case, the results of the present study do not in any way invalidate the results of corpus studies which utilise the measures of association strength which this study has shown to be the least psychologically valid. After all, even though mutual information and log-likelihood are the measures that are least strongly correlated with amplitude, the correlation is still strong.

Moreover, it is not the case that there is one measure of collocation strength that is optimal in all studies. That is, just because the z-score measure of association strength has been shown to have the most psychological validity in this study, this does not mean that this measure will automatically be the best measure to use in all corpus studies. After all, the psychological validity of an association measure will not always be a relevant factor to take into account, and other studies might use different conceptualizations of psychological validity, which will mean that other measures are preferable. For example, raw frequency would be a better measure to use in a study which is interested in the level of entrenchedness of a collocation. This is because, as mentioned in Chapter 2 (section 2.11), entrenchedness is determined by the frequency with which a collocation is encountered (Gries 2008b:409-410, 413; 2014:11; 2017:11). The precise measure of association which is chosen should always depend on the purpose of the study.

Secondly, the finding that raw frequency is more strongly correlated with amplitude than transition probability, mutual information, or log-likelihood is in opposition to the results of Ellis and Simpson-Vlach (2009), who find that mutual information is more indicative of psycholinguistic validity than raw frequency (see section 8.2). This finding also opposes the results of Gries et al. (2005), who find that an effect size measure of collocation strength (specifically, collostruction strength) outperforms raw frequency in sentence-completion and reading-time tasks. However, the finding that raw frequency is more strongly correlated with amplitude than mutual information is in line with Wiechmann's (2008) finding that raw frequency outperforms mutual information when correlated with eye-fixation data.

The results of the present study are arguably stronger and more convincing than the results of previous studies, as EEG provides more direct evidence of how language is processed in the brain than sentence-completion and reading-time tasks, eye-tracking tasks, or the psycholinguistic tasks used by Ellis and Simpson-Vlach (2009) (specifically, rate of reading in a grammaticality judgement task, two variations of a read-aloud task, and a semantic judgement task).

Moreover, it is intuitively plausible for frequency to correlate highly with the psychological reality of collocation. As Wray (2002:25) points out:

> the more often a string is needed, the more likely it is to be stored in prefabricated form to save processing effort, and once it is so stored, the more likely it is to be the preferred choice when that message needs to be expressed.

This relates to Gries' claim that the raw frequency of a collocation determines the extent to which that collocation is entrenched in memory (Gries 2008b:409-410, 413; 2014:11;

2017:11). However, although explicitly linking frequency with entrenchedness, Gries (2010:71) himself distances himself from this idea and critiques the use of raw frequency, stating that:

> while cognitive linguists regularly regard frequency data as directly reflecting the degree of routinization or entrenchment, we have shown that (i) frequency alone runs the risk of severely misrepresenting speakers' behavioural patterns and that (ii) collostructional strength [an effect size measure] outperforms frequency as a predictor of speakers' behaviour in both production and comprehension tasks (Gries et al. 2010:71).

Similarly, Wray (2002:30) claims that raw frequency is arguably an inadequate measure of collocation strength because "many word strings are indisputably formulaic, but not frequent". Further issues raised by these results will be discussed in Chapter 9 (section 9.4).

8.6 Chapter summary and conclusion

In this chapter I have presented the final experiment of this thesis. The purpose of this experiment was twofold. First, to replicate and thus strengthen the results of Experiment 2; and, second, to investigate the psychological validity of different measures of collocation strength. The results of this experiment mirror those of Experiment 2 by again showing that the onset latency is higher in the collocational condition, and that reading the second word of a non-collocational bigram elicits an N400. The N400 found in Experiment 4 is different from that of Experiment 2, as it shares the central-posterior scalp distribution of the classical semantic N400. Moreover, unlike in Experiment 2, the results of Experiment 4 show evidence of a potential P600 being elicited in the non-collocational condition. I explain the possible presence of a P600, as well as the classical scalp distribution of the N400, by referring to the comparatively greater strength of the collocations used in Experiment 4.

The findings presented in this chapter also suggest that the statistical measures of collocation strength that are commonly used in corpus linguistics (Hoffmann et al. 2008) *do* have psychological validity. The statistical measures that most strongly correlate with the ERP response are the hybrid measures, which measure collocation strength by combining both statistical significance and effect size. By contrast, the measures which correlate less strongly with the ERP response are those which measure collocation strength using only a single parameter, such as statistical significance, effect size, or raw frequency. Two of the most widely-used measures in corpus linguistics, namely log-likelihood and mutual information (Gries 2014a:37), were found to have the least strong correlations. Therefore, the results of this experiment have important implications for the field of corpus linguistics. These will be discussed further in the following chapter (section 9.3).

CHAPTER 9: GENERAL DISCUSSION AND CONCLUSION

9.1 Overview of the chapter

In section 9.2 of this chapter, I summarise the aims and results of the four ERP experiments conducted in this thesis, and I explicitly state the conclusions that can and cannot be drawn from the data. In this summary of my aims and results, I do not make any reference to the issue of scalp distributions of the ERP components; this issue requires a much longer discussion, and is therefore postponed to section 9.4.1.

In section 9.3, I discuss the implications of the results, first in terms of what the results add to the debate of whether there are one or two linguistic processing systems, and then in terms of the implications that the results have for the network model of language processing introduced in Chapter 2. In section 9.4 I then discuss the issues raised by the results. These relate to the inconsistencies in scalp distribution data (section 9.4.1), the elicitation of a putative P600 in the collocational condition (section 9.4.2), and the way in which the N400 and onset latency results from Experiment 3 suggest conflicting conclusions (section 9.4.3). In section 9.5 I then outline the limitations of the study and suggest some directions for future research. Finally, in section 9.6, I explain how this thesis constitutes a novel contribution to the fields of corpus linguistics and cognitive neuroscience, and I reiterate the main findings of the study.

9.2 Summary of findings

9.2.1 Experiment 1

The aims of Experiment 1 were to pilot a procedure for determining whether or not there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams, and to refine the hypotheses and methods in preparation for Experiments 2 and 3 (see Chapter 5, section 5.1). The experimental procedure was successful for the N400 analysis as well as the onset latency analysis, as these analyses produced clear results that formed the basis of the hypotheses for Experiments 2 and 3.

Specifically, the results of the component-based analysis of Experiment 1 show that there is a slightly positive amplitude in the collocational condition and a slightly negative amplitude in the non-collocational condition. The difference between conditions is statistically significant, suggesting that an N400 is elicited in response to reading the second word of a non-collocational bigram. This therefore provides evidence in support of Hypothesis 1. Furthermore, the results of the component-independent analysis of Experiment 1 show that the onset latency is greater in the collocational bigram condition compared to the non-collocational bigram condition. In other words, the time point at which the amplitude is half of the value of the peak amplitude in the grand average waveforms occurs earlier in the non-collocational bigram condition compared to the collocational bigram condition. This demonstrates that there is a neurophysiological difference between conditions independent of the identification of specific ERP components. I therefore achieved Aim 2, as I was able to use these findings as the basis for the hypotheses in experiments 2 and 3.

The experimental procedure was less successful for the P600 analysis, as I do not have sufficient evidence to claim that there is a P600 in my data. This is potentially due to the way in which I measured the mean amplitude of the P600 in the 500-800 ms latency range, which overlaps with the latency range of the putative P750/the P3a of the next word. I therefore decided to use a tighter time window when measuring the P600 in subsequent experiments. Moreover, although the exploratory analysis of the pilot data reveals conspicuous voltage deflections, which I labelled the P250 and the P750, I decided against exploring the P250 and P750 further in subsequent experiments, due to the uncertainty in how to interpret the results.

9.2.2 Experiment 2

The purpose of Experiment 2 was to replicate the results from the pilot study in order to provide stronger evidence in support of the idea that there is a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams. Specifically, I aimed to replicate the finding that an N400 is elicited in response to reading non-collocational bigrams, and I also aimed to replicate the finding that there is a difference in onset latency between collocational and non-collocational bigrams. In addition, I aimed to find out whether or not a P600 is elicited in response to reading non-collocational bigrams, as I did not find clear evidence for this component in the pilot study due to the overlapping time windows mentioned in the previous sub-section.

The results of Experiment 2 reveal that, in the N400 latency range, there is a slightly positive amplitude in the collocational condition and a slightly negative amplitude in the non-collocational condition. This replicates the N400 results of Experiment 1, and it confirms Hypothesis 1, that reading the second word of a non-collocational bigram will elicit an N400. The onset latency results also replicate those of Experiment 1, by showing that there is a difference in onset latency between collocational and non-collocational bigrams, and that the onset latency is greater in the collocational condition compared to the non-collocational condition. This therefore confirms Hypothesis 3, which states that the onset latency will be greater for the collocational condition compared to the non-collocational condition. Moreover, it provides further evidence in support of the idea that there a neurophysiological difference in the way that the brain processes collocational bigrams compared to non-collocational bigrams, regardless of the identification of specific ERP components.

However, the P600 results disconfirm Hypothesis 2, which states that reading the second word of a non-collocational bigram will elicit a P600, as the amplitude in the P600 latency range is actually greater in the collocational condition compared to the non-collocation condition. In this case, I therefore found the opposite to what I hypothesized, suggesting that a P600 is actually elicited in response to reading the second word of a *collocational* bigram rather than the second word of a non-collocational bigram. This finding will be discussed further in section 9.4.2.

9.2.3 Experiment 3

The purpose of Experiment 3 was to conduct the exact same study as Experiment 2 but, this time, using non-native speakers of English rather than native speakers of English. The aims of Experiment 3 were to find out whether or not there is a neurophysiological difference in the way that collocational and non-collocational bigrams are processed by non-native speakers of English, and to see how the ERP results for the non-native speaker group compare to those of the native speaker group in Experiment 2.

The results of Experiment 3 reveal that, in the N400 latency range, there is a positive amplitude in the collocational condition and a slightly negative amplitude in the non-collocational condition. This replicates the results of Experiments 1 and 2, and it provides evidence in support of Hypothesis 1, which states that reading the second word of a non-collocational bigram will elicit an N400. Moreover, while the difference between conditions in the N400 latency range is not significant in Experiment 2, it is highly significant in Experiment 3, and the difference between conditions in exceptionally large in comparison to the N400s present in Experiments 1 and 2 of this thesis. This therefore provides evidence in support of Hypothesis 4, which states that the ERP responses will be larger than those demonstrated by the native speakers in Experiment 2.

The results of Experiment 3 also reveal that, in line with the results of Experiment 2, there is a higher amplitude in the P600 latency range in the collocational condition compared to the non-collocation condition. This therefore confirms Hypothesis 2, which states that reading the second word of non-collocational bigram will *not* elicit a P600. However, the results of the onset latency analysis refute Hypothesis 3, which states that the onset latency will be greater for the collocational condition compared to the non-collocational condition, as the opposite is found to be true. That said, although the onset latency in Experiment 3 is greater in the non-collocational condition compared to the collocational condition, the difference is only marginal and it is not significant. The range between the onset latency of the collocational condition and the onset latency of the non-collocational condition is actually much smaller for the non-native speakers (2.272 ms) compared to the native speakers (69.21 ms). This suggests that the native speakers are actually more sensitive to the transition probabilities between words than the non-native speakers, thus providing evidence against Hypothesis 4, which states that the ERP responses will be larger than those demonstrated by the native English speakers in Experiment 2. An additional problem with determining whether it is the native or non-native speakers who are most sensitive to transition probabilities in discussed in section 9.4.3.

9.2.4 Experiment 4

The purpose of Experiment 4 was twofold: Part 1 aimed to replicate the results of Experiment 2 in order to strengthen the confidence of its conclusions, and Part 2 aimed to investigate the strength of the correlation between different measures of collocation strength and the amplitude of the ERP response. The measures of collocation strength that I investigated were transition probability, mutual information, log-likelihood, z-score, t-score, Dice coefficient, MI3, and raw frequency.

The results of Part 1 reveal that, in the N400 latency range, the amplitude is lower in the non-collocational condition compared to the collocational condition. This is in line with the results of the previous three experiments, and it provides evidence in support of Hypothesis 1, which states that reading the second word of a non-collocational bigram will elicit an N400. Similarly, the onset latency results mirror those of Experiment 2, and confirm Hypothesis 3, as the onset latency is significantly greater in the collocational condition compared to the non-collocational condition. However, the P600 results are in opposition to those found in Experiments 2 and 3, and they refute Hypothesis 2, which states that reading the second word of a non-collocational bigram will *not* elicit a P600. In this case, reading the second word of a non-collocational bigram does seem to elicit a P600, as the amplitude in the P600 latency range is higher in the non-collocational condition compared to the collocational condition. The P600 results will be discussed further in section 9.4.2.

The results of Part 2 support Hypothesis 4 as they reveal that there *is* a correlation between the transition probability of a bigram and the amplitude of the ERP response. Moreover, as well as there being a strong correlation between transition probability and amplitude, there is a strong correlation between transition probability and *every* association measure investigated in this study. The strongest correlation exists between amplitude and z-score, closely followed by amplitude and MI3, amplitude and Dice coefficient, and amplitude and t-score. Interestingly, these are all hybrid measures of collocation strength. As mentioned in section 1.2, Hoffman et al. (2009:151) explain that "t-score is a hybrid between frequency and significance, z-score is a hybrid between effect size and significance" and "both MI3 and Dice balance frequency and effect size, but do so in diametrically opposed ways". By contrast, the measures of collocation strength that are less strongly correlated with amplitude (namely raw frequency, transition probability, mutual information, and log-likelihood) are all pure association measures, which use just one statistical dimension. Log-likelihood (Dunning 1993) is a significance statistic whereas mutual information (Church and Hanks 1990) is an effect-size statistic. The results of this study thus suggest that it is the hybrid measures which can be seen as having the most psychological validity, i.e. both *strength* and *entrenchedness* matter.

9.2.5 Overall findings

The overarching aim of this thesis was to find out whether or not there is a neurophysiological difference in the way that the brain processes collocational adjective-noun bigrams compared to non-collocational adjective-noun bigrams, and thus contribute to our understanding of whether or not the phenomenon of collocation can be seen as having psychological validity. Taken together, the results of the four ERP experiments provide sufficient evidence to suggest that there *is* a neurophysiological difference in the way that the brain processes collocational adjective-noun bigrams compared to non-collocational adjective-noun bigrams. I therefore conclude that the phenomenon of collocation *can* be seen as having psychological validity.

Across the four ERP experiments conducted in this thesis, the N400 results are stronger and more consistent than the P600 and onset latency results. Therefore, while the P600 and onset latency findings would need to be replicated in future experiments in order to confirm their status as valid ERP effects, I can confidently conclude that reading the second word of a non-collocational bigram elicits an N400. I propose that this component be known as the 'collocational N400' as, although I do not claim that the N400 elicited in my experiments is qualitatively different and functionally distinct from the classical semantic N400, there is a clear quantitative difference between the semantic N400 and the collocational N400 found in this thesis.

Specifically, while the semantic N400 reaches negative amplitudes on the order of -9 µV (Kutas and Hillyard 1984:205), the collocational N400 found in this thesis has a maximum negative amplitude of -0.486 µV. This is not surprising, as a semantic violation is intuitively much more obstructive to language processing than a collocational violation. Moreover, this is in line with the results of Siyanova-Chanturia et al. (2017), who find that N400 amplitude is small (5.12 µV) in response to reading a binomial collocations (e.g. *knife and fork*), large (−0.1 µV) (in the context of their study) in response to reading a binomial non-collocation with a semantic violation (*theme and fork*), and intermediate (3.18 µV) between these values in response to reading a binomial non-collocation with no semantic violation (*spoon and fork*).

9.3 Implications of the results

9.3.1 One or two processing systems?

For the purposes of this thesis, I set up a dichotomy between collocations and non-collocations, with the collocations having a high transition probability in the BNC1994 and the non-collocations being absent from the BNC1994. However, this dichotomy is purely methodological; I am not claiming that collocations and non-collocations exist in a dichotomy in a theoretical sense. Rather, I believe that all sequences of words exist on a sliding scale of collocationality, from less-collocational to more-collocational. This belief is supported by the results of this thesis. After all, if there were truly a dichotomy, we would expect there to be a qualitative difference in the way that the brain processes collocations and non-collocations, with entirely different ERP components and different scalp distributions being present in each condition. But this is not what I find. Rather, as predicted (see Chapter 4, section 4.2.1), the results reveal a quantitative difference between conditions, with the waveforms following the same pattern in both conditions, but with varying amplitudes (particularly in the latency range for the N400).

The fact that the results reveal a quantitative difference rather than a qualitative difference has important implications for theories of language which posit the existence of separate systems for the processing of collocational and non-collocational word sequences. Typically, this is conceptualized as there being one processing system for grammar and another, separate, processing system for lexis/semantics. As mentioned in Chapter 2 (section 2.9), Sinclair (1991) and Wray (2002) both posit the existence of two processing systems, arguing that the grammatical system is secondary to the system which processes language using memorized preconstructed or prefabricated chunks (Sinclair 1991:114; Wray 2002:10). By contrast, the theories of Lexical Priming (Hoey 2005), Pattern Grammar (Hunston and Francis

2000), and Construction Grammar all attempt to account for the whole of language without positing the existence of a separate grammatical system.

Since the results of this thesis suggest that there is a quantitative rather than a qualitative difference in the way that collocations and non-collocations are processed, this thesis provides evidence against the existence of the two processing systems proposed by Sinclair and Wray. Instead, this thesis supports the theories of Lexical Priming, Pattern Grammar, and Construction Grammar, which attempt to account for the whole of language using collocational mechanisms.

In Chapter 3 (section 3.4.5), I noted that the identification in the ERP literature of the N400 and the P600 suggests that there are distinct neurophysiological processes involved in semantic and syntactic processing. This would weigh against theories of language such as Pattern Grammar and Construction Grammar, which argue that syntax and semantics are inseparable. However, I also pointed out that this is not the whole story, as more recent studies have shown that the N400 can be sensitive to morphosyntactic violations (e.g. Severens et al. 2008:141; Nieuwland et al. 2013:151), and the P600 can be sensitive to semantic violations (e.g. Geyer et al. 2006). This suggests that there is not a separate component specifically reflecting syntactic processing and another component specifically reflecting semantic processing. Rather, the N400 and P600 seem to be part of a system that somehow encompasses both. Indeed, based on these results, Kuperberg et al. (2006:527) conclude that “the neural systems supporting syntactic and semantic processing may be linked”, and Swaab et al. (2012:28) state that “the interaction between semantic and syntactic processes in the brain may be more dynamic than was previously suggested”.

However, since the results of three out of the four experiments in this thesis reveal that reading the second word of a non-collocational bigram elicits an N400 but *not* a P600, this thesis provides very little evidence to suggest that these components work together as part of a single processing system. The N400 is sensitive to collocational violations but the P600 is not, suggesting that the possibility of there being a separate grammatical processing system cannot be ruled out. That said, the results of Experiment 4 *do* reveal that reading the second word of a non-collocational bigram elicits a (non-significant) P600. As mentioned in Chapter 8 (section 8.5.2), a P600 may be elicited in the non-collocational condition in Experiment 4 but not in the same condition in other experiments because, compared to the stimuli set used in Experiments 1, 2, and 3, the stimuli set used in Experiment 4 contains more collocational bigrams with an exceptionally high transition probability. The mean transition probability of the bigrams in Experiments 1/2/3 is 0.231, while the mean transition probability of the bigrams in Experiment 4 is almost double that, at 0.452. When there is a higher transitional probability, there is a stronger expectation that word Y will follow word X, and therefore a greater increase in cognitive load when that expectation is violated. It could be the case that a P600 is only elicited when a particularly strong collocation is violated. Therefore, in this way, the P600 could still be seen as working with the N400 as part of a single processing system.

9.3.2 Network model of language processing

In Chapter 2 (section 2.8), I introduced what I refer to as *the network model of language processing*. In this model, the network consists of nodes which, depending on the theory, constitute individual words, collocations, or constructions. These nodes are connected to other nodes via weighted connections, and the weight of the connections represents the strength of the collocation. Collocational strength is conceptualized as psychological transition probability whereby, when an individual produces or hears word X immediately followed by word Y, a connection is formed between those two words (or nodes). Then, on subsequent occasions when that individual produces or hears word X, word Y will be mentally activated, along with any other words that have previously followed word X for that individual. Through repeated exposure to the same sequence X-then-Y, the connection between word X and word Y is strengthened so that there is an increased probability that word Y will occur after word X. Thus, connection strength is determined by the level of prior exposure to the collocation.

In this thesis, I used the statistical measure of (corpus) transition probability as a proxy for the psychological transition probability inherent in the network model of language processing. The results of Experiment 4 reveal that there *is* a strong correlation between the transition probability of a bigram and the amplitude of the ERP response (see Chapter 8, 8.4.2). Specifically, the higher the transition probability of the collocational bigram, the greater the expectation that word Y will follow X, and therefore the greater the increase in cognitive load (and hence the greater the amplitude) when the second word of the bigram does *not* fill the collocational expectation set up by the first word. I have therefore provided strong evidence to suggest that we *do* process language in terms of probabilities, regardless of semantic or syntactic constraints. Thus, I argue that the network model of language processing proposed in this thesis does have psychological validity.

This finding that there *is* neuroscientific evidence for the network model of language processing has important implications for the field of cognitive neuroscience. It was already clear that the network model of language processing was intuitively plausible from a biological perspective, as it is in line with what we know about how neurons work in the brain. Indeed, as mentioned in Chapter 2 (section 2.8), neurons interact and form networks that are capable of representing knowledge (Rosenweig et al. 1996:35; Sternberg 2009:35), and these networks form the basis of complex cognitive behaviour (Hebb 1949). With this biological knowledge, we can assume that linguistic information is represented in the brain as some sort of network. Meanwhile, with experimental evidence from cognitive science, we can build models of how linguistic information might be stored, and evaluate the plausibility of these models based on how well they fit with behavioural data. However, without the use of neuroimagining techniques, it is impossible to bridge the divide between biology and cognitive science. In this thesis, I have begun to bridge this divide using EEG, one of the two electromagnetic neuroimaging techniques used in cognitive neuroscience. Specifically, I have proposed a cognitive model (i.e. the network model of language processing) which is based on biological knowledge (i.e. what we know about how neurons work in the brain), and I have provided electrophysiological evidence which demonstrates that the brain *is* sensitive to transition probabilities, which serve as a point-estimate of the psychological transition probability inherent in the model.

Likewise, the finding that there is neuroscientific evidence for the network model of language processing has important implications for the field of corpus linguistics, and related usage-based approaches to the study of language, such as Construction Grammar. As mentioned in Chapter 2 (section 2.8), the network model of language processing is prevalent across the different approaches to collocation, as they often carry either an explicit or an implicit assumption that collocations are represented in the brain as transitions across a network. This idea is most explicit in Hoey’s (2005) theory of Lexical Priming, as this theory centres around the idea that any word that frequently follows another word is mentally activated whenever the first word is spoken or heard. Similarly, the network model is very explicit in Construction Grammar, as “[t]he collection of constructions ... constitute a highly structured lattice of inter-related information” (Goldberg 1995) – similar to the pattern of interconnected neurons in the brain.

While the network idea is not explicitly discussed by Wray (2002), it is there implicitly, as she argues that frequently encountered sequences are accessed more quickly than sequences that are encountered less frequently (Wray 2002:268). Similarly, although the network idea is not explicitly discussed by Sinclair (2004:161), he does argue that a “word becomes associated with a meaning through its repeated occurrence in similar contexts”. This suggests that the mind is sensitive to frequency information in the input and that the connections between nodes are strengthened through repeated exposure. Thus, since this thesis provides evidence in support of the network model of language processing, it also provides evidence in support of these approaches to collocation, which either explicitly or implicitly acknowledge the existence of a network that serves as a collocational mechanism in the processing of language.

The evidence that I have found in support of the network model of language processing has further implications for how we define and conceptualize collocation. In Chapter 2 (section 2.9), I mentioned three of the key parameters of difference between the different approaches to collocation. The first is whether the network model of language processing is addressed explicitly or implicitly, the second is whether the approach posits one or two processing systems, and the third is whether the approach assumes that collocations have clear beginning and end points, or whether there are no clear boundaries to when a collocation starts and ends. Within the network model of language processing, there are no clear beginning and end points to a collocation because, whenever any word is heard or produced, all words that have previously followed that word in the language experience of the individual are mentally activated for use (to at least some extent). Thus, by providing evidence in support of the network model of language processing, this thesis also provides evidence in support of approaches to collocation which do not require there to be clear boundaries on when a collocation starts and ends.

In both Lexical Priming (Hoey 2005) and Pattern Grammar (Hunston & Francis 2000), no artificial boundaries are imposed on when a collocation starts and ends (see Chapter 2, section 2.9). By contrast, both Wray (2002) and Sinclair (1991) do conceptualize collocations as having clear beginning and end points. This suggests that the former theories are more compatible with the results of this thesis than the latter theories. However, it is important to remember that the nodes within a network do not necessarily have to constitute individual words; rather, depending on the theory, the nodes can constitute whole collocations or constructions (see Chapter 2, section 2.8). This allows for clear boundaries between different collocations or constructions.Accounting for these boundaries is necessary in the context of idioms or other fixed multi-word expressions, which are non-compositional or opaque. Thus, in this way, different definitions and conceptualizations of collocation can be accounted for within the same model. However, more research would be needed in order to clarify how the processing of compositional collocations interacts with the processing of holistically-stored non-compositional collocations.

Another key point of discussion is that, although I used the statistical measure of transition probability as a proxy for the psychological transition probabilities inherent in the network model of language processing, the results of Experiment 4 suggest that this may not be the most appropriate statistical measure for this purpose. Indeed, the results of the correlational analysis reveal that transition probability is *not* the measure of collocation strength that can be seen as having the most psychological validity. Rather, the z-score statistic is the measure of collocation strength that correlates most strongly with the amplitude of the ERP response, closely followed by MI3, Dice coefficient, and t-score (Chapter 8, section 8.4.2).

As mentioned in Chapter 4 (section 4.2.1), I chose to use transition probability (specifically forward transition probability) as the sole measure of collocation strength in this thesis partly because it is a directional association measure. This means that it takes into account the asymmetry of bigrams, i.e. the fact that the strength of the association is not equal in both directions. For example, the bigram *of course* is asymmetric because the set of words that are likely to follow *of* is much larger than the set of words that is likely to precede *course* (Gries 2013:144). Gries (2014a:38; 2014b:291; 2013:139) points out that, although over 25% of collocations are asymmetric, the majority of association measures that are used in corpus linguistics are symmetric, meaning that they do not take into account the fact that the strength of association is not always equal in both directions. Indeed, all of the association measures tested in Experiment 4 Part 2 (aside from transition probability) are symmetrical (also known as *bidirectional*) measures of collocation strength.

Using a directional measure as a proxy for the psychological transition probabilities inherent in the network model of language processing seems appropriate as, according to this model, hearing or producing a particular word automatically activates words that are likely to *follow* that word, rather than precede it (though this linear assumption will be problematized below). Likewise, since ERP experiments involve presenting sentences in a word-by-word fashion, forward transition probability as opposed to backward transition probability seems to be more relevant in this context

Gries (2013:137; 2014b:291; 2015:95) strongly encourages the use of directional association measures in corpus linguistics, arguing that they are "psychologically/psycholinguistically more realistic". It seems intuitively plausible to take a linear approach to the study of language. After all, as mentioned in Chapter 2 (section 2.4.1.2), Brazil (1995:4) convincingly points out that "[s]peech is an activity that takes place in time: speakers necessarily say one word, follow it with another and then with another, and so on". However, while the sound stream is linear, language *production* cannot be wholly linear, due to the way in which a word can precede the word which it is modifying. In other words, it seems plausible that we must select the noun before we can select the adjective which modifies it. The relevant network transition in language production may therefore be N → Adj, rather than Adj → N, even though the adjective will come first in the speech stream. With this in mind, the focus on linearity can be seen as a limitation of the network model of language processing. Nevertheless, it could be argued that the linearity assumption works for language *comprehension*, as we do read in a linear fashion and we can read a modifier before we read the word that it is modifying.

Although my ERP studies focus on language comprehension, rather than production, I decided to check whether or not backward transition probability is more closely correlated with the amplitude of the ERP response than forward transition probability. To do this, I calculated the backward transition probability of each collocational bigram in Experiment 4, and then conducted a Pearson correlation as described in Chapter 8 (section 8.3.4). Whereas forward transition probability is calculated by dividing the frequency of a bigram X-then-Y with the frequency of X, backwards transition probability is calculated by dividing the frequency of a bigram X-then-Y with the frequency of Y.

The results of this additional analysis reveal that there is a strong negative correlation between the backward transition probability of the collocational bigrams and the difference in amplitude between the two conditions: $r = -0.658$, $p = 0.054$. This correlation can be seen on the scatterplot in figure 9.1.

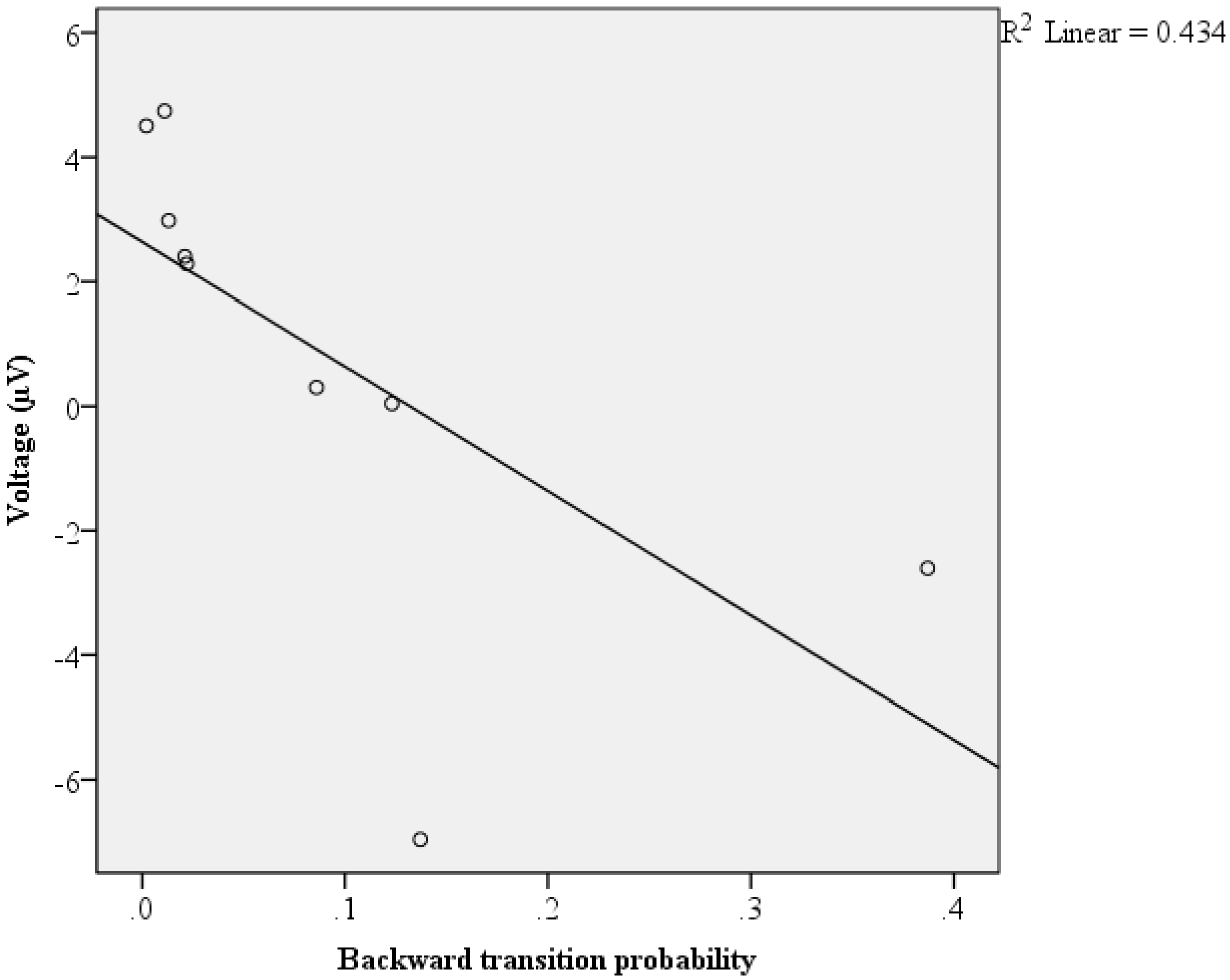


Figure 9.1: Scatterplot showing the relationship between amplitude and backward transition probability

Figure 9.1 shows that the value of R² is 0.434, indicating that backward transition probabilities account for 43.4% of the variance in amplitude. This is much higher than forward transition probability, which accounts for 38.6% of the variance in amplitude (Chapter 8, section 8.4.2). Moreover, looking at table 9.1, we can see that backward transition probability correlates more strongly with amplitude than forward transition probability.

Table 9.1: Association measures ranked by strength of correlation (including backward transition probability)

Association measure	Pearson's *r*	*p*-value
1. Z-score	-0.773	0.014*
2. MI3	-0.772	0.015*
3. Dice coefficient	-0.712	0.031*
4. T-score	-0.679	0.044*
5. Backward TP	-0.658	0.054
6. Frequency	-0.636	0.065
7. Forward TP	-0.621	0.074
8. MI	-0.575	0.105
9. Log-likelihood	-0.566	0.112

We can therefore conclude that, although my ERP studies focused on language comprehension rather than production, backward transition probability is more strongly correlated with the observed brain response than forward transition probability (at least for adjective-noun bigrams). However, the correlation is still stronger (and significant) for the bidirectional hybrid measures, suggesting that the combination of *strength* and *entrenchedness* is still key to the psychological validity of collocation. It may therefore be optimal to use a bidirectional association measure as a proxy for the psychological transition probabilities inherent in the network model of language processing. Furthermore, since we know that a word can precede the word that it is modifying, as in the case of the adjective-noun bigrams used as experimental stimuli in this thesis, we know that language production cannot be completely linear. It is therefore necessary to build into the model the propensity of a language user to construct some sort of mental hierarchy during language production.

9.3.3 Cognitive neuroscience

As mentioned in Chapter 3, the study of collocation and related notions is growing in psychology, with collocation and related notions being studied in a range of contexts from language disorders (e.g. van Lancker Sidtis & Fromkin 2017; Wray 2017) to language acquisition and learning (e.g. Theakston & Lieven 2017; Arnon & Christiansen 2017; McCauley & Christiansen 2017; Ellis & Ogden 2017). A variety of methods are used to measure different responses to (non-)collocational stimuli, including behavioural responses (self-paced reading; see Chapter 3, section 3.2), oculo-motor responses (eye-tracking; Chapter 3, section 3.3) and (in cognitive neuroscience) electrophysiological responses (EEG/ERPs). However, across the studies conducted using these different methods, it is becoming increasingly apparent that the collocations that are studied are predominantly idioms and other fixed or semi-fixed multi-word expressions. This is not surprising, as it is more practical to work with fixed rather than fluid collocations; yet it is problematic in that it limits the generalisability of the results to a small subset of language use.

By investigating the processing of collocations, the work presented in this thesis is part of an ongoing tradition in psychology; yet, by focusing on collocations that are less fixed than those that are typically studied, the work presented in this thesis pushes the boundaries of work that has come before it, and accords collocationality a more central position in language processing than has previously been assumed. While the Lau et al. (2016) study introduced in Chapter 3 (section 3.4.6.2) uses comparable stimuli to that used in the present study (i.e. adjective-noun bigrams that are extracted from a corpus based on their transition probabilities), this study is flawed in that it creates a confound between frequency and collocationality. This makes it impossible to disentangle the effects of word frequency and transition probability on the amplitude of the N400. Therefore, by avoiding this confound, this thesis can be seen as constituting a novel contribution to the fields of psychology and cognitive neuroscience.

The following table details the results of the key ERP studies of collocation and predictability that were introduced in Chapter 3 (sections 3.4.6 and 3.4.7), along with comments concerning whether or not these results are supported by the findings of this thesis.

Only those studies focusing on the N400 and P600 are included here, as these are the only components investigated in this thesis.

Table 9.2: Comparison table of results from previous studies of collocation and the N400/P600, and results from the present study

Author (s)	Date of publication	Observed effects	Observed effects supported by thesis findings?	Scalp distribution of observed effects	Scalp distributions supported by thesis findings?
Laurent et al.	2006	N400 and P600 amplitude is larger in response to reading the last word of an unconventional expression compared to the last word of a conventional idiom.	Yes	N400 - anterior P600 - left central parietal	N400 - partly, but only in Experiment 1 P600 - partly, but only in Experiment 1
Rhodes & Donaldson	2008	N400 is enlarged in response to reading words in a word pair which are not likely to occur together.	Yes	N400 - central	N400 - partly, but only in Experiments 1 and 4
Davenport & Coulson	2011	N400 and P600 amplitude is larger in response to reading sentence-final words in sentences with a low cloze probability compared to sentence-final words in sentences with a high cloze probability.	Yes - N400 results only	N400 - central posterior P600 - right anterior	N400 - yes, in Experiment 4 P600 – no
Lau et al.	2016	N400 amplitude is larger in response to reading an unpredictable adjective-noun bigram compared to predictable adjective-noun bigrams.	Yes	N400 - central posterior	N400 - yes, in Experiment 4
Siyanova-Chanturia et al.	2017	N400 amplitude is smaller is response to reading binomial collocations compared to matched non-collocations.	Yes	N400 - evident across all sites, but strongest at right hemisphere sites	N400 - maximal at right hemisphere sites in Experiments 2 and 3

It is clear from table 9.2 that the N400 results from this thesis support all five of the cited studies which investigate the N400 response to reading unpredictable stimuli. Across these previous five studies as well as the four experiments presented in this thesis, the N400 is larger in response to reading unpredictable words compared to predictable words. This is therefore a very robust finding. However, there is great variation in scalp distribution results, both across previous studies and across the four experiments presented in this thesis. I will now discuss inconsistencies in scalp distribution findings as a separate issue.

9.4 Issues raised by the results

9.4.1 Inconsistencies in scalp distribution findings

One issue raised by the results is that the scalp distributions of the ERP effects are not consistent across experiments. In Experiment 1, the N400 has an anterior-central scalp distribution with no effect of laterality. By contrast, there is no anteriority effect in Experiment 2, but there is a clear effect of laterality, with the N400 being maximal over right hemisphere and midline electrode sites. In Experiment 3, there is no significant interaction between condition and left-to-right electrode position, but the N400 effect is found to be significant at all levels of anteriority. Then, in Experiment 4, I again find that there is no laterality effect, but the N400 is maximal at central-posterior electrode sites.

Each ERP component is typically associated with a particular scalp distribution (Kutas and Dale 1997:205). Therefore, since the scalp distribution of the N400 is inconsistent across my experiments, this calls into question whether the effect in the N400 latency range reflects the same component in each experiment, and whether the effect is comparable with the traditional semantic N400 found in other studies (e.g. Kutas & Hillyard 1980). It could be that there is a functionally distinct ERP component elicited in each experiment, in response to small methodological differences, and that these components are qualitatively different from the traditional semantic N400.

However, as mentioned previously, Kutas and Dale (1997:222) point out that “N400s do differ in latency and scalp distribution, even within presumably similar experimental tasks”. Furthermore, there is “potential for extensive individual and group differences in ERP scalp topography”, especially for endogenous components such as the N400, so “it may not always be the case that a component of interest will be maximal over the expected scalp sites” (Handy 2005:36). This is apparent from the scalp distribution findings of previous N400 studies of collocation (see table 9.2, section 9.3.3). Therefore, I conclude that my results do provide sufficient evidence to show that an N400 is elicited in response to reading the second word of non-collocational bigrams, and that this N400 is *not* qualitatively different from the traditional semantic N400.

It is interesting to see that the scalp distribution of the N400 is central-posterior in Experiment 4, and therefore matches the scalp distribution of the traditional semantic N400 (Swaab et al. 2012:399). In Chapter 8 (section 8.5.2), I questioned why the N400 in Experiment 4 has the expected scalp distribution, whereas the N400s in Experiments 1, 2, and 3 do not. I argued that this can be explained by the properties of the stimuli sets. Compared to the stimuli set used in Experiments 1, 2, and 3, the stimuli set used in Experiment 4 contains more collocational bigrams with an exceptionally high transition probability. Thus, since the bigrams with a higher transition probability set up stronger expectations, and are therefore associated with a greater increase in cognitive load when those expectations are violated, it could be the case that the N400 is only elicited ‘fully’ (i.e. in a way that resembles the classical component) when the cognitive load is exceptionally high. This explanation is also in-fitting with the results of the Lau et al. (2016) study because, as with the results of Experiment 4, the results of the Lau et al. (2016) study reveal that a central-posterior N400 is elicited in response to reading collocational bigrams with an exceptionally high transition probability.

The same could be said for the onset latency results. The onset latency effect is maximal at anterior and central scalp sites in Experiment 1, so the analysis was conducted using only these sites in subsequent experiments. Yet, while there are no significant interactions in Experiments 2 and 3, the effect is maximal at left hemisphere electrode sites in Experiment 4, allowing me to conclude that the onset latency effect occurs at left anterior-central electrode sites. Perhaps no significant interactions were found in the previous experiments because the transition probabilities were not high enough to create very high expectations, and therefore a large increase in cognitive load as a result of breaking those very high expectations. This possibility would need to be explored in future experiments, by comparing the ERP responses to reading violations of collocations with very high association scores to that of violations of collocations with comparatively lower association scores.

9.4.2 Elicitation of a putative P600 in the collocational condition

Another issue raised by the results is that, in Experiments 2 and 3, a putative P600 is elicited in the *collocational* condition, as opposed to the non-collocational condition as expected. In one way this is not problematic, as it still shows that there is a difference between conditions, regardless of the direction of the effect or how well it fits the hypotheses associated with each individual experiment. It therefore provides further evidence in support of the overarching hypothesis that there is a neurophysiological difference in the way that collocational and non-collocational bigrams are processed. However, it *is* problematic in the sense that there is no clear explanation as to *why* a P600 would be elicited in response to reading the second word of a collocational bigram. The collocational condition is the control condition representing conventionalized language use. Moreover, while it is widely known that a small N400 is elicited in response to reading "all meaningful words" (Kumar & Debruille 2004:89), this is not known to be the case for the P600. The P600 result found in Experiments 2 and 3 would need to be replicated in future experiments in order to be considered a valid ERP effect.

9.4.3 Conflicting conclusions from N400 and onset latency results

A final issue raised by the results is that the N400 and onset latency results in Experiment 3 suggest conflicting conclusions, making it unclear whether it is the native or non-native speakers who are the most sensitive to the transition probabilities between words. The N400 results suggest that it is the non-native speakers who are the most sensitive to transition probabilities, as the N400 effect is larger for the non-native speaker group compared to the native speaker group (Chapter 7, section 7.4.1.1). However, the onset latency results suggest that it is the native speakers who are the most sensitive to transition probabilities, as the onset latency is much larger for the native speaker group compared to the non-native speaker group (section 7.4.2).

I am inclined to assign more weight to the N400 results for two distinct reasons. First, the idea that non-native speakers are more sensitive to the transition probabilities between words than native speakers is supported by the results of Hughes and Hardie (forthcoming), who conduct a self-paced reading experiment with native and non-native speakers of English (see Chapter 1, section 1.5). Second, the N400 results can be seen as being more accurate than the onset latency results. I will explain the reason for this in the following section.

9.5 Limitations of the study and directions for future research

In Chapter 3 (section 3.4.8), I discussed the limitations of the ERP technique. For instance, the ERP technique has a low spatial resolution (Hillyard 2000:25; Swaab et al. 2012:22), meaning that I am unable to draw conclusions about *where* in the brain the electrical activity is taking place. To do this, I would need to conduct an fMRI or PET experiment (see Chapter 1, section 1.3), as these haemodynamic neuroimaging techniques have a high spatial resolution (Hillyard 2000:25; Swaab et al. 2012:22).

In addition to the general limitations associated with all ERP studies, there are a number of limitations that are specific to the ERP experiments conducted in this thesis. One limitation is that, since the experiments in this thesis involved reading as opposed to listening, I cannot assume that the results are generalizable to auditory processing. As far as I am aware, at present there are no studies (either electrophysiological or behavioural) which directly address the auditory processing of collocations (Conklin & Schmitt 2008:84). This would therefore be an interesting avenue for future research (Millar 2010:276).

Another limitation of this thesis is that the experimental stimuli consist solely of adjective-noun bigrams. It would be interesting to see if the same results are found when using bigrams from different word classes, such as noun-verb bigrams or adverb-verb bigrams. It would also be interesting to see if reading the second word of non-collocational bigrams (of any word class) produces *spillover effects*. A spillover effect is defined by Millar (2010:162) as “the prolonged processing burden caused by collocation errors”. In the context of the present study, looking for spillover effects would involve investigating whether or not the N400 effect is apparent in the words following the non-collocational bigrams.

Another problem with the stimuli is that, when selecting the bigrams, I relied on my intuition as a native speaker of English to determine whether or not a potential bigram that is absent from the BNC is a non-collocation as opposed to just being “accidentally absent” (Stefanowitsch 2006:62). In retrospect, a more reliable approach would have been to ask a group of people to rate the bigrams according to their perceived collocationality.

An additional limitation of the experiments in this thesis is the small sample size. I used a sample size of 16 participants per experiment, as this is on the higher end of the sample size recommended by Luck (2014a:262). However, it would have been beneficial to have double this number of participants as this would increase the statistical power. Moreover, having 32 participants would have kept the experiments in this thesis more in line with previous studies of collocation and predictability, most of which have at least 30 participants. Indeed, looking back at the 9 studies of collocation and predictability listed in table 3.4 (Chapter 3, section

3.4.7), only 2 of these studies have less than 30 participants (Tremblay and Baayen (2010) have 10 while Davenport and Coulson (2011) have 18). I will therefore consider using a larger sample size for future ERP experiments.

A potential criticism of this work is that I do not fully account for whether or not the adjective and noun in each collocational bigram are more semantically related than the adjective and noun in the matched non-collocational bigram. This makes it difficult to discern whether or not the observed neurophysiological difference between conditions can be attributed to differences in collocational strength, as opposed to differences in semantic relatedness. In Experiment 4, I attempted to account for this by only using bigram pairs where the HSO score is *not* higher for the collocational bigram compared to the non-collocational bigram (see Chapter 8, section 8.3.2). However, a major problem with the HSO metric is that it conflates semantic relatedness with collocation strength by measuring both paradigmatic and syntagmatic relations. For instance, Hirst and St-Onge (1998:306) give the example sentence *The evening prior to admission, take a <u>shower</u> or <u>bath</u>*, and argue that the underlined words are semantically related. Yet, although the words *shower* and *bath* exist in a paradigmatic relationship, as they can be substituted for one another (see Chapter 2, section 2.4.2), they also exist in a syntagmatic relationship as they co-occur in close proximity. The HSO metric therefore relies at least partly on collocational patterning.

The same can be said for the two other semantic relatedness metrics offered in WordNet::Similarity 2.07, namely Lesk, and vector pairs (Gomaa and Fahmy 2013:16). This is inferred from the descriptions of the metrics on the WordNet::Similarity 2.07 webpage (http://maraca.d.umn.edu/similarity/measures.html), where syntagmatic relations are described as being captured through measuring “two consecutive words” in the Lesk metric and “second-order co-occurrence(s)” in the vector pairs metric. Therefore, future studies should control for semantic relatedness using a more robust measure. This could be achieved experimentally by investigating whether or not a collocational N400 is still elicited even when the non-collocational bigrams are preceded by a semantically related prime (e.g. *clinical devices* could be preceded by *machine*). This in itself could be problematic, however, given that the collocational N400 is smaller than the semantic N400 (see section 9.2).

Another potential criticism of this thesis is that I operationalized collocations as bigrams, but bigrams are just one of the many types of collocation. As noted by Mason (2000:269), "[o]ne of the most commonly used window sizes is one word, which leads to the rather narrow definition of collocation as a fixed two word expression". Previous ERP studies have investigated the processing of other forms of collocation such as idioms and fixed multi-word expressions (see Chapter 3, section 3.4.6), but no ERP studies have investigated the processing of longer collocations that are less fixed or non-adjacent. Non-adjacent, or discontinuous, collocation pairs are the object of a *lot* of corpus analyses. I would therefore like to develop an ERP paradigm that would allow me to investigate the processing of non-adjacent collocations, though this would require a major overhaul of the method used in this thesis.

One possible avenue for exploring the processing of longer (though still adjacent) collocations would be to compare the processing of compositional and non-compositional multi-word expressions. This would involve identifying multi-word expressions in a corpus using, for example, the technique for multi-word expression extraction proposed by Wahl and Gries (in press), which is "based on the successive combination of bigrams to form word sequences of various lengths". It would then involve taking into the account the transition probability (or indeed the z-score, since this was found to be more psychologically valid in Experiment 4) of each bigram *within* the multi-word expressions (i.e. in a word sequence A-B-C-D-E, I would need to know the strength of the association for the bigram pairs A-B, B-C, C-D, and D-E). Then, for the non-compositional multi-word expressions, I could explore whether there comes a point in the processing of the expression when the ERP response does *not* correlate with the transition probability (or z-score, or other association measure), and instead seems to be elicited by the semantics of the expression. This could indicate the time point at which the processing of the non-compositional multi-word expression switches to the retrieval of the expression as a holistic unit, rather than as a sequence of individual words, and it could also provide an insight into how the semantic N400 and the collocational N400 interact. However, investigating longer collocations would bring with it additional practical problems, such as the issue of finding compositional and non-compositional multi-word expressions whose component words are matched for frequency and length.

Another direction for future research would be to explore how the network model of language processing interacts with networks from other domains of cognitive processing. After all, Hoey (2005:163), positions words as being part of a network that consists not only of linguistic information but also information from other cognitive domains, and we know from cognitive neuroscience that language processing is linked to wider cognition (see Chapter 2, section 2.8). This could be explored using the ERP technique by investigating the interaction of ERP components from diverse cognitive domains.

Finally, in retrospect, I realise that the accuracy of the onset latency results could have been improved by filtering the data using a low-pass cutoff of 10 Hz rather than 30 Hz, as this is in line with Luck's (2012:246) recommendation for filtering highly noise-sensitive measures. While it would have been ideal to re-do my onset latency analyses using this alternative filter, this was out of the scope of this thesis, as it would have required extensive re-processing of the data as well as re-analysis. This would therefore be an interesting direction for future research.

9.6 Conclusions

In this thesis, I have demonstrated that there is a neurophysiological difference in the way that the brain processes corpus-derived collocational bigrams compared to matched non-collocational bigrams, and I have thereby provide evidence to suggest that the phenomenon of collocation *can* be seen as having psychological validity. This constitutes a novel contribution to the fields of corpus linguistics and cognitive neuroscience because, while a very small selection of previous ERP studies have investigated the processing of idioms or other fixed multi-word expressions (e.g. Molinaro & Carreiras 2010:179-180), up until now, the only ERP study which has investigated the processing of collocations which are *not* fixed inadvertently confounds collocationality with frequency (Lau et al. 2016). This makes it impossible to isolate the effect of interest. Moreover, there are no previous ERP studies which investigate the processing of non-fixed collocations in non-native speakers, and there are no previous studies which attempt to correlate the neural markers of collocation with the different association measures. Each experiment presented in this thesis therefore constitutes unique and original research

An important finding of this thesis is the discovery of the ‘collocational N400’. This ERP component is elicited in response to reading the second word of non-collocational bigrams. The collocational N400 can be said to reflect the increase in cognitive load associated with reading a collocational violation. This increase in cognitive load is greater for non-native speakers compared to native speakers, as non-native speakers have less flexibility than native speakers in their use of (non-)collocational patterns. Moreover, while there is a strong correlation between the amplitude of the collocational N400 and *all* of the measures of collocation strength that I investigated in Experiment 4, the strongest correlations exist between amplitude and the hybrid association measures, including z-score, MI3, and Dice co-efficient. This suggests that mutual information and log-likelihood, which are two of the most commonly used association measures in corpus linguistics (Gries 2014a:37), are not necessarily always the optimal choice.

There is no evidence to suggest that this collocational N400 is qualitatively different and functionally distinct from the traditional semantic N400, especially since the collocational N400 elicited in Experiment 4 shares the same scalp distribution as the semantic N400 (Swaab et al. 2012:399). However, there is a clear quantitative difference between the collocational N400 and the semantic N400, as research on the semantic N400 has shown that the negative amplitudes reach around -9 µV (Kutas & Hillyard 1984:205), whereas the present study shows that the amplitude of the collocational N400 has a maximum negative amplitude of -0.486 µV. This is not surprising, as a semantic violation is intuitively much more obstructive to language processing than a collocational violation. Thus, since the N400 is elicited in response to collocational violations as well as semantic violations, we can infer that the N400 is associated more broadly with the violation of expectations. This has important implications for the field of cognitive neuroscience, as it provides further evidence to show that the usefulness of the N400 is not necessarily restricted to the linguistic domain.

In light of the issues raised by the results (section 9.4) and the limitations of the experiments conducted in this thesis (section 9.5), I have outlined several suggestions for potential future research. It would be interesting to investigate the auditory processing of collocations (Millar 2010:276), as this has not yet been explored using either behavioural or electrophysiological techniques (Conklin & Schmitt 2008:84). It would also be interesting to see if the results of this thesis are replicated using bigrams composed of words from different grammatical categories, or indeed collocations that extend to three or more words. In addition, it would be interesting to explore how the network model of language processing interacts with networks from other domains of cognitive processing. Finally, an important avenue for future research would involve investigating *where* in the brain collocational processing takes place. This could be achieved using haemodynamic neuroimaging techniques such as PET or fMRI.

REFERENCES

Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies*. London: Routledge.

Aguirre, G. K. (2014). Functional neuroimaging: Technical, logical, and social perspectives. Interpreting neuroimages: An introduction to the technology and its limits, a special report. *Hastings Centre Report*, *45*(2):8-18.

Alario, F. X, Segui, J., and Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology*, *53a*(3):741-764.

Ambridge, B., Theakston, E., Lieven, V. M., and Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development*, *21*(2):174-193.

American Clinical Neurophysiology Society (2006). Guideline 5: Guidelines for Standard Electrode Position Nomenclature. *American Journal of Electroneurodiagnostic Technology*, *46*(3):222-225

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th edn.). New York: Worth Publishers.

Andreassi, J. L. (2007). *Psychophysiology: Human behaviour and physiological response* (5th edn.). London: Lawrence Erlbaum Associates.

Andrewes, D. (2001). *Neuropsychology: From theory to practice*. New York: Psychology Press.

Arnon, I. and Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science*, *9*(3):621-636.

Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., and Zeschel, A. (2010). Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora*, *5*(1):1-27.

Ashcraft, M. H. and Radvansky, G. A. (2010). *Cognition* (5th edn.). London: Pearson.

Aston, G. and Burnard, L. (1998). *The BNC handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Atkinson, R. L., Atkinson, R. C., Smith, E. E., Bem, D. J., and Nolen-Hoeksema, S. (1996). *Hilgard's introduction to psychology*. London: Harcourt Brace.

Baker, P., Hardie, A. and McEnery, T. (2006). *Glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Baldwin, T. and Kim, S. N. (2010). Multiword expressions. In N. Indurkhya and F. R. Damerau (Eds.). *Handbook of natural language processing* (2nd edn.). Boca Raton, FL: CRC Press.

Banich, M. T. (1997). *Neuropsychology: The neural bases of mental function*. New York: Houghton Mifflin Company.

Beaumont, J. G. (2008). *Introduction to neuropsychology* (2nd edn.). London: The Guilford Press.

Berry, R. B. and Wagner, M. H. (2015). *Sleep medicine pearls* (3rd edn.). Philadelphia, PA: Elsevier Saunders.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3):275-311.

Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*. The Hague: Mouton.

Bloom, P. A. and Fisher, I. (1980). Completion norms for 329 sentence contexts. *Memory and Cognition*, *8*(6):631-642.

Brandeis, D. and Lehmann, D. (1986). Event-related potentials of the brain and cognitive processes: Approaches and applications. *Neuropsychologia*, *24*(1):151-168.

Brazil, D. (1995). *The grammar of speech*. Oxford: Oxford University Press.

Bridger, E. K., Bader, R., Kriukova, O., Unger, K., and Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *NeuroImage*, *63*(3):1334-1342.

Brown, C. and Hagoort, P. (2000). On the electrophysiology of language comprehension: Implications for the human language system. In M. W. Crocker, M. Pickering, and C. Clifton (Eds.). *Architectures and mechanisms for language processing*. Cambridge: Cambridge University Press, pp. 213-237.

Brunia, C. H. M., van Boxtel, G. J. M., and Böcker, K. B. E. (2012). Negative slow waves as indices of anticipation: the Bereitschaftspotential, the contingent negative variation, and the stimulus-preceding negativity. In E. S. Kappenman and S. J. Luck (Eds.). *The Oxford handbook of event-related potentials*. Oxford: Oxford University Press, pp. 189-207.

Budanitsky, A. and Hirst, G. (2001). Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. In *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources*. Pittsburgh, PA, pp. 1-6.

Butler, D. L., Mattingley, J. B., Cunnington, R. and Suddendorf, T. (2013). Different neural processes accompany self-recognition in photographs across the lifespan: an ERP study using dizygotic twins. *PloS one*, *8*(9):1-9.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Caharel, S., Poirou, S., Bernard, C., Thibaut, F., Lalonde, R., and Rebai, M. (2002). ERPs associated with familiarity and degree of familiarity during face recognition. *International Journal of Neuroscience*, *112*(12):1499–1512.

Chaudhari, F., Damani, O. P. and Laxman, S. (2010). Lexical co-occurrence, statistical significance, and word association. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, pp. 1058-1068.

Chen, J. J., Wieckowska, M., Meyer, E., and Pike, G. B. (2008). Cerebral blood flow measurement using fMRI and PET: A cross-validation study. *International Journal of Biomedical Imagining*, article 516359:1-12.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chung, K. K. H., Tong, X. H., Liu, P. D., McBridge-Chang, C., and Meng, X. (2010). The processing of morphological structure information in Chinese coordinative compounds: An event-related potential study. *Brain Research*, *1352*:157-166.

Church, K. W. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1):22-29.

Church, K., Gale, W., Hanks, P., and Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Ed.). *Lexical acquisition: Using on-line resources to build a lexicon*. Hillsdale, NG: Lawrence Erlbaum, pp. 115–164.

Christiansen, M. H. and Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*(3):542-551.

Churchland, P. S. and Sejnowski, T. J. (2000). Perspectives on cognitive neuroscience. In M. S. Gazzaniga (Ed.). *Cognitive neuroscience: A reader*. Oxford: Blackwell, pp. 14-24.

Conklin, K. and Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1):72-89.

Conklin, K. and Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*:45-61.

Coulson, S., King, J. W. and Kutas, M. (1998a). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes*, *13*(6):653-672.

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

Cowie, A. P. (1978). The place of illustrative material and collocations in the design of a learner's dictionary. In P. Strevens (Ed.). *In Honour of A.S. Hornby*. Oxford: Oxford University Press, pp. 127–139.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, W. (2007). Construction grammar. In D. Geeraerts and H. Cuyckens (Eds.). *The Oxford handbook of cognitive linguistics*. Oxford: Oxford University Press, pp. 463-508.

Croft, W. and Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Crowell, B. (2003). *Electricity and magnetism*. Fullerton: Light and Matter.

Darby, D. and Walsh, K. (2005). *Neuropsychology: A clinical approach* (5th edn.). London: Elsevier.

Davies, Mark. (2008-) The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. Available online at https://corpus.byu.edu/coca/.

Davenport, T. and Coulson, S. (2011). Predictability and novelty in literal language comprehension: An ERP study. *Brain Research*, *1418*:70-82.

de Saussure, F. (1966). *Course in general linguistics*. New York: McGraw-Hill.

Delorme, A. and Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134*(1):9-21.

Dennis, S. F. (1965). The construction of a thesaurus automatically from a sample of text. In M. E. Stevens, V. E. Giuliano, and L. B. Heilprin (Eds.). *Proceedings of the Symposium* *on Statistical Association Methods for Mechanized Documentation*. Washingston, DC, pp. 61-148.

Dien, J. and Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In T. C. Handy (Ed.). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press, pp. 57-82.

Divjak, D. S. and Gries, S. Th. (2008). Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon*, *3*(2):188-213.

Donald, M. W. (1983). Neural selectivity in auditory attention: Sketch of a theory. In W. Ritter and A. W. K. Gaillard (Eds.). *Tutorials in event-related potential research: Endogenous components*. Amsterdam: North-Holland Publishing Company, pp. 37-78.

Donchin, E. and Heffley, E. F. III. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.). *Multidisciplinary perspectives in event-related brain potential research*. Washington, DC: US Government Printing Office, pp. 555-572.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context-updating? *Behavioural and Brain Sciences*, *11*(3):355–372.

Duffin, W. J. (1968). *Advanced electricity and magnetism for undergraduates*. London: McGraw Hill.

Duffy, F. H., Iyer, V. G. and Surwillo, W. W. (1989). *Clinical electroencephalography and topographic brain mapping: Technology and practice*. London: Springer-Verlag.

Duncan-Johnson, C. and Donchin, E. (1979). The time constant in P300 recording. *Psychophysiology*, *16*(1):53-55.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1):61–74.

Ehrlich, S. F. and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behaviour*, *20*(6):641-655.

Elias, L. J. and Saucier, D. M. (2006). *Neuropsychology: Clinical and experimental foundations*. Boston: Pearson.

Ellis, A. W. and Young, A. W. (2004). *Human cognitive neuropsychology: A textbook with readings*. New York: Psychology Press.

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning*, *48*(4):631-664.

Ellis, N. C. and Ogden, D. C. (2017). Thinking about multiword constructions: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science*, *9*(3):604-620.

Engbert, R. and Kliegl, R. (2004). Microsaccades keep the eyes' balance during fixation. *Psychological Science*, *15*(6):431-6.

Evert, S. (2005). The statistics of word cooccurrences: Word pairs and collocations (published doctoral thesis). University of Stuttgart.

Evert, S. (2008). Corpora and collocations. In A. Lüdeling and M. Kytö (Eds.). *Corpus linguistics: An international handbook*. Berlin: Mouton de Gruyter, pp. 1212-1248.

Evert, S. and Kermes, H. (2003). *Experiments on candidate data for collocation extraction*. Paper presented at the 10th Conference of The European Chapter of the Association for Computational Linguistics.

Fedorenko, E. and Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, *18*(3):120-126.

Fillmore, C. J., Kay, P. and O'Connor, M. C. (2003). Regularity and idiomaticity in grammatical constructions: The case of let alone. In M. Tomasello (Ed.). *The new* *psychology of language: Cognitive and functional approaches to language structure* (Vol. 2). London: Lawrence Erlbaum, pp. 243-270.

Firth, J. R. (1957). A synopsis of linguistic theory. In *Special volume of the Philological Society: Studies in linguistic analysis*. Oxford: Blackwell, pp. 1-32.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, F. Francis, and E. Tognini-Bonelli (Eds.). *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins, pp. 137-156.

Friederici, A. D. and Männel, C. (2013). Neural correlates of the development of speech perception and comprehension. In K. N. Ochsner and S. Kosslyn (Eds.). *The Oxford handbook of cognitive neuroscience, volume 1: Core topics*. Oxford: Oxford University Press, pp. 171-192.

Frisson, S., Rayner, K., and Pickering, M. J. (2005). Effects of contextual predictability and transition probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5):862-877.

Gazzaniga, M. S. (2000). Life with George: The birth of the Cognitive Neuroscience Institute. In M. S. Gazzaniga (Ed.). *Cognitive neuroscience: A reader*. Oxford: Blackwell, pp. 4-13.

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2014). *Cognitive neuroscience: The biology of the mind*. London: Norton.

Geyer, A., Holcomb, P., Kuperberg, G., and Pearlmutter, N. (2006). Plausibility and sentence comprehension: An ERP Study. *Cognitive Neuroscience Supplement*, Abstract 19.

Gilquin, G. and Gries, S. Th. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory, 5*(1):1-26.

Goldberg, A. E. (1995). *A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldstein, E. B. (2011). *Cognitive psychology: Connecting mind, research and everyday experience* (3rd edn.). Belmont, CA: Wadsworth.

Gomaa, W. H. and Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, *68*(13):13-18.

Grainger, J., Lopez, D., Eddy, M., Dufau, S. and Holcomb, P. J. (2012). How word frequency modulates masked repetition priming: An ERP investigation. *Psychophysiology, 49*(5): 604-616.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.). *Phraseology: Theory, analysis, and applications*. Oxford: Clarendon Press, pp. 145-160.

Gray, H. (1918). *Anatomy of the human body* (20th edn.). New York: Lea and Febiger.

Greenbaum, D. (1974). Some verb-intensifier collocations in American and British English. *American Speech*, *49*(1/2): 79-89.

Grefenstette, G. (1992). Use of syntactic context to produce term association lists for test retrieval. *Proceedings of the 15th Annual International Conference on Research and Development in Information Retrieval*. New York: ACM, pp. 89-97.

Gries, S. Th. (2008a). Dispersion and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, *13*(4):403-437.

Gries, S. Th. (2008b). Corpus-based methods in analyses of SLA data. In P. Robinson and N. C. Ellis (Eds.). *Handbook of cognitive linguistics and second language acquisition*. New York: Routledge, Taylor & Francis Group, pp. 406-431.

Gries, S. Th. (2013). 50-something years of work on collocations: What is or should be next. *International Journal of Corpus Linguistics*, *18*(1):137-165.

Gries, S. Th. (2014a). Quantitative corpus approaches to linguistic analysis: seven or eight levels of resolution and the lessons they teach us. In I. Taavitsainen, M. Kytö, C.

Claridge, and J. Smith (Eds.). *Developments in English: expanding electronic evidence*. Cambridge: Cambridge University Press, pp. 29-47.

Gries, S. Th. (2014b). Corpus and quantitative methods. In J. Taylor and J. Littlemore (Eds.). *The Bloomsbury companion to cognitive linguistics*. London: Bloomsbury, pp. 279-300.

Gries, S. Th. (2015). Some current quantitative problems in corpus linguistics and a sketch of some solutions. *Language and Linguistics*, *16*(1):93-117.

Gries, S. Th. (2015). Quantitative designs and statistical techniques. In D. Biber and R. Reppen (Eds.). *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press, pp. 50-71.

Gries, S. Th. (2017). *Ten lectures on quantitative approaches in cognitive linguistics: Corpus-linguistic, experimental, and statistical applications*. Leiden & Boston: Brill.

Gries, S. Th. and Mukherjee, J. (2010). Lexical gravity across varieties of English: an ICE-based study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics*, *15*(4):520-548.

Gries, S. Th. and Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* *3*(1):182-200.

Gries, S. Th., Hampe, B., and Schönefeld, D. (2005). Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics*, *16*(4):635-676.

Gries, S. Th., Hampe, B., and Schönefeld, D. (2010). Converging evidence II: more on the association of verbs and constructions. In S. Rice and J. Newman (Eds.). *Empirical and experimental methods in cognitive/functional research*. Stanford, CA: CSLI, pp. 59-72.

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*(12):1711-1725.

Haas, S. M., Frei, M. G., Osorio, I., Pasik-Duncan, B., and Radel, J. (2003). EEG ocular artifact removal through ARMEX model system identification using extended least squares. *International Press, 3*(1):19-40.

Halliday, M. (1961). Categories of the theory of grammar. *Word, 17*(2): 241-292.

Hampton Wray, A. and Weber-Fox, C. (2013). Specific aspects of cognitive and language proficiency account for variability in neural indices of semantic and syntactic processing in children. *Developmental Cognitive Neuroscience*, *5*:149-171.

Handy, T. C. (2005). Basic principles of ERP quantification. In T. C. Handy (Ed.). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press, pp. 33-55.

Harley, T. A. (2008). *The psychology of language: From data to theory* (3rd edn.). New York: Psychology Press.

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2):301-307.

Hawtin, A. (forthcoming). The Written British National Corpus 2014: Design, compilation and analysis (Unpublished doctoral thesis). Lancaster University, Lancaster.

Hebb, D. O. (1949). *The organization of behaviour*. New York: Wiley.

Hillyard, S. (2000). Electrical and magnetic brain recordings: Contributions to cognitive neuroscience. In M. S. Gazzaniga (Ed.). *Cognitive neuroscience: A reader*. Oxford: Blackwell, pp. 25-37.

Hillyard, S. and Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Science USA*, *95*:781-787.

Hirst, G. and St-Onge, D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. In C. Fellbaum (Ed.). *WordNet: An electronic lexical database*. Cambridge, MA: The MIT Press, pp. 305-332.

Hoehl, S., Reid, V., Mooney, J., and Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science, 11*(1):10-16.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.

Hoffman, S. and Evert, S. (2006). BNCweb (CQP-Edition): The marriage of two corpus tools. In S. Braun, J. Mukherjee, and K. Kohn (Eds.). *Corpus technology and language pedagogy: New resources, new tools, new methods*. Frankfurt: Peter Lang, pp. 177-195.

Hoffman, S., Evert, S., Smith, N., Lee, D., and Berglund Prytz, Y. (2008). *Corpus Linguistics with BNCweb: A practical guide*. Frankfurt: Peter Lang.

Holmqvist, K. Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Howell, D. C. (2014). *Fundamental statistics for the behavioural sciences*. Belmont, CA: Wadsworth.

Huang, P., Y. Wible, D., and Ko, H. W. (2012). Frequency effects and transition probabilities in L1 and L2 speakers' processing of multiword expressions. In S. Th. Gries and D. Divjak (Eds.). *Frequency effects in language learning and processing*. Berlin: De Gruyter Mouton, pp. 145-175.

Hughes, J. and Hardie, A. (forthcoming). The effect of collocational strength on the speed of processing of adjective-noun bigrams in native speakers and learners of English: Evidence from a self-paced reading experiment.

Hunston, S. and Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Hyland, K. (2015). Corpora and written academic English. In D. Biber and R. Reppen (Eds.). *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press, pp. 292-308.

IBM Corp. (2013). IBM SPSS statistics for Windows (version 22). Armonk, NY: IBM Corp.

IELTS Partners (2018). IELTS test takers information. https://www.ielts.org/about-the-test/how-ielts-is-scored, last accessed 13/07/2018.

Ingram, J. C. L. (2007). *Neurolinguistics: An introduction to spoken language processing and its disorders*. Cambridge: Cambridge University Press.

Inkscape 0.91 [Computer software] (2015). Retrieved from https://inkscape.org/en/release/0.91/

Johnson, M. H. (1997). *Developmental cognitive neuroscience: An introduction*. Cambridge, MA: Blackwell.

Johnson, R. (1995). Event-related potential insights into the neurobiology of memory systems. In F. Boller and J. Grafman (Eds.). *The handbook of neuropsychology, volume 10*. Amsterdam: Elsevier Science Publishers, pp. 135-164.

Jones, L. L. and Estes, Z. (2012). Lexical priming: Associative, semantic, and thematic influences on word recognition. In J. S. Adelman (Ed.). *Visual word recognition* (Vol. 2). Hove: Psychology Press, pp. 44-73.

Jurafsky, D. and Martin, J. H. (2017). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition (3rd edn. draft). Retrieved from https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4):329-354.

Kaiser, E. (2014). Experimental paradigms in psycholinguistics. In R. J. Podesva and D. Sharma (Eds.). *Research methods in linguistics*. Cambridge: Cambridge University Press, pp. 135-168.

Kappenman, E. S. and Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology, 47*(5):888-904.

Kappenman, E. S. and Luck, S. J. (2011). ERP components: The ups and downs of brainwave recordings. In E. S. Kappenman and S. J. Luck (Eds.). *The Oxford handbook of event-related potentials*. New York: Oxford University Press, pp. 3-30.

Keil, A., Debener, S., Gratton, G., Junghofer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., and Yee, C. M (2013). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, *51*:1-21.

Kennison, S. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin and Review, 8*(1):132-138

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., and Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *The Journal of Neuroscience*, *28*(47):12224-12230.

Kielar, A., Meltzer-Asscher, A., and Thompson, C. K. (2012). Electrophysiological responses to argument structure violations in healthy adults and individuals with agrammatic aphasia. *Neuropsychologia*, *50*(14):3320-3337.

Kiessel, A., Miller, J., Jolicœur, P., and Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*(2):250-274.

Kiloh, L. G., McComas, A. J., Osselton, A. R., and Upton, A. R. M. (1981). *Clinical electroencephalography* (4th edn.). London: Butterworths.

Kingsley, R. E. (2000). *Concise text of neuroscience* (2nd edn.). London: Lippincott Williams and Wilkins.

Kirk, J. (2009). Word frequency use or misuse? In D. Archer (Ed.). *What's in a word list? Investigating word frequency and keyword extraction*. Surrey: Ashgate Publishing.

Kolb, B. and Whishaw, I. Q. (1996). *Fundamentals of human neuropsychology* (4th edn.). New York: W. H. Freeman and Company.

Kramer, A. F., Wickens, C. D., and Donchin, E. (1985). Processing of stimulus properties: Evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(4):393-408.

Kumar, N. and Debruille, J. B. (2004). Semantics and N400: Insights for schizophrenia. *Journal of Psychiatry and Neuroscience*, *29*(2):89-98.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1):23-49.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., and Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*(5):489–530.

Kutas, M. and Dale, A. (1997). Electrical and magnetic reading of mental functions. In M. D. Rugg (Ed.). *Cognitive neuroscience*. Hove, UK: Psychology Press, pp. 197-242.

Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427):203-205.

Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947):161-163.

Kutas, M. and Hillyard, S. A. (1989). An electrophysiological probe on incidental semantic association. *Journal of Cognitive Neuroscience*, *1*(1):38-49.

Kutas, M. and Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.). *Handbook of psycholinguistics*. San Diego, CA: Academic Press, pp. 83-143.

Kutas, M., Lindamood, T.E., and Hillyard, S.A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum and J. Requin (Eds.). *Preparatory states and processes*. Hillsdale, NJ: Lawrence Erlbaum, pp. 217-237.

Kutas, M., Neville, H, J., and Holcomb, P, J, (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening, and signing, *Electroencephalography and Clinical Neurophysiology*, *39*:325-330.

Langacker, R. (1987). *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.

Lau, E., Namyst, A., Fogel, A. and Delgado, T. (2016). A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabra*, *2*(1):1-19.

Laufer, B. and Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2):647-672.

Laurent, J., Denhières, G., Passerieux, C., Iakimovac, G., and Hardy-Baylé, M. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research, 1068*(1):151–160.

Leech, G. (1987). General introduction. In R. Garside, G. Leech, and G. Sampson (Eds.). *The computational analysis of English: A corpus-based approach*. London: Longman, pp. 1-15.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.). *Directions in Corpus Linguistics: Proceedings of Nobel Symposium, 4-8 August 1991*. Berlin: Mouton de Gruyter, pp. 105–122.

Leech, G. (2015). Descriptive Grammar. In D. Biber and R. Reppen (Eds.). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, pp. 146-160.

Leech, G., Myers, G., and Thomas, J. (1995). *Spoken English on computer: Transcription, mark-up and application*. Essex: Longman.

Lei, X. and Liao, K. (2017). Understanding the influences of EEG reference: A large-scale brain network perspective. *Frontiers in Neuroscience*, *11*:1-11.

Lintean, M. and Rus, V. (2012). Measuring semantic similarity in short texts through greedy pairing and word semantics. In Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference. Florida, CA, pp. 244-249.

Longstaff, A. (2005). *Neuroscienc*e (2nd edn.). Oxford: Taylor & Francis.

Lopez-Calderon, J. and Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*(213):1-18.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis and E. Tognini-Bonelli (Eds.). *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins, pp. 157-176.

Louw, B. (2010). “The painting is where the paint is not”: Reflections on the “Bootcamp Debate”. *International Journal of Corpus Linguistics*, *15*(3):344–353.

Love, R. (2017). The Spoken British National Corpus 2014: Design, compilation and analysis (Unpublished doctoral thesis). Lancaster University, Lancaster.

Love, R., Dembry, C., Hardie, A., Brezina, V., and McEnery, T. (2017). The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, *22*(3):319-344.

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. C. Handy (Ed.). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press, pp. 17-32.

Luck, S. J. (2014a). *An introduction to the event-related potential technique* (2nd edn.). Cambridge, MA: MIT Press.

Luck, S. J. (2014b). Online-only content: Chapter 15 Reading, writing, and reviewing ERP Papers. Retrieved from https://mitpress.mit.edu/books/introduction-event-related-potential-technique-0.

Luck, S. J. (2014c). Online-only content: Chapter 16 Setting up and running an ERP lab. Retrieved from https://mitpress.mit.edu/books/introduction-event-related-potential-technique-0.

Luck, S., Lopez-Calderon, J., Huang, S., and Foo, E. (2011). ERPLAB Toolbox tutorial: User's manual version 1.1.

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1):4-18.

Mangun G. R., and Hillyard S. A. (1991). Modulation of sensory-evoked brain potentials provide evidence for changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4):1057–1074.

Mai, L. L., Owl, M. Y., and Kersting, M. P. (2005). *Cambridge dictionary of human biology and evolution*. Cambridge: Cambridge University Press.

Mai, X., Xu, L., Li, M., Shao, J., Zhao, Z., deRegnier, R., Nelson, C. A., and Lozoff, B. (2012). Auditory recognition memory in 2-month-old infants as assessed by event-related potentials. *Developmental Neuropsychology*, *37*(5):400-414.

Majumdar, K. (2018). *A brief survey of quantitative EEG*. London: CRC Press.

Martin, A. E., Nieuwland, M. S. and Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage*, *59*(2):1859-1869.

Martin, E. A. (2010). *Concise medical dictionary* (8th edn.). Oxford: Oxford University Press.

Martin, G. N. (2006). *Human neuropsychology* (2nd edn.). London: Pearson.

Marton, M. and Szirtes, J. (1982). Averaged lambda potential and visual information processing. *Studia Psychologica*, *24*(3):165-170.

Marzi, T. and Viggiano, M. P. (2007). Interplay between familiarity and orientation in face processing: an ERP study. *International Journal of Psychophysiology, 65*(3):182-192.

Mason, O. (2000). Parameters of collocation: The word in the centre of gravity. In J. Kirk (Ed.). *Corpora galore*. Amsterdam: Rodopi, pp. 267-280.

MathWorks (2015). MATLAB version 8.5.1. Natick, Massachusetts: The MathWorks Inc.

Mauranen, A. (1993). Theme and prospection in written discourse. In M. Baker, F. Francis, and E. Tognini-Bonelli (Eds.). *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins, pp. 95-114.

McCauley, S. M. and Christiansen. M. H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science*, *9*(3):637-652.

McDonald, S. A. and Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*(6):648-652.

McDonald, S. A. and Shillcock, R. C. (2003b). Low-level predictive inference in reading: the influence of transition probabilities on eye movements. *Vision Research*, *43*(16): 1735-51.

McDonough, K. and Trofimovich, P. (2012). How to use psycholinguistic methodologies for comprehension and production. In A. Mackey and S. M. Gass (Eds.). *Research methods in second language acquisition: A practical guide.* Chichester: Wiley-Blackwell, pp. 117-138.

McEnery, T. and Wilson, A. (2004). *Corpus linguistics* (2nd edn.). Edinburgh: Edinburgh University Press.

McEnery, T. and Hardie, A. (2012). *Corpus linguistics: Methods, theory and practice*. Cambridge: Cambridge University Press.

Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2):227-234.

Michelizzi, J. (2005). Semantic relatedness applied to all words sense disambiguation (Unpublished MSc dissertation). University of Minnesota, Minnesota.

Millar, N. (2010). The processing of learner collocations (Unpublished doctoral thesis). Lancaster University, Lancaster.

Mitchell, T. F. (1971). Linguistic 'goings-on': Collocations and other lexical matters arising on the syntactic record. *Archivum Linguisticum*, *2*:35-69.

Molinaro, N. and Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology, 83*(3):176-190.

Molinaro, N., Barraza, P., and Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, *72*:120-132.

Moss, H. E., Hare, M. L., Day, P., and Tyler, L. K. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science*, *6*(4):413-427.

Nesselhauf, N. (2004). What are collocations? In D. J. Allerton, N. Nesselhauf, and P. Skandera (Eds.). *Phraseological units: Basic concepts and their application*. Basel: Schwabe, pp. 1-22.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Nieuwenhuism S., Ridderinkhof, K., Blom, J., Guido, B., and Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5):752-760.

Nieuwland, M. S., Martin, A. E., and Carreiras, M. (2013). Event-related brain potential evidence for animacy processing asymmetries during sentence comprehension. *Brain and Language*, *126*:151-158.

Noirhomme, Q. and Lehembre, R. (2012). Electrophysiology and disorders of consciousness. In C. Schnakers and S. Laureys (Eds.). *Coma and disorders of consciousness*. London: Springer, pp. 55-66.

Nolen-Hoeksema, S., Fredrickson, L., Loftus, G. R., and Wagenaar, W. A. (2009). *Atkinson and Hilgard's introduction to psychology* (15th edn.). Hampshire: Cengage Learning.

Núñez-Peña, M. A. and Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, *18*(2):130-141.

Ochoa, C. J. and Polich, J. (2000). P300 and blink instructions. *Clinical Neurophysiology*, *111*(1):93-98.

Osman, A., Lou, L. G., Muller-Gethmann, H., Rinkenauer, G, Mattes, S., and Ulrich, R. (2000). Mechanisms of speed-accuracy tradeoff: Evidence from covert motor processes. *Biological Psychology*, *51*(2):173-199.

Osterhout, L. and Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Language and Cognitive Processes*, *8*(4):785-806.

Otten, L. J. and Rugg, M. D. (2005). Interpreting event-related brain potentials. In T. C. Hardy (Ed.). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press, pp. 3-16.

Otten, L. J. and Rugg, M. D. (2005). Interpreting event-related brain potentials. In T. C. Handy (Ed.). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press, pp. 3-16.

Papadelis, C., Grant, P. E., Okada, Y., and Preissl, H. (2015). Editorial on emerging neuroimaging tools for studying normal and abnormal human brain development. In C. Papadelis, P. E. Grant, Y. Okada, and H. Preissl (Eds.). *Magnetoencephalography: An emerging neuroimaging tool for studying normal and abnormal human brain development*. Lausanne: Frontiers Media, pp. 6-7.

Parker Labs. (2017). Electromedical products: Signagel electrode gel. Retrieved from http://www.parkerlabs.com/signagel.asp

Patel, A. D., Gibson, E., Ratner, J., Besson, M. and Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6):717-733.

Pawley, A. and Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards and R. W. Schmidt (Eds.). *Language and communication*. London: Longman, pp. 191-226.

Pederson, T., Patwardhan, S., Banerjee, S. and Michelizzi, J. (2005-2008). *WordNet::Similarity* (version 2.07). http://wn-similarity.sourceforge.net/, last accessed 03/01/2017.

Picton, T. W. (1993). The P300 wave of the human event-related potentials. *Journal of Clinical Neurophysiology*, *9*(4):456–479.

Picton, T. W., Otavio, G. L. and Scherg, M. (2005). The recording and analysis of event-related potentials. In F. Boller and J. Grafman (Eds.). *Handbook of neuropsychology*. Amsterdam: Elsevier Science, pp. 3-73.

Pinel, J. P. J. (2014). *Biopsychology* (8th edn.). Essex: Pearson.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10):2128-2148.

Polich, J. (2012). Neuropsychology of P300. In E. S. Kappenman and S. J. Luck (Eds.). *The Oxford handbook of event-related potentials*. Oxford: Oxford University Press, pp. 159-188.

Postman, L. and Keppel, G. (1970). *Norms of word association*. New York: Academic Press.

Princeton University (2010). *WordNet* (version 3.0). https://wordnet.princeton.edu/, last accessed 03/01/2017.

Raichle, M. E. (2001). Functional neuroimaging: A historical and physiological perspective. In R. Cabeza and A. Kingstone (Eds.). *Handbook of functional neuroimaging of cognition*. Cambridge, MA: MIT Press, pp. 3-26.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3):372-422.

Rayner, K., Pollatsek, A., Ashby, J., and Clifton, C. (2012). *Psychology of reading* (2nd edn.). London: Psychology Press.

Reis, P. M. R., Hebenstreit, F., Gabsteiger, F., von Tscharner, V., and Lochmann, M. (2014). Methodological aspects of EEG and body dynamics measurements during motion. Kingsley, R. E. (2000). *Concise text of neuroscience*, *8*(156):1-19.

Reiterer, S., Hemmelmann, C., Rappelsberger, P. and Berger, M. L. (2005). Characteristic functional networks in high- versus low-proficiency second language speakers detected also during native language processing: an explorative EEG coherence study in 6 frequency bands. *Cognitive Brain Research*, *25*:566-578.

Reiterer, S., Pereda, E. and Bhattacharya, J. (2009). Measuring second language proficiency with EEG synchronization: How functional cortical networks and hemispheric involvement differ as a function of proficiency level in second language speakers. *Second Language Research*, *25*:77-106.

Reiterer, S., Pereda, E. and Bhattacharya, J. (2011). On a possible relationship between linguistic expertise and EEG gamma band phase synchrony. *Frontiers in Psychology*, *2*:1–11.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *International Joint Conference for Artificial Intelligence*, *1*:448–453.

Rhodes, S. M. and Donaldson, D. I. (2008). Association and not semantic relationships elicit the N400 effect: Electrophysiological evidence from an explicit language comprehension task. *Psychophysiology, 45*(1):50-59.

Richards, J. E. (1995). Infant cognitive psychophysiology: Normal development and implications for abnormal development outcomes. In T. H. Ollendick and R. J. Prinz (Eds.). *Advances in clinical child psychology, volume 17*. New York: Plenum Press, pp. 77-88.

Richards, J. E. (1995). Infant cognitive psychophysiology: Normal development and implications for abnormal development outcomes. In T. H. Ollendick and R. J. Prinz (Eds.). *Advances in clinical child psychology, volume 17*. New York: Plenum Press, pp. 77-88.

Rinkenauer G., Osman A., Ulrich R., Mandüller-Gethmann H., and Mattes S. (2004). On the locus of speed-accuracy trade-off in reaction time: inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, *133*(2):261–281.

Roehm, D., Bornkessel-Schlesewksy, I., Rosler, F., and Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience, 19*(8):1259-1274.

Rosenzweig, M. R., Leiman, A. L., and Breedlove, S. M. (1996). *Biological psychology*. Sunderland, MA: Sinauer Assoviates.

Rossion, B. and Jacques, C. (2012). The N170: Understanding the time course of face perception in the human brain. In E. S. Kappenman and S. J. Luck (Eds.). *The Oxford handbook of event-related potentials*. Oxford: Oxford University Press, pp. 115-141.

Rotschafer, S. E. and Razak, K. A. (2014). Auditory processing in fragile X syndrome. *Frontiers in cellular neuroscience*, *8*(19):18-28.

Rugg, M. D. and Coles, M. G. H. (1995). The ERP and cognitive psychology: Conceptual issues. In M. D. Rugg and M. G. H. Coles (Eds.). *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford: Oxford University Press, pp. 27-39.

Schmitt, N. and Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins, pp. 173-190.

Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools Inc.

Scott, M. and Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.

Seretan, V. (2011). *Syntax-based collocation extraction*. Dordrecht: Springer.

Severens, E., Jansma, B.. M. and Hartsuiker, R. J. (2008). Morphological influences on the comprehension of subject-verb agreement: An ERP study. *Brain Research*, *1228*:135-144.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRead, K., Martin, A. and Barsalou, W. (2007). A common neural substrate for perceiving and knowing about colour. *Neuropsychologia*, *45*(12):2802-2810.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1992). Shared knowledge. Paper presented at the Georgetown Round Table Conference.

Sinclair, J. (1996). The search for units of meaning. *Textus*, *9*(1):75-0106.

Sinclair, J. (2004). *Trust the text: Language, corpus, and discourse*. London: Routledge.

Sinclair, J. and Mauranen, A. (2006). *Linear unit grammar: Integrating speech and writing*. Amsterdam: John Benjamins.

Sinclair, J., Jones, S., Daley, R. and Krishnamurthy, R. (2004). *English collocation studies: The OSTI Report*. London: Continuum.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E. and van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, *175*:111-122.

Smadja, F., McKeown, K. R. and Hatzivassiloglou, V. (1996). Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics*, *22*(1):1-38.

Smith, L. (2009). *ActiveTwo System Operating Guidelines*. Wilmington, NC: Cortech Solutions.

Smulders, F. T. Y. and Miller, J. O. (2012). The lateralized readiness potential. In E. S. Kappenman and S. J. Luck (Eds.). *The Oxford handbook of event-related potentials*. Oxford: Oxford University Press, pp. 209-229.

Spence, D. P and Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, *19*(6):317-330.

Spöttl, C. and McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (Ed.). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins, pp. 191-226.

Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory*, *2*(1):61-77.

Stefanowitsch, A. and Gries, S. Th. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, *8*(2):209-243.

Sternberg, R. J. (2009). *Cognitive psychology*. Belmont, CA: Wadsworth.

Stubbs, M. (1993). British traditions in text analysis: from Firth to Sinclair. In M. Baker, F. Francis, and E. Tognini-Bonelli (Eds.). *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins, pp. 1-36.

Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Sur, S. and Sinha, V. K. (2009). Event-related potentials: An overview. *Industrial Psychiatry Journal*, *18*(1):70-73.

Sutton, S., Braren, M., Zubin, J., and John, E.R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700):1187-1188.

Swaab, T. Y., Ledoux, K., Camblin, C. C., and Boudewyn, M. (2012). Language-related ERP components. In S. J. Luck and E. S. Kappenman (Eds.). *The Oxford handbook of event-related potentials*. New York: Oxford University Press, pp. 397-439.

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

Tanner, D., Morgan-Short, K., and Luck, S. J. (2015). How inappropriate high-pass filters can produce artefactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*(8):997-1009.

Taylor, J. R. (2002). *Cognitive grammar*. Oxford: Oxford University Press.

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, *2*(2):1-11.

Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics 10*(1):1-13.

Teubert, W. (2010). Our brave new world? *International Journal of Corpus Linguistics*, *15*(3):354-358.

Theakston, A. and Lieven, E. (2017). Multiword sequences in first language acquisition. *Topics in Cognitive Science*, *9*(3):588-603.

The MathWorks (2015). *MATLAB Release 2015a*. Natick, MA.

Thompson, R. F. (2000). *The brain: A neuroscience primer* (3rd edn.). New York: Worth Publishers.

Thompson-Schill, S. L., Bedny, M., and Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neuropsychology*, *15*(2):219-224.

Tiago-Costa, E., Quelhas-Costa, E., and Santos-Baptista, J. (2016). Changes in EEG amplitude (alpha and beta waves) with thermal environment. *DYNA*, *83*(197):87-93.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Tomasello, M. (2001). The Item-Based Nature of Children's Early Syntactic Development. In Tomasello, M. and Bates, E. (Eds). *Language development: The essential readings*. Oxford: Blackwell, pp. 169-186.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. and Bates, E. (2001). *Language development: The essential readings*. Oxford: Blackwell Publishing.

Torres Cacoullous, R. and Walker, J. (2011). Collocations in grammaticalization and variation. In B. Heine and H. Narrog (Eds.). *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press, pp. 225-238.

Tremblay, A. and Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.). *Perspectives on formulaic language: Acquisition and communication*. London: Continuum, pp. 151-173.

Tyner, F. S., Knott, J. R. and Mayer, W. B. (1983). *Fundamentals of EEG technology: Basic concepts and methods*. London: Lippincott Williams and Wilkins.

Underwood, G., Schmitt, N., and Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins, pp. 153-172.

van Boxtel, G. J. M. (1998). Computational and statistical methods for analysing event-related potential data. *Behaviour Research Methods, Instruments, and Computers*, *31*(1):87-102.

van der Lubbe, R. H., Jaśkowski, P., Wauschkuhn, B., and Verleger, R. (2001). Influence of time pressure in a simple response task, a choice-by-location task, and the Simon task. *Journal of Psychophysiology*, *15*(4):241–255.

van Lancker Sidtis, D. and Fromkin, V. A. (2017). Formulaic language performance in left- and right-hemisphere damaged patients: Structured testing. *Aphasiology*, *31*(1):82-99.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioural and Brain Sciences*, *11*(3):343-427.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S. and Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8):1682-1700.

Wahl, A. K. and Gries, S. Th. (in press). Multi-word expressions: A novel computational approach to their bottom-up statistical extraction. In P. Cantos (Ed.). *Lexical collocation analysis: advances and applications*. Berlin & New York: Springer.

Webb, S. J., Long, J. D., and Nelson, C. A. (2005). A longitudinal investigation of visual event-related potentials in the first year of life. *Developmental Science*, *8*(6):605-616.

Weisstein, E. W. (1999). Eric Weisstein's World of Mathematics. Wolfram Inc. An online encyclopedia. http://mathworld.wolfram.com/.

Wiechmann, D. (2008). On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory*, *4*(2):253-290.

Woodman, G. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8):2031-2046.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*(1):231-254.

Wray, A. and Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, *20*(1):1-28.

Wray, A. (2017). Formulaic sequences as a regulatory mechanism for cognitive peturbations during the achievment of social goals. *Topics in Cognitive Science*, *9*(3):569-587.

Xue, S., Ren, G., Kong, X., Liu, J. and Qiu, J. (2015). Electrophsyiological correlated related to the conflict adaptation effect in an emotional conflict task. *Neuroscience Letters*, *584*:219-223.

Zillmer, E. A., Spiers, M. V. and Culbertson, W. C. (2008). *Principles of neuropsychology*. Belmont, CA: Thomson Wadsworth.

APPENDICES

Appendix 1: Counterbalanced stimuli lists used in Experiment 1

List A – collocational items come before non-collocational items (experimental bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth century** was a time of religious revival and controversy.
> *There was no religious controversy in the nineteenth century.*

The **head teacher** in the local primary school was featured in the local newspaper.
> *The head teacher in the local primary school was featured in a national newspaper.*

Claudia passed the interview, but had to take an aptitude test before she got the job.
> *Claudia failed the interview.*

The **chief executives** are optimistic about the prospects of their companies.
> *More than one chief executive is optimistic about the prospects of their companies.*

The scientific study showed that female smokers in older age groups were at greatest risk.
> *Young male smokers are at greatest risk.*

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
> *Most people do not lose weight on 1,500 calories per day.*

There is a **wide range** of new products on offer and this is attracting many new customers.
> *There are not many new products.*

The **nineteenth position** in the competition's scoring system is second to last.
> *The competition has a scoring system.*

The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.
> *The legislation that exists to protect endangered species is always adequate.*

Eighty five year old Mrs. Brown can still recite poems she learnt off by heart at primary school.
> *Mrs. Brown has forgotten all of the poems that she learnt at primary school.*

The **chief definitions** are listed at the top of each dictionary entry.
> *The top of each dictionary entry contains the chief definition.*

The **minimum wage** is designed to help people in low pay service industries.
> *The minimum wage is designed to help people in high pay service industries.*

There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.
> *The wide city is on top of the hill.*

The judge found Paul guilty and gave him one hundred hours of community service and a heavy fine.
> *The judge decided that Paul was not guilty.*

The **head character** in the local production of Annie is played by a girl from the local primary school.
> *A girl from the local primary school is involved in a local production of Annie.*

It is a **classic example** of losing sight of the wood for the trees
> *It is a classic example.*

After David lost his job, he just sat around doing nothing all the time.
> *David sat around doing nothing after losing his job.*

The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.
> *The legislation that exists to protect endangered fish lacks proper enforcement.*

The **random sample** of 50 observations was sufficient to carry out statistical tests.

The random sample contains 65 observations.

Most of the people on the island earned a living through tourism or fishing.

There is very little tourism or fishing on the island.

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.

The minimum prize included £500 plus four nights in a luxury hotel.

The high cost of **clinical trials** is delaying progress in medical research.

Clinical trials are expensive.

It is a **classic company** with an excellent reputation and a large workforce.

The company has an excellent reputation.

The brochure said the hotel was a five-minute walk from the beach, but actually it was over a mile.

It took more than 5 minutes to walk from the hotel to the beach.

The **random content** of the article made it difficult to identify the target audience.

The content of the article was random.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

The government announced plans to build a high speed train link from the airport to the city.

The high speed train link from the airport to the city has already been built.

The high cost of **clinical devices** is delaying progress in medical research.

Clinical devices are expensive.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

About 500,000 people stood in the pouring rain to listen to the Pope speak.

It was not raining when the Pope was speaking.

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

Amy was telling John about the **profound person** that she met on holiday many years ago.

John told Amy about a person who he met on holiday.

Andy’s company tried to encourage staff to use public transport instead of driving to work.

Andy’s company prefer it when people drive to work instead of using public transport.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.

The statement concerned the regional distribution of rural growth.

All the sunbathing Sandy did as a teenager has caused serious damage to her skin.

Sandy’s skin is seriously damaged.

The **key issues** associated with the development will be discussed in the monthly meeting.

The key issues will be discussed in the weekly meeting.

It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

Nicole’s dad has very strong opinions on the subject of politics and immigration.

Nicole has very strong opinions about politics and immigration.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.

The competition was in the March issue of the Clothes Show Magazine.

During the meeting with her boss, Linda listened carefully and took notes of the main points.

Linda did not listen during the meeting.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

List B – non-collocational items come before collocational *items* (experimental bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth position** in the competition's scoring system is second to last.

The competition has a scoring system.

The **head character** in the local production of Annie is played by a girl from the local primary school.

A girl from the local primary school is involved in a local production of Annie.

Claudia passed the interview, but had to take an aptitude test before she got the job.

Claudia failed the interview.

The **chief definitions** are listed at the top of each dictionary entry.

The top of each dictionary entry contains the chief definition.

The scientific study showed that female smokers in older age groups were at greatest risk.

Young male smokers are at greatest risk.

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.

The competition was in the March issue of the Clothes Show Magazine.

There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.

The wide city is on top of the hill.

The **nineteenth century** was a time of religious revival and controversy.

There was no religious controversy in the nineteenth century.

The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered fish lacks proper enforcement.

Eighty five year old Mrs. Brown can still recite poems she learnt off by heart at primary school.

Mrs. Brown has forgotten all of the poems that she learnt at primary school.

The **chief executives** are optimistic about the prospects of their companies.

More than one chief executive is optimistic about the prospects of their companies.

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.

The minimum prize included £500 plus four nights in a luxury hotel. FALSE

There is a **wide range** of new products on offer and this is attracting many new customers.

There are not many new products.

The judge found Paul guilty and gave him one hundred hours of community service and a heavy fine.

The judge decided that Paul was not guilty.

The **head teacher** in the local primary school was featured in the local newspaper.

The head teacher in the local primary school was featured in a national newspaper.

It is a **classic company** with an excellent reputation and a large workforce.

The company has an excellent reputation.

After David lost his job, he just sat around doing nothing all the time.

David sat around doing nothing after losing his job.

The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered species is always adequate.

The **random content** of the article made it difficult to identify the target audience.

The content of the article was random.

Most of the people on the island earned a living through tourism or fishing.

There is very little tourism or fishing on the island.

The **minimum wage** is designed to help people in low pay service industries.

The minimum wage is designed to help people in high pay service industries.

The high cost of **clinical devices** is delaying progress in medical research.

Clinical devices are expensive.

It is a **classic example** of losing sight of the wood for the trees.

It is a classic example.

The brochure said the hotel was a five-minute walk from the beach, but actually it was over a mile.

It took more than 5 minutes to walk from the hotel to the beach.

The **random sample** of 50 observations was sufficient to carry out statistical tests.

The random sample contains 65 observations.

Amy was telling John about the **profound person** that she met on holiday many years ago.

John told Amy about a person who he met on holiday.

The government announced plans to build a high speed train link from the airport to the city.

The high speed train link from the airport to the city has already been built.

The high cost of **clinical trials** is delaying progress in medical research.

Clinical trials are expensive.

The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.

The statement concerned the regional distribution of rural growth.

About 500,000 people stood in the pouring rain to listen to the Pope speak.

It was not raining when the Pope was speaking.

It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

Andy's company tried to encourage staff to use public transport instead of driving to work.

Andy's company prefer it when people drive to work instead of using public transport.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

All the sunbathing Sandy did as a teenager has caused serious damage to her skin.

Sandy's skin is seriously damaged.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

Nicole's dad has very strong opinions on the subject of politics and immigration.

Nicole has very strong opinions about politics and immigration.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.
The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
Most people do not lose weight on 1,500 calories per day.
During the meeting with her boss, Linda listened carefully and took notes of the main points.
Linda did not listen during the meeting.
The **key issues** associated with the development will be discussed in the monthly meeting.
The key issues will be discussed in the weekly meeting.

List C – half of the collocational items come before the non-collocational items (experimental bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth century** was a time of religious revival and controversy.
There was no religious controversy in the nineteenth century.
The **head teacher** in the local primary school was featured in the local newspaper.
The head teacher in the local primary school was featured in a national newspaper
Claudia passed the interview, but had to take an aptitude test before she got the job.
Claudia failed the interview.
The **chief executives** are optimistic about the prospects of their companies.
More than one chief executive is optimistic about the prospects of their companies
The scientific study showed that female smokers in older age groups were at greatest risk.
Young male smokers are at greatest risk.
The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
The competition was in the March issue of the Clothes Show Magazine.
There is a **wide range** of new products on offer and this is attracting many new customers.
There are not many new products.
The nineteenth position in the competition's scoring system is second to last.
The competition has a scoring system.
The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.
The legislation that exists to protect endangered species is always adequate
Eighty five year old Mrs. Brown can still recite poems she learnt off by heart at primary school.
Mrs. Brown has forgotten all of the poems that she learnt at primary school.
The **chief definitions** are listed at the top of each dictionary entry.
The top of each dictionary entry contains the chief definition.
The **minimum wage** is designed to help people in low pay service industries.
The minimum wage is designed to help people in high pay service industries.
The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
Most people do not lose weight on 1,500 calories per day.
There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.
The wide city is on top of the hill.
The judge found Paul guilty and gave him one hundred hours of community service and a heavy fine.
The judge decided that Paul was not guilty.
The **head character** in the local production of Annie is played by a girl from the local primary school.

A girl from the local primary school is involved in a local production of Annie.

It is a **classic example** of losing sight of the wood for the trees.

It is a classic example.

After David lost his job, he just sat around doing nothing all the time.

David sat around doing nothing after losing his job.

The legislation that exists to protect endangered fish is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered fish lacks proper enforcement.

The **random content** of the article made it difficult to identify the target audience.

The content of the article was random.

Most of the people on the island earned a living through tourism or fishing.

There is very little tourism or fishing on the island.

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.

The minimum prize included £500 plus four nights in a luxury hotel.

The high cost of **clinical devices** is delaying progress in medical research.

Clinical devices are expensive.

It is a **classic company** with an excellent reputation and a large workforce.

The company has an excellent reputation.

The brochure said the hotel was a five-minute walk from the beach, but actually it was over a mile.

It took more than 5 minutes to walk from the hotel to the beach.

The **random sample** of 50 observations was sufficient to carry out statistical tests.

The random sample contains 65 observations.

Amy was telling John about the **profound person** that she met on holiday many years ago.

John told Amy about a person who he met on holiday.

The government announced plans to build a high speed train link from the airport to the city.

The high speed train link from the airport to the city has already been built.

The high cost of **clinical trials** is delaying progress in medical research.

Clinical trials are expensive.

The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.

The statement concerned the regional distribution of rural growth.

About 500,000 people stood in the pouring rain to listen to the Pope speak.

It was not raining when the Pope was speaking.

It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

Andy's company tried to encourage staff to use public transport instead of driving to work.

Andy's company prefer it when people drive to work instead of using public transport.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

All the sunbathing Sandy did as a teenager has caused serious damage to her skin.

Sandy's skin is seriously damaged.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

Nicole's dad has very strong opinions on the subject of politics and immigration.

Nicole has very strong opinions about politics and immigration.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.

Most people do not lose weight on 1,500 calories per day.

During the meeting with her boss, Linda listened carefully and took notes of the main points.

Linda did not listen during the meeting.

The **key issues** associated with the development will be discussed in the monthly meeting.

The key issues will be discussed in the weekly meeting.

List D – the other half of the collocational items come before the non-collocational items (experimental bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth position** in the competition's scoring system is second to last.

The competition has a scoring system.

The **head character** in the local production of Annie is played by a girl from the local primary school.

A girl from the local primary school is involved in a local production of Annie.

Claudia passed the interview, but had to take an aptitude test before she got the job.

Claudia failed the interview.

The **chief definitions** are listed at the top of each dictionary entry.

The top of each dictionary entry contains the chief definition.

The scientific study showed that female smokers in older age groups were at greatest risk.

Young male smokers are at greatest risk.

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.

Most people do not lose weight on 1,500 calories per day.

There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.

The wide city is on top of the hill.

The **nineteenth century** was a time of religious revival and controversy.

There was no religious controversy in the nineteenth century.

The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered fish lacks proper enforcement.

Eighty five year old Mrs. Brown can still recite poems she learnt off by heart at primary school.

Mrs. Brown has forgotten all of the poems that she learnt at primary school.

The **chief executives** are optimistic about the prospects of their companies.

More than one chief executive is optimistic about the prospects of their companies.

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.

The minimum prize included £500 plus four nights in a luxury hotel.

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.

The competition was in the March issue of the Clothes Show Magazine.

There is a **wide range** of new products on offer and this is attracting many new customers.

There are not many new products.

The judge found Paul guilty and gave him one hundred hours of community service and a heavy fine.

The judge decided that Paul was not guilty.

The **head teacher** in the local primary school was featured in the local newspaper.

The head teacher in the local primary school was featured in a national newspaper.

It is a **classic company** with an excellent reputation and a large workforce.

The company has an excellent reputation.

After David lost his job, he just sat around doing nothing all the time.

David sat around doing nothing after losing his job.

The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered species is always adequate.

The **random sample** of 50 observations was sufficient to carry out statistical tests.

The random sample contains 65 observations.

Most of the people on the island earned a living through tourism or fishing.

There is very little tourism or fishing on the island.

The **minimum wage** is designed to help people in low pay service industries.

The minimum wage is designed to help people in high pay service industries.

The high cost of **clinical trials** is delaying progress in medical research.

Clinical trials are expensive.

It is a **classic example** of losing sight of the wood for the trees.

It is a classic example.

The brochure said the hotel was a five-minute walk from the beach, but actually it was over a mile.

It took more than 5 minutes to walk from the hotel to the beach.

The **random content** of the article made it difficult to identify the target audience.

The content of the article was random.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

The government announced plans to build a high speed train link from the airport to the city.

The high speed train link from the airport to the city has already been built.

The high cost of **clinical devices** is delaying progress in medical research.

Clinical devices are expensive.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

About 500,000 people stood in the pouring rain to listen to the Pope speak.

It was not raining when the Pope was speaking.

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

Amy was telling John about the **profound person** that she met on holiday many years ago.

John told Amy about a person who he met on holiday.

Andy’s company tried to encourage staff to use public transport instead of driving to work.

Andy’s company prefer it when people drive to work instead of using public transport.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.

The statement concerned the regional distribution of rural growth.

All the sunbathing Sandy did as a teenager has caused serious damage to her skin.

Sandy's skin is seriously damaged.

The **key issues** associated with the development will be discussed in the monthly meeting.

The key issues will be discussed in the weekly meeting.

It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

Nicole's dad has very strong opinions on the subject of politics and immigration.

Nicole has very strong opinions about politics and immigration.

In the foreseeable weeks the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

During the meeting with her boss, Linda listened carefully and took notes of the main points.

Linda did not listen during the meeting.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

Appendix 2: Counterbalanced stimuli lists used in Experiments 2 and 3

List A – collocational items come before non-collocational items (bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth century** was a time of religious revival and controversy.
> *There was no religious controversy in the nineteenth century.*

The **head teacher** in the local primary school was featured in the local newspaper.
> *The head teacher in the local primary school was featured in a national newspaper.*

The **chief executives** are optimistic about the prospects of their companies.
> *More than one chief executive is optimistic about the prospects of their companies.*

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
> *Most people do not lose weight on 1,500 calories per day.*

There is a **wide range** of new products on offer and this is attracting many new customers.
> *There are not many new products.*

The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.
> *The legislation that exists to protect endangered species is always adequate.*

The **nineteenth position** in the competition's scoring system is second to last.
> *The competition has a scoring system.*

The **minimum wage** is designed to help people in low pay service industries.
> *The minimum wage is designed to help people in high pay service industries.*

The **chief definitions** are listed at the top of each dictionary entry.
> *The top of each dictionary entry contains the chief definition.*

It is a **classic example** of losing sight of the wood for the trees.
> *It is a classic example.*

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
> *The competition was in the March issue of the Clothes Show Magazine.*

The **head character** in the local production of Annie is played by a girl from the local primary school.
> *A girl from the local primary school is involved in a local production of Annie.*

There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.
> *The wide city is on top of the hill.*

The **random sample** of 50 observations was sufficient to carry out statistical tests.
> *The random sample contains 65 observations.*

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.
> *The minimum prize included £500 plus four nights in a luxury hotel.*

The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.
> *The legislation that exists to protect endangered fish lacks proper enforcement.*

The high cost of **clinical trials** is delaying progress in medical research.
> *Clinical trials are expensive.*

It is a **classic company** with an excellent reputation and a large workforce.
> *The company has an excellent reputation.*

Amy was telling John about the **profound effect** that the book had on her.
> *The book had a profound effect on Amy.*

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key issues** associated with the development will be discussed in the monthly meeting.

The key issues will be discussed in the weekly meeting.

The high cost of **clinical devices** is delaying progress in medical research.

Clinical devices are expensive.

Amy was telling John about the **profound person** that she met on holiday many years ago.

John told Amy about a person who he met on holiday.

The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.

The statement concerned the regional distribution of rural growth.

The **random content** of the article made it difficult to identify the target audience.

The content of the article was random.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

<u>List B – non-collocational items come before collocational *items* (bigrams in **bold**; T/F statements indented and in *italics*)</u>

The **nineteenth position** in the competition's scoring system is second to last.

The competition has a scoring system.

The **head character** in the local production of Annie is played by a girl from the local primary school.

A girl from the local primary school is involved in a local production of Annie.

The **chief definitions** are listed at the top of each dictionary entry.

The top of each dictionary entry contains the chief definition.

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.

The competition was in the March issue of the Clothes Show Magazine.

There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.

The wide city is on top of the hill.

The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered fish lacks proper enforcement.

The **nineteenth century** was a time of religious revival and controversy.

There was no religious controversy in the nineteenth century.

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.

The minimum prize included £500 plus four nights in a luxury hotel.

The **chief executives** are optimistic about the prospects of their companies.

More than one chief executive is optimistic about the prospects of their companies.

It is a **classic company** with an excellent reputation and a large workforce.

The company has an excellent reputation.

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.

Most people do not lose weight on 1,500 calories per day.

The **head teacher** in the local primary school was featured in the local newspaper.

The head teacher in the local primary school was featured in a national newspaper.

There is a **wide range** of new products on offer and this is attracting many new customers.

There are not many new products.

The **random content** of the article made it difficult to identify the target audience.

The content of the article was random.

The **minimum wage** is designed to help people in low pay service industries.

The minimum wage is designed to help people in high pay service industries.

The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered species is always adequate.

The high cost of **clinical devices** is delaying progress in medical research.

Clinical devices are expensive.

It is a **classic example** of losing sight of the wood for the trees.

It is a classic example.

Amy was telling John about the **profound person** that she met on holiday many years ago.

John told Amy about a person who he met on holiday.

It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.

The statement concerned the regional distribution of rural growth.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

The high cost of **clinical trials** is delaying progress in medical research.

Clinical trials are expensive.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

The **random sample** of 50 observations was sufficient to carry out statistical tests.

The random sample contains 65 observations.

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key issues** associated with the development will be discussed in the monthly meeting.

The key issues will be discussed in the weekly meeting.

List C – half of the collocational items come before the non-collocational items (bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth century** was a time of religious revival and controversy.
There was no religious controversy in the nineteenth century.
The **head teacher** in the local primary school was featured in the local newspaper.
The head teacher in the local primary school was featured in a national newspaper.
The **chief executives** are optimistic about the prospects of their companies.
More than one chief executive is optimistic about the prospects of their companies.
The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
Most people do not lose weight on 1,500 calories per day.
There is a **wide range** of new products on offer and this is attracting many new customers.
There are not many new products.
The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.
The legislation that exists to protect endangered species is always adequate.
The **nineteenth position** in the competition’s scoring system is second to last.
The competition has a scoring system.
The **minimum wage** is designed to help people in low pay service industries.
The minimum wage is designed to help people in high pay service industries.
The **chief definitions** are listed at the top of each dictionary entry.
The top of each dictionary entry contains the chief definition.
It is a **classic example** of losing sight of the wood for the trees.
It is a classic example.
The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
The competition was in the March issue of the Clothes Show Magazine.
The **head character** in the local production of Annie is played by a girl from the local primary school.
A girl from the local primary school is involved in a local production of Annie.
There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.
The wide city is on top of the hill.
The **random content** of the article made it difficult to identify the target audience.
The content of the article was random.
The high cost of **clinical devices** is delaying progress in medical research.
Clinical devices are expensive.
Amy was telling John about the **profound person** that she met on holiday many years ago.
John told Amy about a person who he met on holiday.
It was a **crucial night** for basketball in this country and there were many disappointed fans.
The basketball fans were very happy.
The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.
The statement concerned the regional distribution of rural growth.
The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.
The minimum prize included £500 plus four nights in a luxury hotel.
The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered fish lacks proper enforcement.

The high cost of **clinical trials** is delaying progress in medical research.

Clinical trials are expensive.

It is a **classic company** with an excellent reputation and a large workforce.

The company has an excellent reputation.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

The **random sample** of 50 observations was sufficient to carry out statistical tests.

The random sample contains 65 observations.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key issues** associated with the development will be discussed in the monthly meeting.

The key issues will be discussed in the weekly meeting.

List D – the other half of the collocational items come before the non-collocational items (bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth position** in the competition's scoring system is second to last.

The competition has a scoring system.

The **head character** in the local production of Annie is played by a girl from the local primary school.

A girl from the local primary school is involved in a local production of Annie.

The **chief definitions** are listed at the top of each dictionary entry.

The top of each dictionary entry contains the chief definition.

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.

The competition was in the March issue of the Clothes Show Magazine.

There is a **wide city** between the hills with a mixture of modern buildings and old cobbled streets.

The wide city is on top of the hill.

The legislation that exists to protect **endangered fish** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered fish lacks proper enforcement.

The **nineteenth century** was a time of religious revival and controversy.

There was no religious controversy in the nineteenth century.

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.

The minimum prize included £500 plus four nights in a luxury hotel.

The **chief executives** are optimistic about the prospects of their companies.

More than one chief executive is optimistic about the prospects of their companies.

It is a **classic company** with an excellent reputation and a large workforce.

The company has an excellent reputation.

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.

Most people do not lose weight on 1,500 calories per day.

The **head teacher** in the local primary school was featured in the local newspaper.

The head teacher in the local primary school was featured in a national newspaper.

There is a **wide range** of new products on offer and this is attracting many new customers.

There are not many new products.

The **random sample** of 50 observations was sufficient to carry out statistical tests.

The random sample contains 65 observations.

The high cost of **clinical trials** is delaying progress in medical research.

Clinical trials are expensive.

Amy was telling John about the **profound effect** that the book had on her.

The book had a profound effect on Amy.

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.

A very important point was raised in parliament.

The **massive increase** in the volume of traffic in rural areas is a threat to the English countryside.

There is an increase in traffic in rural areas.

The **minimum wage** is designed to help people in low pay service industries.

The minimum wage is designed to help people in high pay service industries.

The legislation that exists to protect **endangered species** is often inadequate and lacks proper enforcement.

The legislation that exists to protect endangered species is always adequate.

The high cost of **clinical devices** is delaying progress in medical research.

Clinical devices are expensive.

It is a **classic example** of losing sight of the wood for the trees.

It is a classic example.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key issues** associated with the development will be discussed in the monthly meeting.

The key issues will be discussed in the weekly meeting.

The **random content** of the article made it difficult to identify the target audience.

The content of the article was random.

Amy was telling John about the **profound person** that she met on holiday many years ago.

John told Amy about a person who he met on holiday.

It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

The **massive statement** concerning the regional distribution of urban growth was followed by two important studies.

The statement concerned the regional distribution of rural growth.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **key costs** associated with the development will be discussed in the monthly meeting.

The key costs will be discussed in the monthly meeting.

Appendix 3: Counterbalanced stimuli lists used in Experiment 4

List A – collocational items come before non-collocational items (bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth century** was a time of religious revival and controversy.
There was no religious controversy in the nineteenth century.
Twenty-four hours have indeed passed but the end of a day does not necessarily mean sleep.
Twenty-six hours have passed.
The **prime minister** was seen by the public to be working against political trends.
The public saw that the prime minister was working against political trends.
The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
Most people do not lose weight on 1,500 calories per day.
An **integral part** of delegation is the setting of targets for those given responsibility for tasks.
Setting targets is an integral part of delegation.
Disposable income will be severely reduced following the recent tax increases.
The recent tax increases will reduce disposable income.
The **nineteenth position** in the competition's scoring system is second to last.
The competition has a scoring system.
The **minimum wage** is designed to help people in low pay service industries.
The minimum wage is designed to help people in high pay service industries.
It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.
A very important point was raised in parliament.
In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.
The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
The competition was in the March issue of the Clothes Show Magazine.
The **prime period** for international expansion by transnational corporations was between 1963 and 1972.
International expansion by transnational corporations mostly took place in the 1980s.
An **integral thought** came into his mind but he was unable to articulate it.
He was able to clearly articulate the integral thought.
Twenty-four patients were recruited for the study into childhood asthma.
The study focused on childhood obesity.
The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.
The minimum prize included £500 plus four nights in a luxury hotel.
Disposable property continues to pile up in the streets yet the council is refusing to collect it.
There is disposable property piling up in the streets.
It was a **crucial night** for basketball in this country and there were many disappointed fans.
The basketball fans were very happy.
In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.

List B – non-collocational items come before collocational *items* (bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth position** in the competition's scoring system is second to last.
The competition has a scoring system.
Twenty-four patients were recruited for the study into childhood asthma.
The study focused on childhood obesity.
The **prime period** for international expansion by transnational corporations was between 1963 and 1972.
International expansion by transnational corporations mostly took place in the 1980s.
The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
The competition was in the March issue of the Clothes Show Magazine.
An **integral thought** came into his mind but he was unable to articulate it.
He was able to clearly articulate the integral thought.
Disposable property continues to pile up in the streets yet the council is refusing to collect it.
There is disposable property piling up in the streets.
The **nineteenth century** was a time of religious revival and controversy.
There was no religious controversy in the nineteenth century.
The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.
The minimum prize included £500 plus four nights in a luxury hotel.
It was a **crucial night** for basketball in this country and there were many disappointed fans.
The basketball fans were very happy.
In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.
The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
Most people do not lose weight on 1,500 calories per day.
The **prime minister** was seen by the public to be working against political trends.
The public saw that the prime minister was working against political trends.
An **integral part** of delegation is the setting of targets for those given responsibility for tasks.
Setting targets is an integral part of delegation.
Twenty-four hours have indeed passed but the end of a day does not necessarily mean sleep.
Twenty-six hours have passed.
The **minimum wage** is designed to help people in low pay service industries.
The minimum wage is designed to help people in high pay service industries.
Disposable income will be severely reduced following the recent tax increases.
The recent tax increases will reduce disposable income.
It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.
A very important point was raised in parliament.
In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.
Plans to build a new railway line have been cancelled.

List C – half of the collocational items come before the non-collocational items (bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth century** was a time of religious revival and controversy.
> *There was no religious controversy in the nineteenth century.*

Twenty-four hours have indeed passed but the end of a day does not necessarily mean sleep.
> *Twenty-six hours have passed.*

The **prime minister** was seen by the public to be working against political trends.
> *The public saw that the prime minister was working against political trends.*

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.
> *Most people do not lose weight on 1,500 calories per day.*

An **integral part** of delegation is the setting of targets for those given responsibility for tasks.
> *Setting targets is an integral part of delegation.*

Disposable property continues to pile up in the streets yet the council is refusing to collect it.
> *There is disposable property piling up in the streets.*

The **nineteenth position** in the competition's scoring system is second to last.
> *The competition has a scoring system.*

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.
> *The minimum prize included £500 plus four nights in a luxury hotel.*

It was a **crucial night** for basketball in this country and there were many disappointed fans.
> *The basketball fans were very happy.*

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.
> *Plans to build a new railway line have been cancelled.*

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.
> *The competition was in the March issue of the Clothes Show Magazine.*

The **prime period** for international expansion by transnational corporations was between 1963 and 1972.
> *International expansion by transnational corporations mostly took place in the 1980s.*

An **integral thought** came into his mind but he was unable to articulate it.
> *He was able to clearly articulate the integral thought.*

Twenty-four patients were recruited for the study into childhood asthma.
> *The study focused on childhood obesity.*

The **minimum wage** is designed to help people in low pay service industries.
> *The minimum wage is designed to help people in high pay service industries.*

Disposable income will be severely reduced following the recent tax increases.
> *The recent tax increases will reduce disposable income.*

It was a **crucial point** that was raised in parliament and it provoked an interesting discussion.
> *A very important point was raised in parliament.*

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.
> *Plans to build a new railway line have been cancelled.*

List D – the other half of the collocational items come before the non-collocational items (bigrams in **bold**; T/F statements indented and in *italics*)

The **nineteenth position** in the competition's scoring system is second to last.

The competition has a scoring system.

Twenty-four patients were recruited for the study into childhood asthma.

The study focused on childhood obesity.

The **prime period** for international expansion by transnational corporations was between 1963 and 1972.

International expansion by transnational corporations mostly took place in the 1980s.

The **vast opportunity** to win a complete makeover was offered in the March issue of the Clothes Show Magazine.

The competition was in the March issue of the Clothes Show Magazine.

An **integral thought** came into his mind but he was unable to articulate it.

He was able to clearly articulate the integral thought.

Disposable income will be severely reduced following the recent tax increases.

The recent tax increases will reduce disposable income.

The **nineteenth century** was a time of religious revival and controversy.

There was no religious controversy in the nineteenth century.

The **minimum wage** is designed to help people in low pay service industries.

The minimum wage is designed to help people in high pay service industries.

Disposable property continues to pile up in the streets yet the council is refusing to collect it.

There is disposable property piling up in the streets.

In the **foreseeable future** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

The **vast majority** of people will achieve a satisfactory weight loss on 1,500 calories a day.

Most people do not lose weight on 1,500 calories per day.

The **prime minister** was seen by the public to be working against political trends.

The public saw that the prime minister was working against political trends.

An **integral part** of delegation is the setting of targets for those given responsibility for tasks.

Setting targets is an integral part of delegation.

Twenty-four hours have indeed passed but the end of a day does not necessarily mean sleep.

Twenty-six hours have passed.

The **minimum prize** in the competition was £500 plus two nights in a luxury hotel.

The minimum prize included £500 plus four nights in a luxury hotel.

Disposable property continues to pile up in the streets yet the council is refusing to collect it.

There is disposable property piling up in the streets.

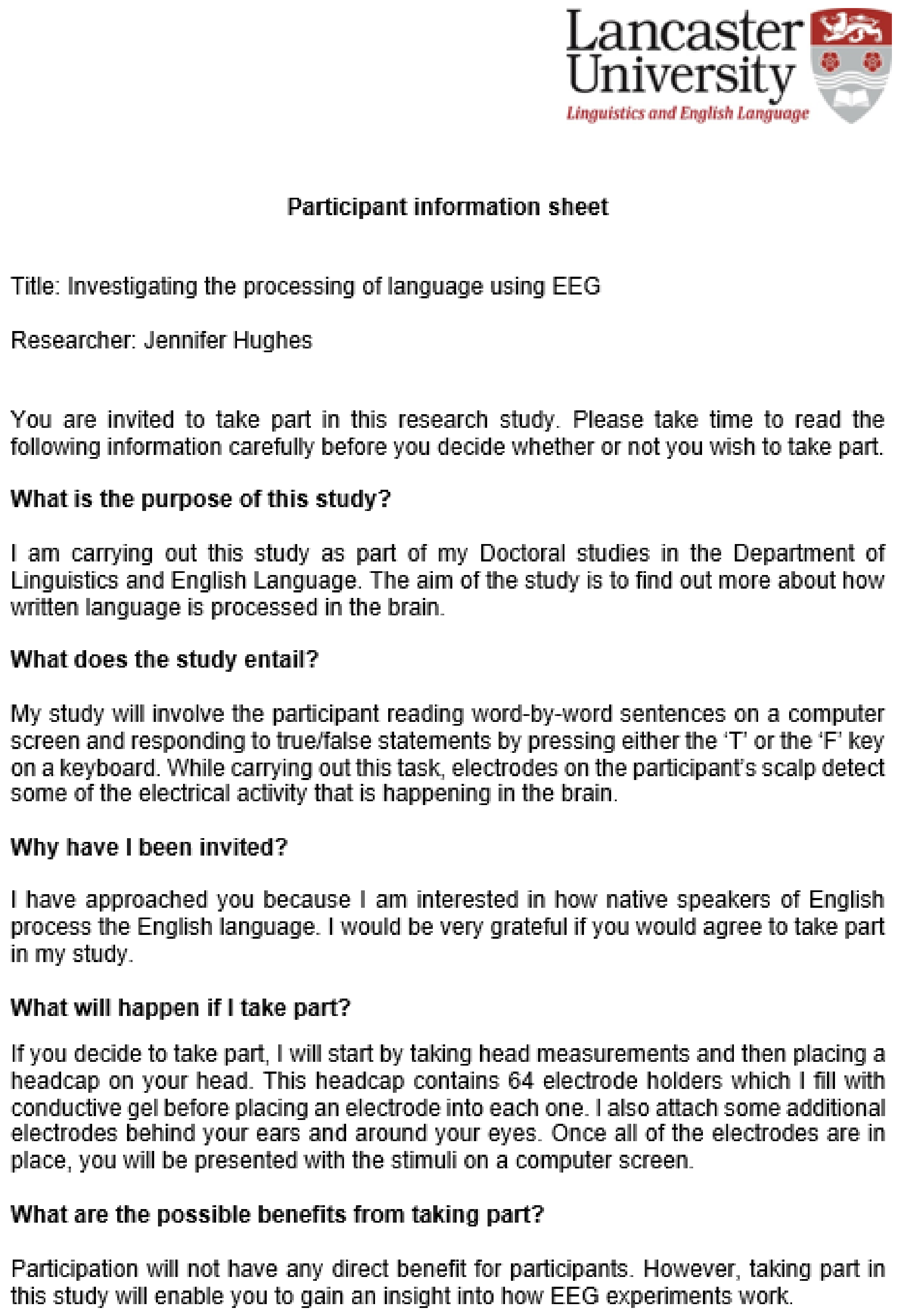
It was a **crucial night** for basketball in this country and there were many disappointed fans.

The basketball fans were very happy.

In the **foreseeable weeks** the new railway line will be built but the completion date has not yet been confirmed.

Plans to build a new railway line have been cancelled.

Appendix 4: Information sheet for native speakers of English



Lancaster University
Linguistics and English Language

Participant information sheet

Title: Investigating the processing of language using EEG

Researcher: Jennifer Hughes

You are invited to take part in this research study. Please take time to read the following information carefully before you decide whether or not you wish to take part.

What is the purpose of this study?

I am carrying out this study as part of my Doctoral studies in the Department of Linguistics and English Language. The aim of the study is to find out more about how written language is processed in the brain.

What does the study entail?

My study will involve the participant reading word-by-word sentences on a computer screen and responding to true/false statements by pressing either the 'T' or the 'F' key on a keyboard. While carrying out this task, electrodes on the participant's scalp detect some of the electrical activity that is happening in the brain.

Why have I been invited?

I have approached you because I am interested in how native speakers of English process the English language. I would be very grateful if you would agree to take part in my study.

What will happen if I take part?

If you decide to take part, I will start by taking head measurements and then placing a headcap on your head. This headcap contains 64 electrode holders which I fill with conductive gel before placing an electrode into each one. I also attach some additional electrodes behind your ears and around your eyes. Once all of the electrodes are in place, you will be presented with the stimuli on a computer screen.

What are the possible benefits from taking part?

Participation will not have any direct benefit for participants. However, taking part in this study will enable you to gain an insight into how EEG experiments work.

What are the possible disadvantages and risks of taking part?

There are no major disadvantages of taking part. However, taking part will involve investing a maximum of 1 hour and 30 minutes of your time.

What will happen if I decide not to take part or if I don't want to carry on with the study?

If you decide not to take part in this study, this will not affect your studies and the way that you are assessed on your course.

You are free to withdraw from the study at any time and you do not have to give a reason. If you withdraw while the study takes place or until 1 month after it finishes, I will not use any of the information or data that you provided. If you withdraw later, I will use the information and data that I obtained from you for my study.

Will my taking part in this project be kept confidential?

All the information collected about you during the course of the research will be kept strictly confidential. Any identifying information, such as names and personal characteristics, will be anonymised in the PhD thesis or any other publications of this research. The data that I will collect will be kept securely. Any paper-based data will be kept in a locked cupboard. Electronic data will be stored on a password protected computer and files containing personal data will be encrypted. The data will be kept securely for at least 10 years.

What will happen to the results of the research study?

The results of the study will be used for academic purposes only. This will include my PhD thesis and other publications, for example journal articles. I am also planning to present the results of my study at academic conferences.

What if there is a problem?

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact myself or my supervisor Dr Andrew Hardie (a.hardie@lancaster.ac.uk; 01534 593024; CASS, FASS Building, Lancaster University, Lancaster, Lancashire, LA1 4YW).

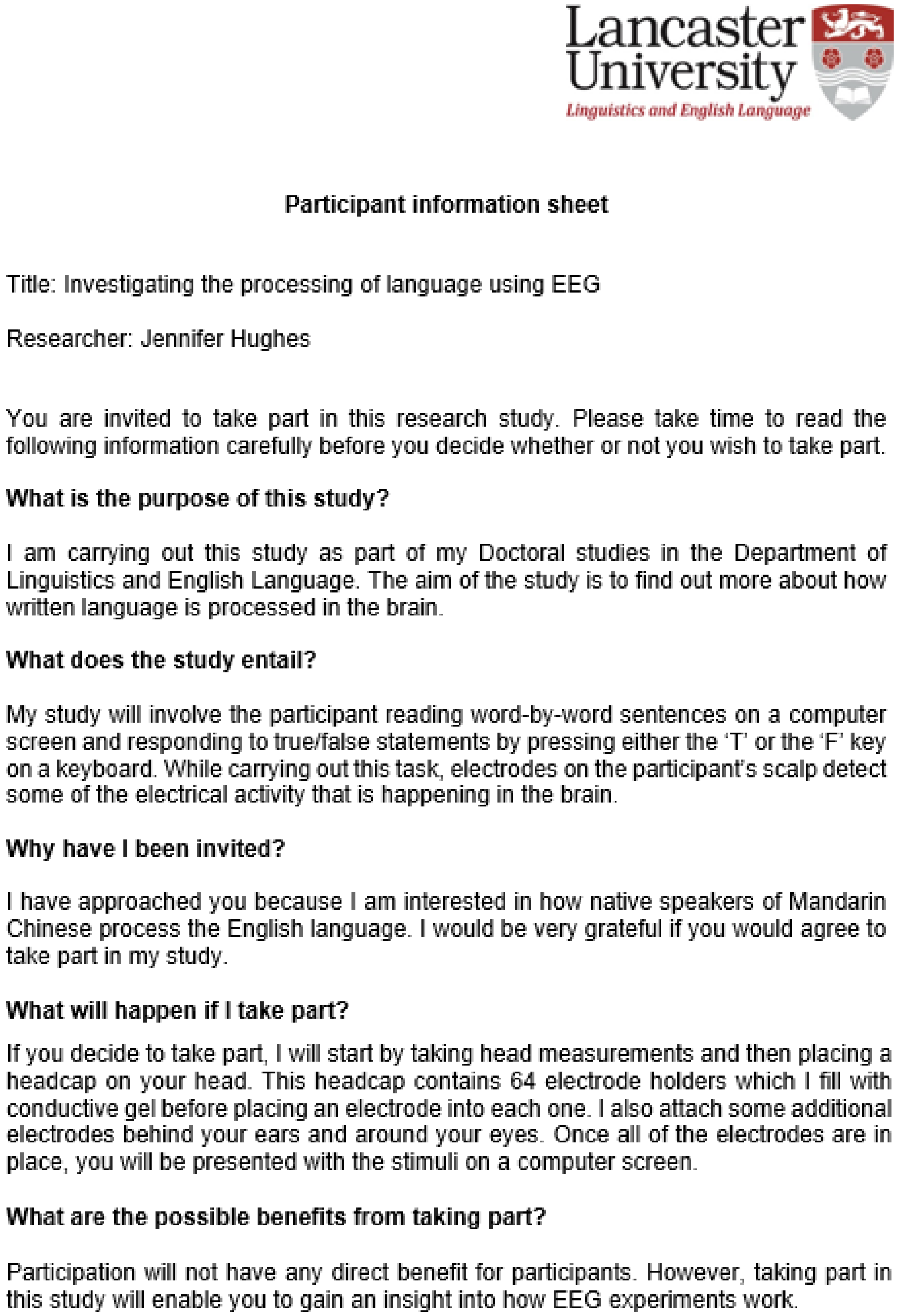
Further information and contact details

I can be contacted at j.j.hughes@lancaster.ac.uk, 07531 802425, or B16, CASS, FASS Building, Lancaster University, Lancaster, Lancashire, LA1 4YW.

This study has been approved by Lancaster University's ethics committee (UREC).

Thank you for considering your participation in this project.

Appendix 5: Information sheet for non-native speakers of English



Lancaster University
Linguistics and English Language

Participant information sheet

Title: Investigating the processing of language using EEG

Researcher: Jennifer Hughes

You are invited to take part in this research study. Please take time to read the following information carefully before you decide whether or not you wish to take part.

What is the purpose of this study?

I am carrying out this study as part of my Doctoral studies in the Department of Linguistics and English Language. The aim of the study is to find out more about how written language is processed in the brain.

What does the study entail?

My study will involve the participant reading word-by-word sentences on a computer screen and responding to true/false statements by pressing either the 'T' or the 'F' key on a keyboard. While carrying out this task, electrodes on the participant's scalp detect some of the electrical activity that is happening in the brain.

Why have I been invited?

I have approached you because I am interested in how native speakers of Mandarin Chinese process the English language. I would be very grateful if you would agree to take part in my study.

What will happen if I take part?

If you decide to take part, I will start by taking head measurements and then placing a headcap on your head. This headcap contains 64 electrode holders which I fill with conductive gel before placing an electrode into each one. I also attach some additional electrodes behind your ears and around your eyes. Once all of the electrodes are in place, you will be presented with the stimuli on a computer screen.

What are the possible benefits from taking part?

Participation will not have any direct benefit for participants. However, taking part in this study will enable you to gain an insight into how EEG experiments work.

What are the possible disadvantages and risks of taking part?

There are no major disadvantages of taking part. However, taking part will involve investing a maximum of 1 hour and 30 minutes of your time.

What will happen if I decide not to take part or if I don't want to carry on with the study?

If you decide not to take part in this study, this will not affect your studies and the way that you are assessed on your course.

You are free to withdraw from the study at any time and you do not have to give a reason. If you withdraw while the study takes place or until 1 month after it finishes, I will not use any of the information or data that you provided. If you withdraw later, I will use the information and data that I obtained from you for my study.

Will my taking part in this project be kept confidential?

All the information collected about you during the course of the research will be kept strictly confidential. Any identifying information, such as names and personal characteristics, will be anonymised in the PhD thesis or any other publications of this research. The data that I will collect will be kept securely. Any paper-based data will be kept in a locked cupboard. Electronic data will be stored on a password protected computer and files containing personal data will be encrypted. The data will be kept securely for at least 10 years.

What will happen to the results of the research study?

The results of the study will be used for academic purposes only. This will include my PhD thesis and other publications, for example journal articles. I am also planning to present the results of my study at academic conferences.

What if there is a problem?

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact myself or my supervisor Dr Andrew Hardie (a.hardie@lancaster.ac.uk; 01534 593024; CASS, FASS Building, Lancaster University, Lancaster, Lancashire, LA1 4YW).

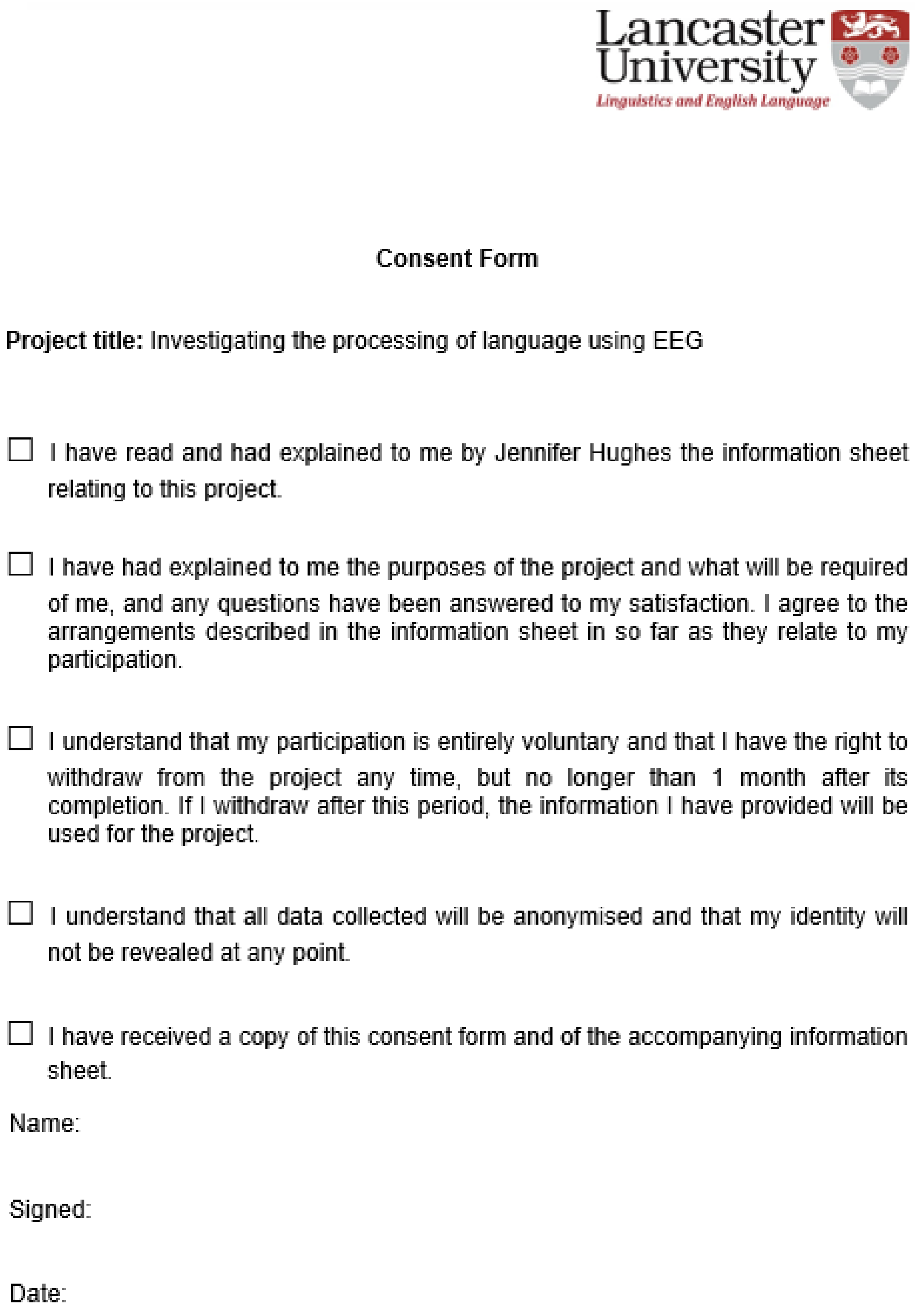
Further information and contact details

I can be contacted at j.j.hughes@lancaster.ac.uk, 07531 802425, or B16, CASS, FASS Building, Lancaster University, Lancaster, Lancashire, LA1 4YW.

This study has been approved by Lancaster University's ethics committee (UREC).

Thank you for considering your participation in this project.

Appendix 6: Consent form for native and non-native speakers of English



Lancaster University
Linguistics and English Language

Consent Form

Project title: Investigating the processing of language using EEG

☐ I have read and had explained to me by Jennifer Hughes the information sheet relating to this project.

☐ I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the information sheet in so far as they relate to my participation.

☐ I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time, but no longer than 1 month after its completion. If I withdraw after this period, the information I have provided will be used for the project.

☐ I understand that all data collected will be anonymised and that my identity will not be revealed at any point.

☐ I have received a copy of this consent form and of the accompanying information sheet.

Name:

Signed:

Date: