The Immediate Influence of Speaker Gaze on Situated Speech Comprehension

Evidence from Multiple ERP Components

Torsten Kai Jachmann

Dissertation
zur Erlangung des akademischen Grades
eines Doktors der Philosophie
an den Philosophischen Fakultäten
der Universität des Saarlandes

vorgelegt von
Torsten Kai Jachmann
aus Dudweiler

Saarbrücken, den 17.01.2020

Dekan der Philosophischen Fakultät II: Univ.-Prof. Dr. Heinrich Schlange-Schöningen

Erstgutachter: Prof. Dr. Matthew W. Crocker

Zweitgutachter: Dr. Maria Staudte

Tag der mündlichen Prüfung: 24.06.2020

In Memoriam Jana Jachmann

Abstract

This thesis presents results from three ERP experiments on the influence of speaker gaze on listeners' sentence comprehension with focus on the utilization of speaker gaze as part of the communicative signal. The first two experiments investigated whether speaker gaze was utilized in situated communication to form expectations about upcoming referents in an unfolding sentence. Participants were presented with a face performing gaze actions toward three objects surrounding it time aligned to utterances that compared two of the three objects. Participants were asked to judge whether the sentence they heard was true given the provided scene. Gaze cues preceded the naming of the corresponding object by 800ms. The gaze cue preceding the mentioning of the second object was manipulated such that it was either Congruent, Incongruent or Uninformative (Averted toward an empty position in experiment 1 and Mutual (redirected toward the listener) in Experiment 2). The results showed that speaker gaze was used to form expectations about the unfolding sentence indicated by three observed ERP components that index different underlying mechanisms of language comprehension: an increased Phonological Mapping Negativity (PMN) was observed when an unexpected (Incongruent) or unpredictable (Uninformative) phoneme is encountered. The retrieval of a referent's semantics was indexed by an N400 effect in response to referents following both Incongruent and Uninformative gaze. Additionally, an increased P600 response was present only for preceding Incongruent gaze, indexing the revision process of the mental representation of the situation. The involvement of these mechanisms has been supported by the findings of the third experiment, in which linguistic content was presented to serve as a predictive cue for subsequent speaker gaze. In this experiment the sentence structure enabled participants to anticipate upcoming referents based on the preceding linguistic content. Thus, gaze cues preceding the mentioning of the referent could also be anticipated. The results showed the involvement of the same mechanisms as in the first two experiments on the referent itself, only when preceding gaze was absent. In the presence of object-directed gaze, while there were no longer significant effects on the referent itself, effects of semantic retrieval (N400) and integration with sentence meaning (P3b) were found on the gaze cue. Effects in the P3b (Gaze) and P600 (Referent) time-window further provided support for the presence of a mechanism of monitoring of the mental representation of the situation that subsumes the integration into that representation: A positive deflection was found whenever the communicative signal completed the mental representation such that an evaluation of that representation was possible. Taken together, the results provide support for the view that speaker gaze, in situated communication, is interpreted as part of the communicative signal and incrementally used to inform the mental representation of the situation simultaneously with the linguistic signal and that the mental representation is utilized to generate expectations about upcoming referents in an unfolding utterance.

Ausführliche Zusammenfassung

Diese Arbeit befasst sich mit der Frage, ob und in welchem Maße Zuhörer in situierten, kommunikativen Interaktionen den Blick des Sprechers nutzen, um Erwartungen über Nennungen in einem sich entfaltenden Satz zu formen und diese, gegebenenfalls, zu evaluieren.

Die in dieser Arbeit präsentierten Experimente mit Ereigniskorrelierten Potentialen (EKP) wurden mit dem Ziel entworfen, die zugrunde liegenden kognitiven Mechanismen, die in der Verwendung und Integration des Sprecherblicks involviert sind, zu untersuchen. Dem gehen in der Literatur beschriebene behaviorale Daten voraus, die darauf hindeuten, dass kommunikationsrelevante Sprecherblicke das Satzverständnis vereinfachen.

Die von den gefunden EKP-Komponenten indizierten Mechanismen wurden in Bezug auf zwei vorgeschlagene Ansätze untersucht, welche die berichteten behavioralen Daten erklären könnten:

Unter dem Prominenzansatz wird davon ausgegangen, dass verbessertes oder verschlechtertes Satzverständnis daher rührt, dass der Blick des Sprechers die Aufmerksamkeit des Zuhörers auf eine bestimmte Position oder ein bestimmtes Objekt lenkt. Dabei führt die Nennung von Objekten innerhalb des Fokus der Aufmerksamkeit des Hörers zu einer einfacheren Verarbeitung, wohingegen die Nennung eines Objektes außerhalb des Fokus dazu führt, dass zunächst die Aufmerksamkeit umgelenkt werden muss. Der durch dieses Umlenken herbeigeführte Aufwand führt dann zu erschwertem Satzverständnis.

Unter dem situativen Integrationsansatz wird, dem entgegengestellt, davon ausgegangen, dass die Verarbeitung des angeblickten Objekts über den reinen Anstieg seiner Prominenz hinaus geht. Dies beinhaltet sowohl semantisches Retrieval als auch die Integration der Bedeutung in eine mentale Repräsentation der Situation. Es wird davon ausgegangen, dass letzteres eine Erwartung dafür hervorruft, dass das betrachtete Objekt als nächstes genannt wird.

Um den Umfang der Verwendung des Sprecherblicks zu untersuchen, wurden drei Experimente entwickelt, in denen Teilnehmern eine Szene, bestehend aus drei unterschiedlich großen oder unterschiedlich hellen Objekten, präsentiert wurde. In der Mitte der Objekte wurde zudem ein stilisiertes Gesicht präsentiert, welches abgestimmt mit einem auditiv präsentierten Satz Augenbewegungen zu den Objekten hin durchführte. Die Sätze stellten Vergleiche zwischen den dargestellten Objekten dar. In den ersten beiden Experimenten waren Sätze wie im folgenden Beispiel strukturiert: "Verglichen mit dem Haus ist das Auto verhältnismäßig klein, denke ich." Der Blick zu dem jeweiligen Objekt ging der Nennung um 800ms voraus (Griffin and Bock, 2000). Während in diesem Satz der linguistische Kontext keine Grundlage bot um Vorhersagen über kommende Nennungen zu treffen, konnte der Sprecherblick potentiell im Voraus genutzt werden, um die folgende

Nennung zu antizipieren. Der Blick vor der Nennung des zweiten Objektes im Satz wurde so manipuliert, dass er entweder kongruent auf das nachfolgend genannten Objekt gerichtet wurde, inkongruent auf das Objekt, das im Satz nicht genannt wurde, oder uninformativ entweder zum unteren Rand des Bildschirms, wo kein Objekt abgebildet war (Experiment 1), oder zurück in die Mitte des Bildschirms, zum Zuhörer (Experiment 2).

Es wurden drei zugrundeliegende Mechanismen prognostiziert, die bei der Sprachverarbeitung involviert sind: a) ein phonologischer Abgleichmechanismus, der das wahrgenommene Signal mit der erwarteten Wortform abgleicht und im EKP-Signal durch die sogenannte Phonological Mapping Negativity (PMN) Komponente indiziert wird (Connolly and Phillips, 1994; Hagoort and Brown, 2000). b) ein semantischer Retrieval-Mechanismus, der den Abruf der entsprechenden semantische Bedeutung des wahrgenommenen Wortes ausführt, welcher von einer N400-Komponente indiziert wird (Van Berkum, 2009; Kutas and Federmeier, 2011). c) ein Integrationsmechanismus, der das wahrgenommene Signal inkrementell nutzt, um die mentale Repräsentation der Situation zu informieren. Dieser Mechanismus wird durch die P600-Komponente indiziert (Burkhardt, 2007; Brouwer et al., 2017).

Unter dem Prominenzansatz wurden in erster Linie Effekte des semantischen Retrieval-Mechanismus erwartet, da der Zugriff auf die semantische Bedeutung eines Wortes außerhalb des Fokus der Aufmerksamkeit erschwert ist. Gleichzeitig wurden keine Effekte des phonologischen Abgleichmechanismus und des Integrationsmechanismus erwartet, da beide Mechanismen voraussetzen, dass Erwartungen über den Verlauf des Satzes geformt wurden. Im Detail benötigt der phonetische Abgleichmechanismus die Antizipation einer bestimmte Wortform, wobei der Integrationsmechanismus nur dann gehemmt sein sollte, wenn die Integration der Wortbedeutung erschwert wird, da eine andere Bedeutung bereits in der mentalen Repräsentation vorhanden ist. Wenn der Sprecherblick folglich lediglich den Fokus der Aufmerksamkeit lenkt, würde dies nicht zu den für diese Mechanismen notwendigen Erwartungen führen.

Unter dem situativen Integrationsansatz hingegen wird angenommen, dass der Sprecherblick als Teil des kommunikativen Signals verstanden wird und semantisches Retrieval so wie Integration bereits vor der Nennung des angeblickten Objektes stattfinden. Darauf basierend wurden neben Effekten des semantischen Retrieval-Mechanismus auch Effekte des phonologischen Abgleichmechanismus sowie des Integrationsmechanismus erwartet. Dies bedeutet, dass die erwartete Wortform basierend auf dem Sprecherblick mit dem wahrgenommenen Signal abgeglichen wird und, wenn diese nicht übereinstimmen, zu einem Anstieg in der Reaktion der PMN-Komponente führen. Gleichermaßen wurde ein Anstieg in der Reaktion der P600-Komponente erwartet, wenn die zu integrierende Bedeutung nicht mit der bereits in der mentalen Repräsentation der Situation übereinstimmt und dadurch zu einer Revision der Repräsentation führt.

Die Ergebnisse der ersten beiden Experimente zeigen, dass der Sprecherblick genutzt wird, um Erwartungen an den bevorstehenden Referenten zu formulieren, was für den situativen Integrationsansatz spricht. Eine nur für kongruenten Blick abgeschwächte PMN-Reaktion vor dem N400-Zeitfenster deutet darauf hin, dass Erwartungen nicht nur auf konzeptioneller Ebene, sondern auch über die konkrete lexikalische Form gebildet werden, wenn ein einzelnes Objekt hervorgehoben wird. Ein N400-Effekt, wenn der Blick uninformativ oder irreführend ist, deutete auf erhöhte Kosten für das semantische Retrieval hin. Darüber hinaus wurde, wie unter dem situativen Integrationsansatz angenommen, eine erhöhte P600-Reaktion nur für die inkongruente Kondition gefunden, was mit der zuvor genannten Interpretation übereinstimmt. Das zweite Experiment liefert zusätzliche Unterstützung für die Unterscheidung der PMN und N400 als unterschiedliche Komponenten. Während der gegenseitige Blick, der in Experiment 2 als die uninformative Kondition diente, einen PMN-Effekt hervorrief, wurde kein signifikanter Unterschied im N400-Zeitfenster zwischen der uminformativen und kongruenten Kondition festgestellt. Diese Ergebnisse weisen auch auf ein starkes Zusammenspiel dieser beiden Komponenten hin. Die relativ kurzlebige N400 (300 – 450ms) nach der PMN zeigt an, dass das Retrieval der Wortbedeutung von dem phonologischen Abgleich profitiert, der durch die PMN indiziert wird. Diese Annahme ist in Übereinstimmung mit Funden von Hagoort and Brown (2000), wo der PMN der Ausdruck eines lexikalischen Auswahlprozesses zugeschrieben wurde. Es könnte argumentiert werden, dass einige Eigenschaften des Referenten in diesem früheren phonologischen Stadium abgerufen werden können, was den Mechanismus des semantischen Retrievals erleichtert. Daher kann man spekulieren, dass die Anwesenheit und Abwesenheit eines PMN-Effekts einen Einfluss auf die Stärke der N400 haben könnte. Dies erfordert jedoch weitere Untersuchungen.

Im Zeitfenster der P600 deuten die Ergebnisse darauf hin, dass die visuelle Szene und das Sprachsignal sowie der Blick des Sprechers verwendet werden, um eine mentale Repräsentation der Situation zu konstruieren. Dies ist im Einklang mit der Ansicht, dass Zuhörer den Sprecherblick so interpretieren, dass er Teil des kommunikativen Signals ist und Referenzabsichten ausdrückt (Staudte and Crocker, 2011).

In den hier vorgestellten Experimenten lieferte die erste Blickbewegung in jeder experimentellen Kondition einen korrekten Hinweis auf den bevorstehenden Referenten. Im Falle eines nachfolgenden inkongruenten Blicks wurden die Teilnehmer zu dem Glauben veranlasst, dass das betrachtete Objekt tatsächlich nachfolgend genannt würde, wodurch die verbleibenden Objekte in der Szene als wahrscheinliche Referenten eliminiert wurden. Die folgende Nennung wiederum zwang den Teilnehmer, die Repräsentation so zu revidieren, dass der gemeinte Referent nicht derjenige war, der angeschaut wurde, sondern der tatsächlich genannte. Diese grundlegende Aktualisierung der situativen Repräsentation spiegelt sich dann in einer P600-Modulation wider, welche die (Re-)Integrationsschwierigkeit darstellt. Ein solcher Unterschied wird in den uninformativen (nicht objektorientierten) Bedingungen nicht induziert, da beide möglichen Referenten — die beiden noch nicht genannten Objekte — gleichermaßen wahrscheinlich sind und daher keine Überarbeitung der Repräsentation erforderlich ist. Während die erhöhte Reaktion im N400-Zeitfenster sowohl für die inkongruenten als auch für die uninformativen Bedingungen unter dem Prominenzansatz wie auch dem situativen Integrationsansatz vorausgesagt wird, wurde die nur in der inkongruenten Bedingung beobachtete erhöhte P600-Reaktion allein durch den situativen Integrationsansatz vorhergesagt.

Während die ersten beiden Experimente untersuchten wie Sprecherblick verwendet wird, um linguistische Elemente des kommunikativen Signals zu antizipieren, wurde sich in Experiment 3 der Frage gewidmet, ob linguistische Elemente des kommunikativen Signals auch zur Antizipation von subsequenten Blickbewegungen genutzt werden und wie diese sich auf eine folgende Nennung des betrachteten Objekts auswirkt. Zu diesem Zweck wurden die den Teilnehmern präsentierten Sätze angepasst, während die visuellen Stimuli gleich blieben. Wie in dem Beispielsatz "Das Auto ist kleiner als das abgebildete Haus, denke ich" zu erkennen ist, können bereits nach dem Hören des Komparativs Erwartungen über den folgenden Referenten des Satzes geformt werden, respektive der drei unterschiedlich großen, dargestellten Objekte. Der Sprecherblick wurde so manipuliert, dass er entweder auf das nachfolgend genannte Objekt, oder das im Satz nicht erwähnte Objekt gerichtet wurde oder uninformativ in die Mitte des Bildschirms, zum Zuhörer.

Die Ergebnisse dieses Experiments zeigten, dass der Blick des Sprechers genutzt wurde, um Erwartungen basierend auf dem linguistischen Kontext zu validieren. Dies drückte sich durch einen N400 und P3b-Effekt in Reaktion auf den Sprecherblick aus, welcher der Nennung eines Referenten voraus ging. Eine PMN-Komponente wurde in Reaktion auf den Sprecherblick nicht gefunden oder erwartet, da kein phonologischer Abgleich stattfinden kann. Der N400-Effekt wurde erneut in Zusammenhang mit semantischem Retrieval interpretiert, wohingegen die P3b-Komponente in Relation mit der Integration in eine mentale Repräsentation der Situation interpretiert wurde (Donchin, 1981) und damit stark der vorherigen Interpretation der P600 ähnelt. Dies ist auch in Übereinstimmung mit der P600-as-P3 Hypothese (z.B. Sassenhagen et al. (2014)). Zusätzliche Unterstützung für diese Interpretation der P3b in der Region des Sprecherblicks fand sich auch in der nachfolgenden Region der Objektnennung. Nur wenn der Blick des Sprechers uninformativ war, waren ähnliche Effekte wie in den vorangegangenen Experimenten in Reaktion zur Nennung der Objekte (PMN, N400 und P600) vorhanden. Die Anwesenheit von Effekten in Reaktion auf die Objektnennung ausschlieSSlich wenn der Sprecherblick zuvor uninformativ war, lässt darauf schließen, dass der Blick des Sprechers tatsächlich als vollwertiges Element des kommunikativen Signals verstanden wird, so dass Erwartungen über Nennungen von visuellen sowie linguistischen Elementen gleichermaßen sowohl aufgebaut wie auch validiert werden. Wäre dies nicht der Fall, sollten unabhängig von den Effekten auf dem Sprecherblick auch Effekte auf der Nennung der Objekte gefunden werden.

Über diese Funde hinaus, wiesen die Ergebnisse von Experiment 3 auch auf eine weitere Funktion der P600-Komponente hin: In den ersten beiden Experimenten tauchte eine erhöhte P600-Reaktion nur in Verbindung mit widersprüchlichen Informationen auf, welche die Revision der mentalen Repräsentation der Situation erforderten. In Experiment 3 tauchte eine Positivierung immer dann auf, wenn das aktuelle Element des kommunikativen Signals die Repräsentation vervollständigte. Basierend auf der Satzstruktur, wie sie in Experiment 3 präsentiert wurde, führte die Identifizierung des zweiten Objektes auch zu einem evaluierbaren Zustand der mentalen Repräsentation der Situation; beide zu vergleichenden Objekte wie auch der Komparativ sind an diesem Punkt bekannt, wodurch Teilnehmer eine Interpretation des Satzes, wie auch eine Reaktion auf diese Information formulieren können (in diesem Fall die Entscheidung ob der Satz wahr oder falsch war). Dies lässt darauf schließen, dass die P600, sowie auch die P3b, nicht nur die Überarbeitung der mentalen Repräsentation indizieren, sondern auch deren Evaluierung. Dies legt die Interpretation eines Monitoring-Mechanismus nahe, was auch mit der LC/NE-P3 Interpretation (z.B. Sassenhagen et al. (2014)) übereinstimmt. Dieser Monitoring-Mechanismus überwacht dauerhaft den aktuellen Zustand der mentalen Repräsentation und führt zu dessen Auswertung sobald dies möglich ist. Dieser Auswertungsprozess drückt sich im Signal dann durch eine Positivierung aus. Basierend auf dieser Interpretation ist auch anzunehmen, dass die Evaluation der Repräsentation deren Überarbeitung subsumiert, so dass eine Positivierung in beiden Fällen zu erwarten ist.

Die experimentellen Ergebnisse dieser Dissertation präsentierte Evidenzen für eine gleichwertige Verwendung von Sprecherblick und linguistischem Inhalt in ihrer Rolle, die mentale Repräsentation einer Situation zu informieren. Dies ist Evidenz für den vorgeschlagenen situativen Integrationsansatz und zeigt auf, dass Sprecherblick, wie vorgeschlagen, als Teil des kommunikativen Signals verstanden wird. Darüber hinaus wurden Evidenzen für eine inkrementell aufgebaute, mentale Repräsentation der Situation, die dynamisch von unterschiedlichen Informationsquellen beeinflusst wird (wie zum Beispiel Sprecherblick und linguistischem Satzinhalt), herausgearbeitet.

Basierend auf den Ergebissen kann für einen Modellansatz argumentiert werden, in dem der bisherige Kontext in einer mentalen Repräsentation der Situation zu jedem Zeitpunkt abgebildet ist. Der aktuelle Zustand dieser Repräsentation wird verwendet um Erwartungen über kommende Elemente des kommunikativen Signals zu formen. Diese Erwartungen werden zum einen verwendet, um eine erweiterte Repräsentation zu formen, und gleichzeitig mit dem nächsten Element des kommunikativen Signals verglichen. Basierend auf der Art des Signals sind verschiedene kognitive Mechanismen in diesem Abgleich involviert: Im Falle eines auditiven Signals wird das erste Phonem des gehörten

Wortes mit dem (basierend auf den Erwartungen) antizipierten Phonem abgeglichen. Im Falle einer Übereinstimmung führt dies zu einer reduzierten PMN-Amplitude, wohingegen anderenfalls eine erhöhte PMN-Amplitude in Relation zum Informationsgehalt des Phonems auftritt. Dabei gelten erwartete Phoneme als wenig informativ, da sie keine neue/nicht erwartete Information beinhalten, wohingegen unerwartete (inkongruent) oder nicht erwartbare (uninformativ) Phoneme einen höherene Informationsgehalt haben, da sie für die Selektion des korrekten Referenten notwending sind. Der Output dieses Mechanismus wird dann an den semantischen Retrieval-Mechanismus weitergereicht. Dieser Mechanismus ist involviert, unabhängig von der Art des Elements (visuell oder linguistisch). Der Retrieval-Mechanismus gleicht den semantischen Inhalt des aktuellen Elements mit dem bereits basierend auf den Erwartungen antizipierten semantischen Inhalt ab, wobei eine Übereinstimmung wieder zu einer reduzierten Reaktion der N400 führt, wohingegen bei Nichtübereinstimmung der semantische Inhalt des neuen Elements abgerufen werden muss, was sich durch eine erhöhte N400-Reaktion darstellt. Schlussendlich überprüft der Monitoring-Mechanismus ob das neue Element des kommunikativen Signals die mentale Repräsentation vervollständigt, sodass die kommunizierte Botschaft verstanden werden kann. Sollte dies der Fall sein, drückt eine erhöhte Postivierung die Evaluation der Repräsentation aus. Sollte hingegen noch keine vollständige Repräsentation erreicht sein, wird das Element in die Repräsentation eingebettet. Dies führt dann zu einer erhöhten Positivierung, wenn die erweiterte Repräsentation überarbeitet werden muss, da das Element nicht den Erwartungen entsprach, was dann zu einer überarbeiteten Repräsentation führt. In beiden Fällen wird die erweiterte/überarbeitete Repräsentation der Situation verwendet, um neue Erwartungen für folgende Elemente zu generieren, die wiederum mit dem nächsten Element des kommunikativen Signals verglichen werden, bis eine vollständige Repräsentation erreicht ist.

Acknowledgment

First and foremost I would like to express my gratitude toward my supervisor Prof. Dr. Matthew Crocker, who provided me with the opportunity to pursue a PhD in his research group. I am thankful for your support and helpful feedback throughout both the experimental design and writing phase, but also for the trust in my abilities and the space you provided for personal growth and the opportunity to implement my own ideas.

Further, special thanks go to Dr. Maria Staudte, with whom I had the pleasure to work since my time as a master's student and who sparked my interest in experimental psycholinguistics. Thank you for supporting me over all those years.

Also, I want to thank Dr. Heiner Drenhaus, who provided valuable feedback and help throughout the past years and who always had an open door for discussing my experiments and ERPs. Thank you for helping me out in various occasions.

I also want to express my appreciation to all other current and past members of the Psycholinguistics Group who made the journey educational, fun and memorable. I especially want to name Yoana Vergilova with whom I not only shared a lab, but also joy and frustrations alike. Thank you for all the work(un)related conversations we shared.

Last but not least I want to thank my parents, Ilse and Rolf Jachmann, as well as my friends and family. While many things — good and bad — happened along the way, I could always count on your support. Vielen Dank für eure Unterstützung. I especially want to thank my dear friend Sina Michels, who had to listen to most of my complaints and frustrations, and who sometimes lost her temper in my my stead. Thank you for the therapeutic Fridays.

This work was supported by the Cluster of Excellence Multimodal Computing and Interaction and SFB/CRC 1102 Information Density and Linguistic Encoding (Project-ID 232722074), funded by the Deutsche Forschungsgemeinschaft (DFG).

The Immediate Influence of Speaker Gaze on Situated Speech Comprehension

Evidence from Multiple ERP Components

Torsten Kai Jachmann

Contents

Contents **xvii**

List of Figures **xxi**

List of Tables **xxv**

1 General Introduction **1**
- 1.1 Overview . . . 6
- 1.2 Publications . . . 6

2 The Role of Gaze in Face to Face Interactions **7**
- 2.1 Gaze as a Social Cue . . . 8
- 2.2 Gaze and Reflexive Orientation . . . 8
  - 2.2.1 Gaze in Comparison to other Attentional Cues . . . 9
- 2.3 Gaze and Language . . . 10

3 Background of Neurolinguistics **15**
- 3.1 Basics of Electroencephalography . . . 15
- 3.2 ERPs and Linguisitc Components . . . 16
  - 3.2.1 P300 . . . 17
  - 3.2.2 P600 . . . 20
  - 3.2.3 N400 . . . 22
  - 3.2.4 PMN (N200) . . . 25
- 3.3 Interpretations and Interplay of the N400 and P600 . . . 28
  - 3.3.1 Semantic Attraction . . . 28
  - 3.3.2 Monitoring Theory . . . 29
  - 3.3.3 Continued Combinatory Analysis . . . 29
  - 3.3.4 Extended Argument Dependency Model . . . 31
  - 3.3.5 Retrieval-Integration account . . . 32
- 3.4 The P600-as-P3 Hypothesis . . . 34

3.5 Component Overlap 35

4 Speaker Gaze as a Predictive Cue for Linguistic Content 37
4.1 Stimulus Materials 38
4.1.1 Pre-test 39
4.2 Experimental Design 41
4.3 Hypotheses 42
4.4 Experiment 1 48
4.4.1 Participants 49
4.4.2 Data Analysis and Results 50
4.4.3 Preliminary Discussion 63
4.5 Experiment 2 66
4.5.1 Participants 69
4.5.2 Data Analysis and Results 69
4.5.3 Preliminary Discussion 79
4.6 Discussion 81
4.6.1 PMN 81
4.6.2 N400 82
4.6.3 P600 83
4.6.4 Summary 84
4.6.5 Conclusion 85

5 Linguistic Content as a Predictive Cue for Speaker Gaze 87
5.1 Experimental Design 88
5.2 Hypotheses 91
5.3 Experiment 3 92
5.3.1 Participants 93
5.3.2 Data Analysis and Results 93
5.4 Discussion 105
5.4.1 PMN 106
5.4.2 N400 106
5.4.3 Positive Deflections 107
5.4.4 Summary 107
5.4.5 Conclusion 108

6 General Discussion 109
6.1 Phonological Matching Mechanism / PMN 112
6.2 Semantic Retrieval Mechanism / N400 114
6.2.1 Monitoring Mechanism / P3b/P600 115
6.2.2 Differences in the Positive Deflections 116

7 Conclusion **125**

Appendices **129**

A Additional Plots - Experiment 1 **131**

B Additional Plots - Experiment 2 **141**

C Additional Plots - Experiment 3 **149**

D Example of Model Reduction Process (PMN in Experiment 1) **161**

Bibliography **171**

List of Figures

4.1 Distribution of complexity ratings of the used objects 39
4.2 EXP1: Timeline of an experimental trial 43
4.3 EXP1: Example screen 49
4.4 EXP1: Difference waves in Noun region 52
4.5 EXP1: ERPs in the Noun region seperated by Experimental Conditions 53
4.6 EXP1: Difference waves of Gaze Informativity 60
4.7 EXP1: Difference waves of Gaze Directions 61
4.8 EXP1: ERPs in the Gaze region separated by Vertical and Horizontal gaze cues . 63
4.9 EXP1: ERPs in the Gaze region separated by Upward and Downward gaze cues 64
4.10 EXP2: Timeline of an experimental trial 67
4.11 EXP2: ERPs in Noun region seperated by Experimental Conditions 71
4.12 EXP2: Difference wave of Gaze informativity 78
4.13 EXP2: ERPs in Gaze region seperated by Gaze informativity 80

5.1 EXP3: Timeline of an experimental trial 90
5.2 EXP3: Difference waves in the Gaze region 96
5.3 EXP3: ERPs in the Gaze region separated by Gaze Presence 97
5.4 EXP3: ERPs in the Gaze region separated by the 4 experimental conditions . . . 99
5.5 EXP3: ERPs in Noun region separated by Expectability under Present Gaze . . . 100
5.6 EXP3: ERPs in Noun region separated by Expectability under Absent Gaze . . . 101
5.7 EXP3: ERPs spanning both ROIs, separated by the 4 experimental conditions . . 102

6.1 ROIs aligned by their appearance across the experiments 117
6.2 ROIs aligned by their role across the experiments 118
6.3 Outline of the proposed Model 121

A.1 Experiment 1 - Fixed Effects for the Positivity in the Gaze Region 132
A.2 Experiment 1 - Plot of Means for Horizontal and Vertical Gaze Direction in the Gaze Region 133

A.3 Experiment 1 - Plot of Means for Upward and Downward Gaze Direction in the Gaze Region . . . 133
A.4 Experiment 1 - Fixed Effects for the PMN Time-window in the Gaze Region . . . 134
A.5 Experiment 1 - Plot of Means for the PMN Time-window in the Noun region . . 134
A.6 Experiment 1 - Topographical Scalp Map for the PMN Time-window in the Noun Region . . . 135
A.7 Experiment 1 - Fixed Effects for the N400 Time-window in the Gaze Region . . . 136
A.8 Experiment 1 - Plot of Means for the N400 Time-window in the Noun region . . 136
A.9 Experiment 1 - Topographical Scalp Map for the N400 Time-window in the Noun Region . . . 137
A.10 Experiment 1 - Fixed Effects for the P600 Time-window in the Gaze Region . . . 138
A.11 Experiment 1 - Plot of Means for the P600 Time-window in the Noun region . . 138
A.12 Experiment 1 - Topographical Scalp Map for the P600 Time-window in the Noun Region . . . 139

B.1 Experiment 2 - Fixed Effects for the PMN Time-window in the Gaze Region . . . 142
B.2 Experiment 2 - Plot of Means for the PMN Time-window in the Noun region . . 142
B.3 Experiment 2 - Topographical Scalp Map for the PMN Time-window in the Noun Region . . . 143
B.4 Experiment 2 - Fixed Effects for the N400 Time-window in the Gaze Region . . . 144
B.5 Experiment 2 - Plot of Means for the N400 Time-window in the Noun region . . 144
B.6 Experiment 2 - Topographical Scalp Map for the N400 Time-window in the Noun Region . . . 145
B.7 Experiment 2 - Fixed Effects for the P600 Time-window in the Gaze Region . . . 146
B.8 Experiment 2 - Plot of Means for the P600 Time-window in the Noun region . . 146
B.9 Experiment 2 - Topographical Scalp Map for the P600 Time-window in the Noun Region . . . 147

C.1 Experiment 3 - Fixed Effects for the N400 time-window in the Gaze Region . . . 150
C.2 Experiment 3 - Plot of Means for the N400 Time-window in the Gaze Region . . 150
C.3 Experiment 3 - Topographical Scalp Map for the N400 Time-window in the Gaze Region . . . 151
C.4 Experiment 3 - Fixed Effects for the P300 time-window in the Gaze Region . . . 152
C.5 Experiment 3 - Plot of Means for the P300 Time-window in the Gaze Region . . 152
C.6 Experiment 3 - Topographical Scalp Map for the P300 Time-window in the Gaze Region . . . 153
C.7 Experiment 3 - Fixed Effects for the PMN time-window in the Noun Region . . . 154
C.8 Experiment 3 - Plot of Means for the PMN Time-window in the Noun Region . . 154
C.9 Experiment 3 - Topographical Scalp Map for the PMN Time-window in the Noun Region . . . 155

C.10 Experiment 3 - Fixed Effects for the N400 time-window in the Noun Region . . . 156
C.11 Experiment 3 - Plot of Means for the N400 Time-window in the Noun Region . . 156
C.12 Experiment 3 - Topographical Scalp Map for the N400 Time-window in the Noun Region 157
C.13 Experiment 3 - Fixed Effects for the P600 time-window in the Noun Region . . . 158
C.14 Experiment 3 - Plot of Means for the P600 Time-window in the Noun Region . . 158
C.15 Experiment 3 - Topographical Scalp Map for the P600 Time-window in the Noun Region 159

List of Tables

4.1 Summary of the objects presented in the experiment 40
4.2 Summary of the ratings for naturalness for the two different voices 41
4.3 Summary of the assumed mechanisms and the corresponding indexing components 45
4.4 Summary of the predicted effects for the two proposed accounts 47
4.5 EXP1: Summary of the number of eye-movements in the noun region per condition 51
4.6 EXP1: Contrasts for Factor Condition 53
4.7 EXP1: Contrasts for Factor Longitude 54
4.8 EXP1: LMER output for PMN time-window 55
4.9 EXP1: LMER output for N400 time-window 56
4.10 EXP1: LMER output for 250-350ms time-window 57
4.11 EXP1: LMER output for P600 time-window 58
4.12 EXP1: Summary of the model results 58
4.13 EXP1: Contrasts for Factor Gaze Informativity 59
4.14 EXP1: Contrasts for Factor Gaze Direction 61
4.15 EXP1: LMER output Gaze Direction 62
4.16 EXP2: Summary of the adjusted objects presented in the experiment 66
4.17 EXP2: Summary of the distribution of gaze patterns 68
4.18 EXP2: Summary of the number of eye-movements in the Noun region 70
4.19 EXP2: Contrasts for Factor Condition in the Control comparison 71
4.20 EXP2: LMER output for PMN time-window 72
4.21 EXP2: LMER output for PMN time-window (control) 73
4.22 EXP2: LMER output for N400 time-window 74
4.23 EXP3: LMER output for N400 time-window (control) 75
4.24 EXP2: LMER output for P600 time-window 76
4.25 EXP2: LMER output for P600 time-window (control) 77
4.26 EXP2: Summary of the model results 77
4.27 EXP2: Contrasts for Factor Informativity 77

4.28 EXP2: LMER output for 300 – 500ms time-window . . . 79

5.1 EXP3: Summary of the predicted effects for the two proposed accounts . . . 92
5.2 EXP3: Contrasts for Condition Factor . . . 95
5.3 EXP3: LMER output for P300 time-window in Gaze region . . . 97
5.4 EXP3: LMER output for N400 time-window in the Gaze region . . . 99
5.5 EXP3: LMER output for PMN time-window in the Noun region . . . 103
5.6 EXP3: LMER output for N400 time-window in the Noun region . . . 104
5.7 EXP3: LMER output for P600 time-window in the Noun region . . . 105

6.1 Summary of the Effects for Experiments 1 and 2 . . . 110
6.2 Summary of the Effects for Experiment 3 . . . 111

A.1 Contrasts for Factor Gaze Direction . . . 131
A.2 Contrasts for Factor Longitude . . . 131
A.3 Contrasts for Factor Condition . . . 131

B.1 Contrasts for Factor Longitude . . . 141
B.2 Contrasts for Factor Condition . . . 141
B.3 Contrasts for Factor Condition in the Control Comparison . . . 141

C.1 Contrasts for Condition Factor . . . 149
C.2 Contrasts for Factor Longitude . . . 149

CHAPTER 1

General Introduction

In face-to-face spoken interactions, interlocutors are not only exposed to the linguistic signal but also to contextual information provided by their surroundings, as well as language-external cues that can be used to enhance language comprehension. Especially cues that are provided by the interlocutors can potentially enrich and facilitate comprehension, as for example gestures, facial expressions, or gaze. As a cue that is usually always accessible in such interactions, this thesis will focus on speaker gaze as an potentially informative cue for listeners' sentence comprehension.

While many studies on verbal language comprehension haven been conducted over the years, those studies mainly focused on how the linguistic signal is processed. Importantly, it has been shown that language comprehension is an incremental process that utilizes the incoming signal word by word (e.g., Tanenhaus et al. (1995)) and that listeners anticipate incoming sentential content (e.g., Altmann and Kamide (1999)). In addition, however, language external behaviour that is still related to the linguistic content of an utterance, such as gaze, has been shown to reveal underlying processes involved in language production as well as comprehension. In face-to-face communication, gaze therefore may serve as an important informative cue that can be considered to be part of the communicated signal.

As an example of a situated face-to-face interaction, one could imagine an everyday situation such as a breakfast scenario in which the table is set. When the speaker's gaze falls onto a mug while saying *'Could you hand me ...'* already at this point the listener may anticipate the mug being the desired object. In the very same scenario, however, a contextually also valid continuation might be *'... the plate.'* Such a situation, in which the continuation — even though being valid given the discourse — is not supported by visual cues, has been shown to lead to comprehension difficulties (Staudte and Crocker, 2011). While some studies have shown that speaker gaze to objects in a co-present scene can influence listeners' sentence comprehension (e.g., Hanna and Brennan (2007)), the precise source of the observed difficulties and facilitatory effects remains unclear. One possible explanation for the observed effects could be that the gaze of the speaker simply draws the listeners attention to the mug and, hence, away from other possible referents in the scene. Alternatively, listeners may use their prior experience that speakers often fixate a referent just before naming it (Griffin and Bock, 2000), so that the mug could already be processed as the anticipated upcoming referent. Hence, the naming of an object other than the mug may lead to the necessity to adjust the believes about the speaker's intended message.

These two possibilities can be viewed as different accounts, equally explaining the behavioral findings from previous research. The possibility that the facilitatory and inhibitory effects stem from the allocation of focus to the gazed-at object, so that the naming of an object other than the attended-to object requires the reallocation of the focus toward the actual referent, will be referred to as the Prominence Account (PA). In other words, the PA assumes that the linguistic signal and gaze are fully independent so that the allocation of attention to an object or position does not interfere with the processing of the linguistic signal up to the point where the linguistic signal itself requires an allocation of focus to the mentioned object. If the mentioned object is not in the vicinity of the current focus, inhibitory effects are explained as costs of the reallocation of focus. The alternative explanation assuming that speaker-gaze is interpreted to reveal the referential intention of the speaker even before a referent is named will be referred to as the Situational Integration Account (SIA). Under the SIA, speaker-gaze is taken as part of the incremental comprehension system to anticipate the content and meaning of the unfolding utterance. This would mean that listeners could form expectations about the upcoming referents on a word-form level, retrieve the semantic information of the anticipated referent and already integrate the referent into a mental representation of the situation.

As behavioral data have not yet offered clear evidence for one account over the other, to answer the question to what extent gaze is actually utilized by interlocutors, it is important to investigate the underlying cognitive mechanisms in the perception and processing of gaze.

Considering the implications of both accounts, it is reasonable to assume the involvement of a semantic retrieval mechanism that benefits from a gaze cue toward an object before its mentioning: Either because of the increased prominence of the gazed-at object (PA), or because the semantic meaning of the gazed-at object is already retrieved (SIA). Also, both accounts predict an inhibition of the retrieval mechanism if the preceding gaze was directed toward another object than the subsequently named object: Either because the named object lies outside the focus of attention and possibly even requires a reallocation of focus (PA), or because the wrong referent was anticipated, requiring the retrieval of the meaning of the actual referent. In the event that gaze does indeed initiate a deeper processing of the cued object (SIA), we might expect other aspects and mechanisms of processing to be engaged: Firstly, if interlocutors in fact form expectations about the upcoming word-form, a phonological matching mechanism could utilize the information provided by the first phoneme of a referent to match this input with the expected phoneme of the anticipated word-form. If the phoneme matches the expectation (the gazed-at object matches the named object), the mechanism successfully matches the perceived phoneme with the anticipated word-form. If, however, the perceived phoneme does not match the expected phoneme, the mechanism is inhibited. Secondly, the SIA assumes a mechanism that maintains and updates a mental representation of the unfolding situation. If a gazed-at object is immediately integrated into this assumed mental representation, the mentioning of an object other than the gazed-at object should lead to the necessity to revise the mental representation.

By investigating the influence of gaze on established indices of these underlying aspects of linguistic processing, we can better understand the status of gaze in the communication signal, as reflected by the two accounts described above. To assess the cognitive mechanisms involved in the integration of speaker-gaze with speech, this thesis presents a series of Event-Related Potential (ERP) experiments. Certain components in the ERP signal have been demonstrated to index the previously named mechanisms: The PMN component was identified to index phonological matching mechanisms, the retrieval mechanism has been found to be indexed by the N400 component, while the P600 component has been found in the context of integration and updating. The underlying accounts connecting the mentioned mechanisms with the ERP components will be discussed in more detail in Chapter 3.

With respect to the outlined accounts, therefore, for the PA, only an N400 effect indexing the retrieval difficulties of a referent outside of the focus of visual attention would be expected. In terms of the breakfast scenario, this would be the case when the *'plate'* was mentioned, although the preceding gaze fell on the *'mug'*, drawing the listeners attention. As the focus would have to be redirected toward the *'plate'*, retrieval of the word could reasonably be inhibited. The SIA, however, predicts a PMN effect indexing the matching of the perceived phoneme with the anticipated first phoneme of the incoming referent followed by an N400 effect indexing the retrieval difficulty of an unexpected referent that differs from the possibly already retrieved gazed-at object and, finally, a P600 effect indexing the necessity to update and adjust the mental representation of the situation. Again, this would be the case in the same situation as described above, however, with different reasoning: As the SIA assumes the speaker-gaze to be part of the conveyed message, the gaze toward the *'mug'* would not only draw the listeners attention toward the object, but would further possibly be interpreted as the upcoming referent in the speaker's utterance. As such, the listener is assumed to anticipate the *'mug'* as the upcoming referent. If however the perceived initial phoneme of the actual referent 'plate' does not fit the anticipated initial phoneme of the word *'mug'*, a PMN effect is reasonably expected. Further, the *'mug'*, as the anticipated referent would, under the SIA, possibly already be retrieved. Hence, the actual naming of the *'plate'* would require the listener to retrieve the word from semantic memory, which should be expressed by an N400 effect. Lastly, as the 'mug' was anticipated as the next referent, it could already be integrated into a mental representation of the situation. The necessity to update this mental representation with the actual referent *'plate'*, then could reasonably be expected to be expressed by a P600 effect. The outlined hypotheses can be found in Chapter 4.3.

These predictions are evaluated in experiments 1 and 2, described in Chapter 4. The results of these experiments strongly suggest a processing of the gazed-at object beyond the increase of its prominence in the scene, supporting the SIA.

While the support of the SIA already provides an interesting addition to the understanding of the integration of gaze with language, it does not exhaust the question about the extent of gaze utilization. Specifically, a case in which the linguistic context creates an expectation for a specific referent and, thus, gaze to that referent prior to its naming could possibly be utilized to evaluate these referential expectations and, eventually, the intended message of the speaker. Recalling the previously described breakfast scenario, the first two experiments investigated the influence of gaze when it was presented in a way so that it can be utilized to anticipate the continuation of the sentence: Observers can use the gaze of the speaker to form expectations about the upcoming linguistic content of the sentence and, hence, accelerate comprehension. It is, however, also possible to imagine that gaze could inversely be utilized instead to evaluate expectations formed on linguistic (and situational) context, so that a gaze cue fitting the linguistic expectations could be perceived differently from such a gaze cue that does not fit the expectations. For example, while having dinner, one could be told to *'please empty ...'*, while the linguistic context will already limit the expectations of the upcoming sentential content to objects that can be emptied, the situation could potentially limit the expectations further. It would be reasonable to expect a continuation such as *'... your plate'*. However, if the gaze of the speaker falls instead on a dishwasher in the kitchen, these expectations could possibly be overwritten by *'... the dishwasher later'*. While both are linguistically plausible sentences, the situational and linguistic context most likely renders a related utterance more expected than an unrelated utterance. The utilization of speaker gaze, however, could potentially early on help to adjust those expectations and, hence, ease sentence comprehension.

In order to more broadly address the question about the extent of gaze utilization, a following experiment investigated whether gaze is also utilized to evaluate expectations formed on the base of the linguistic content of a sentence. To approach this question, the experimental setup used in the previous experiments was adjusted, so that the linguistic context of the presented sentences provided grounds for expectations about the upcoming referent. Hence, both the referent as well as the gaze cue preceding its naming could be used to evaluate the linguistic expectation. Therefore, unlike the previous experiments where effects of the speaker's gaze on the listeners comprehension was only expected on the referent following the gaze cue, this question offers two possible affected regions: namely the gaze cue as well as the referent. If gaze is utilized to evaluate expectations early on, it is reasonable to assume that the processing of linguistic content following the gaze cue is facilitated. In contrast, if the evaluation of linguistic expectations does not utilize external cues and is solely dependent on linguistic evaluation, gaze could ultimately be ignored. As for the previous question, the evaluation of expectations is expected to incorporate mechanisms of phonological matching (for the linguistic signal only), a semantic retrieval mechanism, as well as a mechanism of integration with a mental representation of the situation.

Assuming that gaze is utilized to evaluate referential expectations formed on the base of linguistic content, the mechanisms of semantic retrieval as well as integration with the mental representation should be involved in the processing of the gaze cue. This is based on the earlier findings indicating a deeper processing of gazed-at objects. In turn, the N400 and P600 component indexing the involvement of the assumed mechanisms should consequently be reduced on the subsequent linguistic signal when the referent matches the gazed-at object. In terms of the dinner scenario above, this would be the case when the gaze of the speaker fell on the *'dishwasher'* after uttering *'please empty the ...'*, the listener could already adjust the situational expectations for the continuation to be *'plate'*. Hence, even before hearing *'dishwasher'*, the listener could retrieve the meaning of 'dishwasher', reasonably evoking an N400 effect, and update the mental representation of the situation from *'plate'* to *'dishwasher'*, expressed by a P600 effect. In contrast, if gaze was not utilized to evaluate expectations, the anticipated mechanisms should solely be involved in the processing of the linguistic signal, resembling the findings from the previous experiments. Additionally, in the absence of speaker gaze to an specific object, the evaluation of the referential expectations should also be found on the referent. This would be the case if, in the dinner scenario, the speaker would not look at any specific object in the scene, but, instead, at the listener for example. In this case, the listener would possibly, driven by the situational context, assume the sentence to be continued with *'plate'* as the referent after hearing *'please empty the ...'*. Only when hearing the actual referent *'dishwasher'*, the listener would be able to evaluate the referential expectations, leading to the same effects as found in the first two experiments. In more detail, a PMN effect would be expected in response to the mismatching initial phoneme of the perceived *'dishwasher'* compared to the expected first phoneme of *'plate'*; An N400 effect would be expected to be indexing the retrieval of the word *'dishwasher'* as it was contextually unexpected; And, finally, a P600 effect would be expected as the mental representation of the situation would need to be updated. The same reasoning would hold in the event that gaze was not used to evaluate linguistic expectations. Experiment 3 (Chapter 5) was conducted to asses the involvement of the expected mechanisms in the respective regions (gaze cue and subsequent referent). The results of the experiment strongly suggest that gaze not only is used to evaluate expectations, but that it even was used as a substitute for the linguistic signal, at least to some extent.

Taken together, the results of the presented experiments provide evidence from a range of ERP components that gaze is used to anticipate as well as evaluate linguistic expectations. This supports a view in which speech-related visual cues, that are time locked to the production of referential expressions, should be considered a part of the communication signal, and are processed accordingly. As such, it can further be said that models of language comprehension focusing mostly, if not entirely on the verbal aspect of communication need to be extended to account also for language external cues that enhance communication and are utilized as part of the incrementally utilized stream of input to anticipate the continuation of a sentence or even conversation. This will be discussed in greater detail in Chapter 6 on page 109.

1.1 Overview

Chapter 2 will review the literature on the utilization, perception and influence of gaze cues. Chapter 3 subsequently will give a brief introduction into the ERP method and the components that are relevant to the presented research. The experiments presented in Chapter 4 will investigate the effect of gaze when it can be used to form linguistic expectations (breakfast scenario above). The experiment presented in Chapter 5 investigated the impact of gaze when it could be utilized to evaluate linguistic expectations instead of being used to form them (dinner scenario above). Chapter 6 will then summarize and discuss the overall results of all experiments. Finally, conclusions are drawn in Chapter 7.

1.2 Publications

This thesis is in part adapted from a peer-reviewed journal publication (Jachmann et al., 2019):

- Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2019). Influence of speakers gaze on situated language comprehension: Evidence from Event-Related Potentials. Brain and cognition, 135, 103571.

CHAPTER 2

The Role of Gaze in Face to Face Interactions

As an easily accessible cue, to which people in face-to-face interactions are naturally exposed, it is not surprising that gaze has been investigated by various researchers. It has been shown that gaze plays a multi-faceted role in situated communication: studies provided evidence that listeners fixate mentioned or anticipated referents (e.g., Emery (2000); Flom et al. (2007)), as well as that speakers gaze at objects they are about to mention (e.g., Griffin and Bock (2000); Kreysa (2009)), and that gaze can express joint attention of interlocutors (e.g., Thomsen (1974); Tomasello et al. (1998)). Following those findings, it seems to be clear and reasonable that gaze, utilized by speakers as well as listeners, can facilitate interactions between interlocutors. However, the research on gaze and, more specifically, how and to what extent it is utilized, is not yet exhausted. In the underlying research presented in this dissertation, I aimed to investigate the effect of speaker-gaze on listeners' sentence comprehension. More precisely, the conducted experiments tried to shed light on the underlying mechanisms that are involved in the perception of speech-aligned gaze by listeners and to what extent such speaker-gaze is utilized as a cue that potentially influences downstream sentence comprehension. Additionally, I investigated how speaker-gaze is used and integrated when it can be reversibly used not to form expectations about the unfolding sentence, but to validate expectations build on the linguistic context. Further information on the exact scope and aim of the different experiments will be explained in Part 4.4.3.2.

The current chapter reviews the findings of research on gaze in cognitive and communicative context before discussing the relation of language and gaze in more depth. Section 2.1 will introduce the social aspect of gaze, while Section 2.2 will focus on the broader topic of reflexive reactions shown by people when they were exposed to gaze cues.

The following section 2.3 will focus more strongly on the immediate interplay of gaze and language.

2.1 Gaze as a Social Cue

The importance of gaze as a social cue is evident not only in humans but also in animals (Emery, 2000). As such, it can be said that gaze, also from an evolutionary perspective, holds a profound importance. Studies have provided evidence that even dogs utilize gaze following (Miklösi et al., 1998; Hare and Tomasello, 1999) and that with increasing cognitive capabilities, the function and utilization of gaze further increases. It was, for example, shown, that monkeys and apes further utilize gaze to establish joint attention (Thomsen, 1974; Tomasello et al., 1998). Also in human development, an increasing functionality of gaze can be witnessed. Eyes are among the first points of contact between mother and child (Haith et al., 1977). Various researchers have collected evidence to establish a 'time-line' of how and at what age gaze can be utilized by infants (e.g., Baron-Cohen (1997); Butterworth and Jarrett (1991); Mundy and Gomes (1998); Symons et al. (1998)). With as early an age as 3 months, infants are capable to recognize eye presence. Starting around 9 months, the utilization of simple gaze can be found. With 12 months, joint attention can be established and after 18 months gaze following behaviour. Eventually, starting around 48 months, children are already capable of assigning mental attribution to gaze.

A study by Becchio et al. (2008) provided evidence that, in adults, the gaze to an object can even influence the perception of that object. In their functional magnetic resonance imaging (fMRI) study, they presented participants with scenes in which a human either physically interacted with an object (e.g., grasping) or simply gazed at an object. In the latter condition, they found activations in brain regions that are commonly related to an involvement in hand-object interactions. This attributed graspability of the object was reported to even persist when the gaze to the object and even the observer where no longer present. Other studies further showed that the gaze to an object increases that objects likeability (Bayliss et al., 2006) or can decrease it when paired with according facial expressions (Bayliss et al., 2007). While these studies attested for the communicative attributes of gaze, they don't answer the question whether gaze is utilized volitional or reflexive.

2.2 Gaze and Reflexive Orientation

Although the emphasis of this thesis is situated gaze in face-to-face spoken interactions, its role as an informative cue in non-linguistic communication, as described above, as well as the deliberateness to which gaze is utilized are important factors. When investigating the influence of gaze on language comprehension, it is also necessary to investigate to what extent the information provided by gaze is utilized when a potentially more informative signal is present. A mainly volitional utilization could simply be overwritten or ignored by observers when presented with a linguistic signal, while a reflexive utilization would lead to a second stream of input that unavoidably requires its integration with other sources of information, such as language.

While it is sometimes argued that gaze leads to volitional, goal-based — sometimes also called endogenous — shifts of attention (Vecera and Rizzo, 2006), it is broadly excepted that gaze, at least initially, leads to reflexive — sometimes also called exogenous — attention or orientation shifts. Various studies describe the rapid cuing effect of gaze, even for misleading cues (Driver IV et al., 1999; Friesen and Kingstone, 1998; Langton and Bruce, 1999). Additionally, there is some evidence that these reflexive shifts are resistant to volitional control, which was shown in situations where participants were — or became — aware of the unreliability of the cue (Driver IV et al., 1999; Friesen and Kingstone, 1998; Friesen et al., 2004).

A study by Uono et al. (2014) showed that gaze, as well as arrows, toward an at the time empty position facilitates reaction times for targets that appear in the cued position, and, inversely, inhibited reactions to targets appearing in non-cued positions. While the previously mentioned study utilized pictures of actual human faces, a study by Friesen and Kingstone (1998) additionally showed that reflexive orientation is also induced by stylized faces. This is beneficial for experiments investigating the influence of gaze as stylized faces are easier to manipulate and integrate in experimental setups than actual human faces. However, both of the aforementioned studies used behavioral measures (reaction times) in their studies, which does not allow for drawing conclusions about underlying mechanisms.

While the reviewed work shows the — at least to some extent — reflexive nature of the utilization of gaze cues, it also possible that any directional cue (e.g., arrows) might serve a similar purpose. Therefore, whether and how gaze cues differ from other cues, such as arrows, has been a topic of considerable interest.

2.2.1 Gaze in Comparison to other Attentional Cues

Hietanen et al. (2008), in addition to response times, collected Event-Related Potential (ERP) data in order to compare the perception of gaze and arrow cues. They report differences in the responses to the both cue types, so that arrow cues elicited early directing attention negativities (EDAN) over parietal cites that were not found for gaze cues. They interpret their findings in support of the claim that the attentional following of arrows is of a more volitional nature compared to a more reflexive attention shift induced by gaze.

While Hietanen et al. (2008) provided evidence for a difference between gaze and arrow cues, whether the perception and utilization of the two types of cues actually differs is still a widely debated topic. The inherently reflexive orientation following gaze cues was named to be an indicator for the difference between gaze and arrow cues. For the latter, it was often assumed that volitional processes could overwrite the reflexive processes, which was reported to be hindered for gaze cues (e.g., Driver IV et al. (1999)). Other studies, however provided evidence for a similar reflexive behaviour also found for arrow cues (Hommel et al., 2001; Ristic et al., 2002; Tipples, 2008). It was therefore assumed that for both, gaze and arrow cues, attention allocation is mediated by an interaction of involuntary and voluntary processes (Folk and Remington, 1998; Tipples, 2008). It was, however, also proposed that while gaze and arrow cues induce similar behavioural effects, the underlying mechanisms differ (Ristic et al., 2002). Studies on patients with brain lesions in the superior temporal gyrus (STG) (Akiyama et al., 2006) showed that, while an orienting according to arrow cues was still possible, orienting based on gaze cues was not. Similarly, it was also shown that, while arrow cues induced reflexive orienting in both hemispheres of a split-brain patient, misleading gaze cues produce reflexive orienting only in the face-processing hemisphere (Friesen and Kingstone, 1998; Kingstone et al., 2000).

Additional to the reviewed literature, it is also worth noting that biological cues such as gaze, unlike arrows, are more easily accessible in face-to-face interactions. This underlines the importance of gaze as a communicative cue.

2.3 Gaze and Language

The so far reviewed works provided evidence for the non-verbal communicative value of gaze as well as the — at least partial — reflexive nature of its utilization. This section will review studies investigating the effect of gaze in the context of verbal communication, which form the starting point for the questions addressed in this thesis.

Many researches have investigated whether and to what extent gaze — from both speakers and listeners — influences or facilitates language comprehension (e.g., Tanenhaus et al. (1995); Meyer et al. (1998); Griffin and Bock (2000); Knoeferle et al. (2005)). Interlocutors in face-to-face interactions have been found to pay close attention to each other's gaze. Besides looking at each other, interlocutors also direct their gaze toward objects under discussion (Argyle and Cook, 1976). It has also been reported that listeners tend to gaze at the speaker more that the speaker at the listener (Kendon, 1967), but also that speakers monitor listeners gaze, such that a perceived absence of the listener's visual attention even leads to interruptions and restarted utterances in speakers (Goodwin, 1981). While the gaze toward an object can potentially signal various processes (e.g., a first time processing, the search for a particular target, the preparation of a motor action, etc.), it usually expresses the focus of the speaker's attention and, in linguistic interaction, can very well express the preparation of a referring expression or allocate relevance for the current discourse to the object (Griffin and Bock, 2000; Hanna and Brennan, 2007). As such, it has been stated that gaze as a visual cue expresses the speaker's focus of visual attention and may draw the listener's attention as well (Emery, 2000; Flom et al., 2007).

Previous eye-tracking studies have shown that, if visual context is provided, speakers are orienting their gaze toward an object about 800 – 1000ms before mentioning it (Meyer et al., 1998; Griffin and Bock, 2000; Kreysa, 2009). It was also found, on the other hand, that listeners direct their gaze to objects around 200 – 400ms after hearing the onset of a referent (Tanenhaus et al., 1995; Allopenna et al., 1998).

It has been stated, that the visual information provided by speaker gaze can be used by the listener to ground and disambiguate referring expressions and allow the listener to infer the speakers intentions and goals, which can facilitate comprehension (Hanna and Brennan, 2007). In their study, they presented participants with an array of objects — colored shapes with dots — while a partner identified one of the objects. While the interlocutors where unable to see each other's array of objects, both arrays per trial contained the same objects. It was shown that, when a competitor for a target was present, the speakers gaze was used to disambiguate the referring expression to identify the target quicker. In an extended version of the experiment, an additional condition was introduced in which the arrays of the speaker and listener where reversed. Hence, a gaze to the left in that condition indicated a target to the right of the listener and vice versa. An inhibited reaction time in this condition indicated that listeners reflexively followed the gaze of the speaker and, hence, where hindered in identifying the actual target. Similar automated orienting properties have also been reported by Friesen and Kingstone (1998) to the extent that gaze was followed even if it was uninformative or misleading. However, Hanna and Brennan (2007) also observed that, in the course of the experiment, participants were able to remap the misleading gaze cue over time to some extent indicating a somewhat flexible utilization of speaker gaze that goes beyond reflexive attention allocation. Inversely, it has also been reported that speakers use listeners' gaze to affirm the understanding of the listener and, if necessary, adjust or repeat their utterance (Clark and Krych, 2004).

Eye-tracking studies with gaze cues aligned in time to auditory presented utterances have shown that listeners use gaze to predict a referent in the same order of occurrence (Staudte and Crocker, 2010, 2011; Staudte et al., 2014). Listeners presented with gaze cues toward the named objects in reversed order of mention were slower to validate the sentence than in neutral and original conditions. Also the behavior of the gazer before the utilization of a gaze cue has influence on the gaze following behavior (Meltzoff et al., 2010; Böckler et al., 2011). A mutual gaze preceding the gaze toward an object leads to a higher salience of the cued position. This was revealed by a lower response time in this condition compared to a preceding averted gaze pattern. In all presented experiments, the initial gaze of the stylized face is therefore a straight gaze toward the listener.

Staudte and Crocker (2011) provided evidence that speaker gaze is interpreted by listeners as conveying referential intentions. In their study, participants were presented with videos of a robot performing gaze cues toward objects time-aligned to sentences that compared those objects with one-another, e.g. *'The cylinder is taller than the pyramid that is* *pink.'* The target object additionally had a competitor that was a same type object with different size and color (e.g.: two pyramids were present in the visual scene). Thereby, the linguistic point of disambiguation (LPoD) was not sooner than in the adjective. The gaze cues preceded the naming of the object by 1000ms providing an early visual point of disambiguation (VPoD). The results showed that participants used that early VPoD to disambiguate the sentence as soon as the gaze cue was provided, expressed by a higher inspection rate of the gazed at object compared to the competitor. Furthermore, a misleading gaze cue led to an elevated reaction time while judging whether the heard sentence was true or false given the visual scene. In a following study, Staudte et al. (2014) used a more human-like virtual agent. The experimental setup was similar to the setup of their previous experiment with some changes. (1) In this study, all objects in the visual scene were unique, which shifted the LPoD to the noun region. (2) The order of the gaze cues was manipulated: In a Congruent condition, the gaze was directed toward the objects in the same order as they were named in the sentence, a Neutral condition did not provide any gaze cues, and a Reversed condition provided gaze cues toward the objects named in the sentence, but in reversed order (e.g.: the gaze cue preceding the first noun was directed toward the noun named second and vice versa). The inspection rate of the objects showed that participants initially follow the early visual cue and that the Reversed condition seems to disrupt the understanding of the sentence. This is additionally expressed by the response times when judging whether the sentences was true or false. Participants needed significantly more time to give a response in the Reversed condition compared to the Congruent condition. The Neutral condition elicited an intermediate response time.

In sum, the reviewed studies strongly hint toward an utilization of gaze by both speakers and listeners to facilitate comprehension and enhance the often ambiguous linguistic signal so that a common ground can be established and ambiguity can be resolved quickly.

Further, although it is not entirely clear if or to what extent gaze cues differ from non-biological cues, such as arrows, it can be said that gaze — from a social as well as conversational perspective — holds a special status. While symbolic meaning, such as the meaning of arrows, has to be learned, gaze is used by infants already at an early age and acquired naturally. Also in face-to-face conversations, gaze cues are more easily and naturally accessible than symbolic cues and do not necessarily require the volition of the speaker to reveal (referential) intention.

While the discussed results concerning gaze cues in situated language comprehension are already intriguing, the question remains precisely how gaze influences the underlying mechanisms of comprehension. While it is possible that the reallocation of attention alone already could explain some of the reported results (prominence account), it is also possible that listeners utilize gaze beyond an increase of prominence of gazed-at objects to the extent that a perceived referential intention of the speaker allows to retrieve the semantics of the gazed-at objects and possibly even to integrate the object with sentence meaning before the actual referring expression is uttered (situational integration account).

As most research conducted on the influence of speakers' gaze on listeners' language comprehension acquired behavioral data (e.g.: reaction times, eye-tracking), the afore mentioned questions can not be answered. In order to gain deeper insight into the extent of speaker gaze utilization by listeners, it is necessary to investigate the involved mechanisms in situated language comprehension. As mentioned before, two accounts are identified and tested in this thesis that possibly could explain the reviewed behavioural data: 1) The prominence account, based on an increase in a gazed-at objects prominence, would predict the involvement of a mechanism of semantic retrieval that benefits from the focus of attention on the gazed-at object, but wouldn't necessarily predict the involvement of any other language-related mechanism. 2) The situational integration account similarly would also predict the involvement of a semantic retrieval mechanism, which benefits from referential expectations based on preceding gaze cues to objects. The rationale behind this assumption is that, if gaze is used incrementally as part of language comprehension, the gaze to an object preceding its mentioning could possibly be utilized to retrieve the gazed-at objects semantics before its actual mentioning. Additionally, such anticipatory effects could very well also lead to the involvement of a matching mechanism: If a certain referent is anticipated to be named, the incoming acoustic signal can be matched with the anticipated word-form to early on validate expectations. Finally, the account would also predict an integration or monitoring mechanism that maintains a mental representation of the situation: If the gaze to an object leads interlocutors to anticipate this object as the upcoming referent, it can be integrated into the mental representation before its naming.

To go beyond the behavioral evidence, I utilized the event-related potential (ERP) paradigm, which has demonstrated to differentially index various aspects of linguistic processing. The following chapter will briefly provide the background of the method and introduce the so-called ERP components that are relevant for this thesis and have been found to index the predicted mechanisms.

CHAPTER 3

Background of Neurolinguistics

The origins of modern neurolinguistics — connecting the brain with language — are rooted in the research of aphasias or linguistic deficits that are connected to brain damage. Already in the 19th century, Paul Broca and Carl Wernicke discovered the connection of specific brain regions with certain aphasias. The discoveries of these areas of the brain, called Broca's area (1861) and Wernicke's area (1874) respectively, are to this date considered landmarks in the field (Phillips and Sakai, 2005). In more recent years, technical advancements gave rise to brain imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) as well as time-sensitive electrophysiological imaging techniques as Magnetoencephalography (MEG) and Electroencephalography (EEG) that allowed to go beyond aphasiology (Brown and Hagoort, 1999). The studies presented in this dissertation employ EEG, due to its temporal sensitivities and differential indices of relevant linguistic processes. In the following section I will briefly introduce this methodology and the relevant event-related potential (ERP) components in more detail.

3.1 Basics of Electroencephalography

EEG was first discovered by Richard Caton in 1875 and first recorded on human scalps by Hans Berger in 1924. Berger stated, that changes in a patient's state — as for example shifts from relaxation to alertness — could be reliably observed by utilizing EEG (Bronzino, 1995) when comparing the signal in correspondence to the different stimuli that induce the shift in a patients state. This can be considered as an important first step toward the broad variety of applications of EEG in modern research.

The communication of the on average about 86 billion neurons in the human brain (Herculano-Houzel, 2009) is one of its main function. This communication governs the the processing of and reaction to external influences such as visual perception, language processing, or tactile input, to name a few. So called pyramidal cells or pyramidal neurons form populations along cortical columns. When pyramidal neurons are activated in synchronicity, they generate electric and magnetic fields. These fields can in turn be recorded by the utilization of electrodes that are placed on a subjects scalp. The recorded EEG expresses the fluctuations from positive to negative current induced by excitation and inhibition of neurons. Hence, it can be said that, a potential is generated based on the neurons' plasma membranes permeability to sodium and potassium, that is directly connected to a membrane's charge. The resting state of the pyramidal cells holds a negative charge and a more positive change in this charge leads to the firing of the neuron when a certain threshold is reached. Positive charges are induced by a flow of sodium into the cell, creating an excitatory postsynaptic potential (EPSP). In contrast, an inhibatory postsynaptic potential (IPSP) is induced when the membrane allows the flow of potassium into the cell, increasing the negative charge of the cell and, thereby, distancing the charge even further from the more positive firing threshold. In both cases a dipole is formed: In the case of EPSPs positive charge — the source of the dipole — is found in and near the cell body — or soma — while negative charge — the sink of the dipole — is found around the apical dendrites that rise from the apex of the pyramidal cell's soma. For IPSPs, these charges are inverted. Whether the recorded EEG shows negative or positive deflections is dependant on the the positioning and formost on the orientation of the dipoles. Both EPSPs and IPSPs could lead to negative and positive deflections. For example, a negative deflection can be induced by either the negatively charged soma in case of an IPSP or the negatively charged dendrites in case of an EPSP depending on which lies in the vicinity of the recording electrode. If an electrode is equidistant to both the source and sink of a dipole the electrode will record a neutral charge and therefore neither dispaly a negative nor positive deflection. Also, based on the nature of dipoles, the negative deflections found in one region of the scalp results in a positive deflection in a different region of the scalp. It is also important to note that dipoles induced by single neurons cannot be measured on the scalp: Parallel aligned neurons must fire simultaneously in order for their summed dipoles in a region to be detectable on the scalp. The magnitude of the recorded dipole reflects the number of neurons that sum together (Dugdale, 1993; Kandel et al., 2000). If neurons were not arranged in parallel, the different charges would cancel each other out.

More detailed information about the neurobiological underpinnigs of the EEG methodology can be found in work from, for example, Da Silva (2009) and Jackson and Bolger (2014).

3.2 ERPs and Linguisitc Components

While EEG provides a useful tool to investigate brain activity, a direct link to specific cognitive functions and processes was not trivial based on EEG data alone. Mostly because many of these processes lead to only small temporally and spatially overlapping deflections in the EEG and are thereby also prone to be shrouded by noise in the signal. A solution to this problem was introduced around 1935 by Pauline Davis in the form of event-related potentials (ERPs)(Davis, 1939). The ERP method delivers a measurement of brain responses alligned to the onset of specific events. These events, or stimuli, can vary in nature, such as sensory, motor or cognitive events. ERPs are computed by averaging EEG data that is time-locked to these stimuli. As EEG basically is expressed by waves with different frequencies, such waves that are induced as a brain response to certain stimulus types would add to each other, whereas noise waves that are unrelated to the stimulus — given a high enough number of stimuli — would cancel each other out (Luck, 2014). Following the early introduction of ERPs into scientific research, the discovery of ERP components in the 1960s that could be directly linked to certain cognitive functions progressed modern day neurolinguistic research tremendously. Some of the first isolated components were the contingent negative variation (CNV)(Walter, 1964) and the P3 component (Chapman and Bragdon, 1964; Sutton et al., 1965). In the 1980s electrophysiological techniques gained more interest in the linguistic field with the discovery of the N400: The first language-relevant ERP component shown to be sensitive to semantic aspects of language comprehension (Kutas and Hillyard, 1980a,b,c).

Following this ERP component, more language-related components were discovered over the years. In the following, I will introduce the ERP components that are relevant for the course of my studies.

3.2.1 P300

The P300 or P3 component belongs to one of the most researched ERP components and was first observed by Chapman and Bragdon (1964) in response to meaningful stimuli (numbers) as compared to meaningles stimuli (flashes of light). The component was then described by Sutton et al. (1965) as a 'late positive complex' linked to the probability of a stimulus type to appear, so that less probable stimuli elicited a larger positivity.

Over the years, it was discovered that the P300 response could be separated into two types (Squires et al., 1975). In the oddball paradigm, first utilized in their study, participants were presented with a sequence of stimuli that are identical to one another. The sequences is then 'interrupted' by stimuli of a different structure. Squires et al. (1975), for example, presented sequences of either 'loud' (90 dB) or 'soft' (70 dB) tone burst in a constant stream of 65 dB background noise with equal distance of occurrence. The sequence was then 'interrupted' by the non-sequential type. For example, if the sequence consisted of 'soft' tone bursts, the interruption was induced by a 'loud' tone burst and vice versa. They found that these interruptions or 'oddballs' elicited a P300 effect. However, the effects differed in distribution depending on whether participants were instructed to ignore the tones or to actively attend to them. The former task elicited a positive deflection in frontal electrodes with a peak between 220 and 280ms (P3a), whereas the latter was found to peak between 310 to 380ms in posterior electrodes (P3b). Although first used with the loudness of auditory stimuli, the oddball paradigm became widely used with a variety of stimuli properties, such as other acoustic properties like tone pitch (Duncan-Johnson and Donchin, 1977), but also including visual properties such as object size and shape (Comerchero and Polich, 1999). Overall, a P300 was linked to the probability or expectability for a stimulus in context. The less expected a stimulus is, the larger is the P300 evoked (Donchin, 1981). The P3a later was also described as the 'novelty P3', whereas the P3b, which was the component originally described by Chapman and Bragdon (1964) and Sutton et al. (1965), became also known as the 'classic P300'. As mentioned, the two components differ in scalp distribution and peak latency, and are also elicited by somewhat different tasks, which will be discussed in more detail in the following.

3.2.1.1 P3a

The novelty P3, or P3a, peaks around 250-280ms in fronto-central electrodes. It has been linked to attention and the processing of novelty Polich (2003). It was, for example, shown by Yamaguchi and Knight (1991) that 'tactile novel' and 'shock novel' stimuli elicit a somatosensory novelty P300. In their study, participants were tapped on fingers. Taps to the second finger occurred frequently (76% of the stimuli – standard) whereas taps on the fifth finger were rarer in occurrence (12% – target). Additionally, taps on the third and fourth finger as 'tactile novel' and slight shocks to the wrist as 'shock novel' stimuli occurred only 6% of the time each. They found a novelty P300, or P3a, only in response to the 'novelty' conditions, whereas responses to the targets elicited a more parietal P300.

The main difference between the P3a and P3b is that only the former decreases in strength when the eliciting stimulus is repeatedly presented. This indicates that the stimulus is encoded into memory and no longer is perceived as novel. It is therefore considered that the P3a represents the orienting response, which is also known to decrease with exposure (Soltani and Knight, 2000). Further support for the link between the P3a and novelty was provided by Grillon et al. (1990). In their study, they presented participants with either repeating targets deviating from the standard, or with targets that were novel each time. They found that the P3a was stronger in the latter case.

Although this indicates a difference between the P3a and P3b component, it is argued that factors influencing the P3b also influence the P3a such as the probability of a stimulus or its evaluation difficulty (Polich and Kok, 1995).

3.2.1.2 P3b

The classical P300, or P3b, peaks around 300ms in parietal areas, with variations of the peaks latency between 250 and 600ms dependent on the task (Polich, 2007). It was shown to be elicited by task-relevant acoustic, visual and somatosensory stimuli (Heinze et al., 1986; Comerchero and Polich, 1999). The P3b has also been shown to be task dependent and requires the attention to — and evaluation of — a stimulus. Eliciting tasks include attention tasks, memory tasks, or visual search tasks (Verleger, 1997; Kok, 2001). The P3b can, as the P3a, be elicited in the oddball paradigm (Polich, 2007), but also in tasks that increase the working memory load, such as a dual-task paradigm (Kok, 2001). In the latter it was shown that tasks requiring less working memory in comparison to a preceding task elicit a smaller P3b (Luck, 1998).

Among other fields, the P3b has been widely used in the field of cognitive science in the area of information processing. In 1981, Donchin described the link between the P3b amplitude and the probability of a stimulus, so that less probable stimuli elicit stronger P3b responses. The probabilities can be derived from different sources. For example, the overall distribution of the less frequent target events in comparison to the standard events, where rarer events elicit stronger P3b effects, or the sequence in which they appear so that the reaction to a target following another target is lower than to a target following a standard stimulus (Duncan-Johnson and Donchin, 1977). Also the frequency of a target in a certain amount of time can influence the P3b amplitude (Polich and Margala, 1997). Johnson (1986) stated that the P3 amplitude is influenced by subjective probabilities as described above in addition to human judgment on the relevance of a stimulus, but also by stimulus meaning (Chapman and Bragdon, 1964) and the transmitted information of the stimulus. He defined the latter in terms of the amount of information a stimulus transmits to the perceiver and the originally contained information of the stimulus, so that the more information a stimulus transmits, the stronger the positivity.

The processes underlying the P3b have been broadly discussed. Verleger (1997) proposed a model of task related decision making, so that the P3b indexes the decision on whether a stimulus belongs to a task-relevant category. Kok (2001) similarly proposed a model of event categorization, so that the P3b indexes the matching of a stimulus with an internal representation of a certain category. Another model was proposed by Donchin (1981), who described the P3b as an effect of 'context model updating'. This model includes the idea that context is constantly utilized to form a mental representation of a given situation and to use this representation to form expectations about the upcoming information. According to this model, the P3b indexes the necessity to change these expectations or, more broadly, the current state of the working memory. This falls in line with the assumption of a mental representation of the situation as described, for example, by Zwaan and Radvansky (1998), who described this representation in terms of a situation model (see also van Dijk and Kintsch (1983) or Zwaan et al. (1995)). This type of mental representation was also described under different names such as 'mental model' (Johnson-Laird, 1983), message-level representation (Morris, 1994), or discourse representation (Kamp and Reyle, 1993). This interpretation of the P3b also falls in line with some interpretations of the functionality of the P600 (Burkhardt, 2007), another late positive ERP component that will be discussed later. A commonality between the interpretations remains the event probability as a main eliciting factor of the P3b, where less probable events elicit larger P3b amplitudes.

In summary, although sharing common factors that influence their amplitudes, the P3a and P3b index different processes. While the P3a mainly stands in relation to the novelty of a stimulus, the P3b seems to be related to other properties of a stimulus, such as information content (Johnson, 1986), task-relevance (Verleger, 1997; Kok, 2001) and contextual probability (Donchin, 1981).

3.2.2 P600

The P600 component, first described by Osterhout and Holcomb (1992), is often interpreted as indexing syntactic violations in both written and spoken language (Hagoort, 2007).Due to its sensitivity to syntactic violations, it is also refered to as a syntactic positive shift (SPS) (Hagoort et al., 1993; Coulson et al., 1998). The component with a positive deflection usually has an onset around 500ms and is long lasting with a peak at around 600ms in centro-parietal regions (Hagoort et al., 1999; Gouvea et al., 2010). Some sources, however, report earlier or later onsets, such as 400ms (Kaan and Swaab, 2003) or 600ms (Friederici, 2002).

As mentioned, the component was originally linked to syntactic violations. It was for example found in context with ungrammatical stimuli, as in a study by Kaan and Swaab (2003) where the 'are' in the sentence 'The man in the restaurant doesnt like the hamburger that are on his plate' elicited a P600 effect due to agreement violation. Further findings that indicated the syntactic nature of the P600 in context of ungrammatical sentences linked it to agreement violations of tense, gender, number, case and phrase structure (Gouvea et al., 2010). A P600 effect was also observed for garden-path sentences (Osterhout and Holcomb, 1992). These sentences, although in fact grammatical, are often initially parsed incorrectly. This is often due to different grammatical roles a word can take when parsed incrementally as in the sentence 'The horse raced past the barn fell' where 'raced' could either be a finite verb or, as eventually correct, a passive participle. Osterhout and Holcomb (1992) proposed that the necessity to backtrack and reanalyze the sentence is eliciting the P600 effect.

Besides linguistic evidence, the P600 also has been found in non-linguistic sequences that still follow certain rules. For example, in music, errors in harmony have been found to also elicit P600 effects (Patel et al., 1998).

Although the reported sources strongly hint toward a link between syntactical violations and the P600, the component has also been found in contexts where no syntactic violations were present. For example, grammatically correct sentences with high complexity, as for example such sentences that contain a high number of noun phrases, also have been shown to elicit a P600 (Kaan and Swaab, 2003; beim Graben et al., 2008). Kim and Osterhout (2005) also showed the presence of a P600 effect in context with semantic violations — such as role reversal anomalies — rather then syntactic violations. They found the effect on the word 'devouring' in the sentence 'The hearty meal was devouring the kids'. However, they suggest that this violation is not treated as a semantic violation by participants, but rather leads to syntactic repairs to achieve a preferred construct such as 'The hearty meal was devoured by the kids'. In the literature, such a P600 effect that appears in the context of semantic violations but is interpreted as actually indexing a syntactic repair toward a semantically more sound sentence is often referred to as a 'semantic P600' or 'semantic illusion'. This is similar to the explanation of the P600 given by Kolk et al. (2003). In their study participants were presented with sentences of the form 'De stropers die op de vos joegen slopen door het bos' (literal: 'The poachers who at the fox hunted stalked through the woods'; paraphrased: 'The poachers who hunted the fox stalked through the woods') and their semantically inverted version 'De vos die op de stropers joeg sloop door het bos' (literal: 'The fox that at the poachers hunted stalked through the woods'; paraphrased: 'The fox that hunted the poachers stalked through the woods'). Similarly, they found a P600 effect on the verb ('hunted') although the sentence is syntactically correct. As Kim and Osterhout (2005), they argue that this effect is due to syntactic repairs in order to form a semantically expected sentence. Van Herten et al. (2005) replicated these results with similar stimuli as used in Kolk et al. (2003) with the adjustment of the number of the nouns appearing together ('The poachers who hunted the fox stalked through the woods' was changed to 'The poacher who hunted the fox stalked through the woods', etc.).

A stronger link to semantic information of the P600 was proposed by Burkhardt (2007). In their study, participants were presented with a 'mini discourse' of the form: *'Yesterday a Ph.D. student was shot downtown. The press reported that the pistol was probably from army stocks.'* While the second sentence remained the same across conditions, the verb in the first sentence was changed so that it led to a) a 'Necessary context' where the verb and instrument are directly linked as in the example above (*shot – pistol*), b) a 'Probable context' in which the pistol is a possible but not necessary instrument (*Yesterday a Ph.D. student was killed downtown.*), or c) an 'Inducible context' in which not only the instrument but also the underlying action had to be induced (*Yesterday a Ph.D. student was found dead downtown.*). Their data provided evidence that the P600 found on the noun in the second sentence (*pistol*) can be linked to discourse memory demands, as both context for which inferences have to be drawn (conditions b and c) elicited an effect. They interpret the effect in terms of mental model updating costs (Bornkessel and Schlesewsky, 2006a; Burkhardt, 2006), so that new information that leads to the necessity to update the mental model impedes discourse memory capacity. This interpretation is inline with interpretations concerning the link between the updating cost of a mental model and the P3b component (Donchin, 1981; Bornkessel and Schlesewsky, 2006a; Delogu et al., 2018). It is therefore sometimes considered to belong to the same component family that is domain-general and elicited by rare and/or informative events (Pritchard, 1981; Coulson et al., 1998). This will be discussed in more detail later on.

Drenhaus et al. (2011) also interpreted the P600 found for violations of exhaustivness in only-foci in terms of a process of context updating. In their study, participants were presented with sentences of the form *"Nur Maria kann das Klavier spielen und auerdem noch die Geige"* (*'Only Mary can play the piano and, besides, the violin'*), which are exhaustive so that *'Mary'* remains the only person capable of playing the piano throughout the sentence. Such sentences were contrasted with sentences of the form *"Nur Maria kann das Klavier spielen und außerdem noch Luise und Jana"* (*'Only Mary can play the piano and, besides, Luise and Jana*). The addition of *'Luise and Jana'* violates the exhaustiveness required by *'only'*.

As evident from the summarized research, the role of the P600 component is not entirely clear. This holds especially, when put into context with the N400 component. Considering that the N400 component is often interpreted to express semantic integration, the very same interpretation for the P600 — as proposed by, for example, Burkhardt (2006) — seems unlikely when regarding cases in which an N400 effect was reported in the absence of a P600 effect. The interplay of N400 and P600 and the connected interpretations will be discussed later.

3.2.3 N400

As noted above, the N400 component can be considered as one of the first language-related components discovered. In 1980, Kutas and Hillyard adjusted the oddball paradigm, which was known to elicit P300 effects, for linguistic materials. In their study, participants were presented with short sentences that were to a large number (75%) semantically congruent sentences such as 'I shaved off my mustache and beard'. The remaining sentences (25%) ended either with an unexpected but possible word as in 'He planted string beans in his car' or entirely semantically incongruent as in 'I take coffee with cream and dog' in two experiments respectively. In both experiments, the anomalous word was meant to be the 'oddball'. However, despite their goal to replicate the P300 on linguistic stimuli, no such effect was found. Instead, they found a negative going effect that peaked at around 400ms and that was strongest for the semantically entirely incongruous words.

Subsequently, research was conducted to determine the concrete underlying mechanisms to this effect. Early on, the N400 was believed to be elicited by linguistic anomalies or violations, despite the initial findings of an N400 also in the context of non-anomalous but less expected words as in 'He shaved off his mustache and eyebrows' Kutas and Hillyard (1980c). It was shown, that the N400 is also not only elicited by written words, but could also be evoked by stimuli of different nature, such as spoken and signed words and even pseudo-words that are pronounceable but do not hold meaning in a language (for example 'pank') (Holcomb and Neville, 1990; Kutas et al., 1987). Eliciting stimuli also included known acronyms (Laszlo and Federmeier, 2007) and non-linguistic but meaningful stimuli such as environmental sounds (Chao et al., 1995; Van Petten and Rheinfelder, 1995), drawings of objects and scenes (Nigam et al., 1992; Ganis et al., 1996; Ganis and Kutas, 2003), faces and gestures (Barrett and Rugg, 1989; Bobes et al., 1994; Kelly et al., 2004; Wu and Coulson, 2005) as well as movies (Sitnikova et al., 2008). Therefore, the interpretation of the N400 shifted, so that it is linked to any potentially meaningful information. It is also to be noted that the N400 component is present independently of a stimulus' congruence. Both congruent and incongruous events elicit the component, but to a different intensity. Therefore, studies on the N400 speak of the N400 effect as a difference between the mean amplitude of the N400 between congruent and incongruent events.

The vast variety of stimuli types that can elicit an N400 effect also comes with a number of factors that were found to influence the N400. It was shown that a word's frequency in a language is correlated with the N400 size, so that less frequent words elicit a stronger N400 effect (Rugg, 1990; Van Petten and Kutas, 1990). These frequency effects, however, have also been shown to be, at least in part, overwritten by other factors that influence the N400. Van Petten and Kutas (1990), for example, showed that supportive sentential context can lower the N400 effect size. In other words: the expectability of a word, provided with context, also influences the size of the N400. This expectability can again be derived from different sources. Cloze probability of a word counts as one of the strongest predictors for the N400 effect (Kutas and Hillyard, 1984; Dambacher et al., 2006; Delong et al., 2011). Additionally, the more support a sentence offers for a certain continuation, the easier this continuation is processed by the subject as evident by an more attenuated N400 (Van Petten and Kutas, 1990, 1991). Influences can further be derived from sources beyond the sentential context. It was shown that also world knowledge (Hagoort et al., 2004), discurse information (George et al., 1994; Van Berkum et al., 1999) and even information about the speaker (Van Berkum et al., 2008) influence the size of the N400.

Delogu et al. (2019b) showed expectability related N400 effects for questions under discussion (QUD). In their study, participants were presented with two sentences. The first sentence was either providing a non-actual context (*"Peter hatte einen langen Tag und wollte ein Bier"; 'Peter had a long day and wanted a beer'*). or actual context (*"Peter hatte einen langen Tag und trank ein Bier"; 'Peter had a long day and drank a beer'*). It was hypothesized that the non-actual context introduced a QUD (Did Peter get a beer?). The initial NP of the second sentence then either commented on the context (*"Die Kneipe war bis Mitternacht geöffnet"; 'The bar was open till midnight'*), or not (*"Das Essen war bereits auf dem Tisch"; 'The meal was already on the table'*). They found an N400 effect only when a non-actual context introducing a QUD was not commented on. This indicates that participants had expectations for a commenting continuation of the context and that a violation of those expectations lead to difficulties in retrieving the word.

Another factor influencing the amplitude of the N400 is priming. Priming occours when a presented stimulus or prime activates some features of the target. Again, these features can be of a broad variety as, for example, similar physical or functional features as well as semantic association (Bentin et al., 1985; Kellenbach et al., 2000). Prime as well as target do not necessarily have to be words. Prime – target pairs of the structure picture – word (Barrett and Rugg, 1990a,b), picture – picture (McPherson and Holcomb, 1999), and word – picture (Ganis et al., 1996; Pratarelli, 1994) alike have shown to influence the N400 amplitude on the target.

The distribution of the N400 on the scalp is mostly found to be centro-posterior. This holds especially for visual word presentation (Kutas and Hillyard, 1982). However, depending on the modality, the distribution can differ. Auditory word presentation, for example, can lead to a more central or even global scalp distribution (Connolly et al., 1990, 1992; Connolly and Phillips, 1994; Bentin et al., 1993; Ackerman et al., 1994; McCallum et al., 1984; Holcomb and Anderson, 1993) and N400s on pictures can have a more frontal distribution (Holcomb and Mcpherson, 1994; Ganis et al., 1996). It was argued that component overlap with preceding components might influence the observed shifts in distribution. For example, an overlap with the N300 component (McPherson and Holcomb, 1999) found in picture processing that has a frontal distribution or the PMN component (Connolly and Phillips, 1994) in spoken word recognition that usually shows a fronto-central distribution might be the reason for a different observed distribution of the N400 effect. This argument however could raise the question why such a shift is not induced by, for example, the N270 component (Newman and Connolly, 2004) present in written word recognition, which also has a more frontal distribution. Therefore, it is still an open question what exactly leads to these different distributions.

The usual time-window reported for the N400 effect lies between 300 and 500ms (Kutas and Hillyard, 1980a, 1983). Some studies, however, also reported N400 effects for sub-time-windows as 300-400ms (Caldara et al., 2004), 350-450ms (Heimberg, 2002) and 350-500ms (Ganis and Schendan, 2013; Jończyk, 2016). In auditory sentence presentation, Van Petten et al. (1999) even mentioned an onset of the N400 at 200ms. This early onset however could possibly be explained by component overlap with the phonological mismatch/mapping negativity PMN — a preceding ERP component present for auditory stimuly.

It is still an ongoing debate whether the N400 expresses semantic integration (Friederici et al., 1993, 1999; Van Berkum et al., 1999), as also proposed for some interpretations of the P600 (Burkhardt, 2007; Brouwer et al., 2012; Delogu et al., 2019a), or whether it indexes semantic memory retrieval expressing long-term memory access (Federmeier and Kutas, 1999; Kutas and Federmeier, 2000). The interplay of N400 and P600 and the connected interpretations will be discussed in section 3.3.

3.2.4 PMN (N200)

The N200 component as a phonological mismatch negativity (PMN) usually peaking between 250-300ms was first observed in works by Connolly et al. (1990, 1992). In their experiments on contextually constrained spoken sentences this component preceded the N400 on sentence terminal words. Although appearing on every such word, the component showed a significantly larger negative deflection to semantically congruous words at the end of sentences with low contextual constraint. It was proposed that this early component reflects phonological processing of a stimulus that unfolds over time, such as spoken language. In a later study, Connolly and Phillips (1994) showed that the PMN is independent from later components (specifically the N400). In their experiment they presented participants with four conditions. The presented sentences either ended on a semantically correct or anomalous word. The correct words again were split in words with a high-cloze probability opposed by low-cloze probability. For example, the sentence 'Don caught the ball with his...' was either completed by 'hand' (high-cloze) or 'glove' (low-cloze). Anomalous sentence endings were split between endings that started with the same phoneme as an expected high cloze word ('The gambler had a streak of bad LUggage' instead of 'LUck') and such that were not similar to the expected high-cloze word ('The dog chased the cat up the queen' instead of 'tree'). The PMN was present in both conditions where the onset of the perceived word did not match with the onset of the high-cloze word. More precisely, it was observed for semantically correct, but unexpected words ('glove' instead of 'hand') and for anomalous words that did not begin with a phoneme similar to the expected high-cloze word. The PMN was not observed, however, for anomalous words that started with similar phonemes as the high-cloze word. Interestingly, a strong PMN to the initial phoneme of low-cloze words was followed by a reduced, almost absent N400, as the word was still congruent with the sentantial context. This was perceived as additional evidence for a phonologically related effect expressed by this component. Van Petten et al. (1999) performed a similar experiment to Connolly and Phillips (1994). They contrasted sentence final words with a high-cloze probability with semantically anomalous words that either rhymed with the high-cloze words, shared the same initial phoneme with the high-cloze word, or did not share the same initial phoneme. For example, the sentence 'It was a pleasant surprise to find that the car repair bill was only seventeen...' ended either with 'dollars', 'scholars', 'dolphins', or 'bureaus' respectively. Although they did not find morphologically distinct PMN and N400 patterns, they report that all anomalous words elicited an N400 effect. Most importantly, however, was that the onset of the reported N400 effect differed for the word starting with the same phoneme, so that the onset of the N400 was delayed by about 200ms compared to the other two anomalous conditions. As a result, they stated that the onset of the N400 was related to the first divergence of the perceived word from the expected word. As such, they argued that the N400 reflects the violation of semantic expectations formed on the basis of the context.

The most prominent difference between the works from Van Petten et al. (1999) and Connolly and Phillips (1994) lies in the perception of the ERPs in terms of the interpretation of the early effects as either an early onset of an N400 effect or an distinct PMN component preceding the N400 that, possibly, interacts with the N400 to the extent that it is morphologically indistinguishable. This interaction of components is often times called component overlap, which will be briefly discussed later on. The argument of a possible component overlap was also made by Praamstra and Stegeman (1993), who reported that the morphological distinction of the N200 and N400 is more prominent in frontal electrodes and less distinguishable in centro-parietal regions.

Further evidence for the distinction between the N200 and N400 was presented by Hagoort and Brown (2000). In their study, they presented participants with spoken sentences that either ended on semantically correct or anomalous words. They, as well, reported two distinct early effects, with the earlier peaking around 250ms followed by an N400. They argue that the early 'effect might reflect the lexical selection process that occurs at the interface of lexical form and contextual meaning' (p. 1528). Following the interpretation of Hagoort and Brown (2000), the PMN might not only express a mismatch of the perceived phoneme with the expected word-form, but could very well index an overall matching mechanism. The PMN was later renamed to a phonological mapping negativity rather than a phonological mismatch negativity. Although this was done foremost to avoid confusion with the established mismatch negativity (MMN) (Steinhauer et al., 2008), the renaming also is in line with findings that the PMN is always present when perceiving phonological input (Newman et al., 2003; Desroches et al., 2009). Further support for the distinct N200 component was provided in Dutch by Van Den Brink et al. (2001) and also for sentence internal words (Brink and Hagoort, 2004).

The studies reported so far would allow for an interpretation of the PMN to reflect some early form of lexical access. However, in addition to the findings for unexpected words, it was also shown that the PMN can be elicited by auditory presented non-words as well. In studies by Connolly et al. (2001) and Kujala et al. (2004) participants were visually presented with a word or non-word, followed by a single letter and finally an acoustically presented word or non-word. Participant then were asked to replace the initial sound of the read word with the sound of the presented single letter. The following spoken word then either was the word participants should end up with when replacing the sound or an unrelated word with a different onset. An example for a matching condition would be the sequence 'telk' followed by 'w' and then hearing 'welk', whereas a not-matching condition would be the sequence 'telk', 'w' and then hearing 'ket'. An increased PMN was found only for the conditions for which the heard word was not matching the word formed on the base of the visually perceived input. The effect, however was independent of whether the heard stimulus was an actual word or a non-word. This shows that the PMN is not only sensitive to the semantic content of a word but rather to the physical properties of the unfolding information. The study by Kujala et al. (2004) further provided evidence for the distinction of the the PMN from the N400. As they utilized the MEG method for their study, they showed that the PNMm is located in the left anterior temporal cortex, whereas the following N400m, while also lateralized to the left, is located in posterior temporal regions. Interestingly, they also observed that the region related to the PMN remained active throughout the recording period. This could possibly reflect processes of a phonological working memory.

Studies by Newman et al. (2003); Newman and Connolly (2009) and Desroches et al. (2009) provide further evidence for the stronger bottom-up functionality of the PMN. Newman et al. (2003) observed that the PMN is present in any condition that contains phonological processing but is strongest in such conditions that contain a violation of expectations for the anticipated word form. They presented their participants with a phoneme deletion paradigm that provided no semantic context. Participants listened to instructions to delete the initial phoneme of a word and then listen to one of four conditions consisting of a single word. For example, after hearing 'blink' the task was to erase the 'b'. They then heard either the correctly corresponding word ('link'), a word with the wrong phoneme deleted ('bink'), a word with the initial consonant cluster erased ('ink') or an entierly unrelated word ('tell'). The PMN was equally strong for all three non-matching conditions independently of whether the heard was an actual word or non-word ('ink' and 'bink' respectively). Desroches et al. (2009) further added results that show the incrementality of the effect and show that the expectations do not depend on linguistic predictors. In their study, they presented participants with pictures of objects and subsequently played a word in one of four conditions. The word either was the name of the depicted object (a picture of a cone followed by the word 'cone'), a word sharing a similar onset ('comb'), a word rhyming with the name of the depicted object ('bone') or an unrelated word ('fox'). The PMN was strongest for both the rhyming and the unrelated condition, but attenuated for the other two conditions. The following N400 time-window, however, showed slightly different effects, such that the N400 effect was also attenuated in the rhyme condition, suggesting that the rhyme competitor ('bone') was activated by the presentation of a picture ('cone'). Besides adding further evidence for the distinct processes underlying the PMN and the N400, this also suggests that in speech recognition bottom-up and top-down effects work in parallel consistent with models including such parallelism as for example the TRACE model (McClelland and Elman, 1986).

In summary, based on the reported studies, the PMN can be interpreted to reflect early phonological matching mechanisms that monitor the incoming signal on a word form level. The effect is strongest when the perceived signal does not match the anticipated word form. Expectations can be derived from a variety of sources such as linguistic context (e.g. cloze probabilities), visual information (e.g. images of objects) and experimental task (e.g. phoneme deletion paradigm). The time-window of the PMN is slightly different depending on the source. The reported time-windows usually are 250 – 300ms (Connolly et al., 1990, 1992), 150-250ms (Van Den Brink et al., 2001) and overarching both of them from 150 to 300ms (Praamstra and Stegeman, 1993). Although not always identified as a PMN, all reported findings link the N200/N250 to similar processes that are related to the phonological structure of stimuli. For the remainder of this thesis, I will refer to this early phonological effect as the phonological mapping/matching negativity (PMN).

3.3 Interpretations and Interplay of the N400 and P600

In regard to the different interpretations of both the N400 and P600, it is clear that at least some interpretations cannot be unified, as, for example, the interpretation as indexing semantic integration for both components. If both components were indexing the same mechanism, an effect of one component should not be found in the absence of an effect of the other. While effects of these components often are observed together, this is not always the case as will be discussed later. Various accounts have been proposed over the years that try to account for the often observed biphasic occurrence of the N400 and P600 components. Many of those accounts interpret the N400 as indexing semantic integration, but propose different explanations for the P600 component, often indicated by the aforementioned semantic illusion effect.

3.3.1 Semantic Attraction

As mentioned earlier, Kim and Osterhout (2005) proposed a model of semantic attraction. In their study, they found a so called semantic illusion effect (SIE) in sentences that violate animacy-based thematic roles (TR) as shown in the earlier example 'The hearty meal was devouring the kids'. Based on the properties of an inanimate object, such objects are not able to take the role of an agent with action verbs such as 'devour'. On the other hand, 'meal' is a noun often taking the patient role with a verb like 'devour', thus leading to a semantic attraction of the 'meal' being the patient in the example sentence. This condition elicited a P600 effect on the critical target word ('devouring') in their study. A control condition that contained the passive participle ('The hearty meal was devoured by the kids'), in contrast, did show neither a P600 modulation, nor an N400 modulation. Lastly, a third condition, named the 'no-attraction' condition, replaced 'meal' with 'tabletops' ('The dusty tabletops was devouring the kids'), an object usually not being the patient to verbs as 'devour' and, hence, not leading to a semantic attraction. This condition elicited an N400 effect, but no P600 effect in their study. They interpreted the component patterns across conditions, so that the P600 expresses a syntactic repair of the finite verb form to the passive participle in order to form a coherent sentence. In turn, they interpreted the N400 to express the semantic integration difficulties that a noun-verb combination without a semantic attraction to a specific role of the noun induces. This model however was challenged by findings from

Van Herten et al. (2005). In their study, they presented sentences that are not violated by TR assignment, but rather by world-knowledge. The sentence 'De vos die op de stroper joeg sloop door het bos' (literal: 'The fox that at the poacher hunted stalked through the woods' - paraphrased: 'The fox that hunted the poacher stalked through the woods') contains a verb that can be associated with both a fox and a hunter. However, when set into context, world-knowledge dictates that a hunter is more likely to hunt a fox than the other way around. As the verb additionally is in number agreement with both noun phrases, a P600 due to a syntactic repair in order to create an expected sentence should not be present. However, they found a P600 modulation on 'joeg' ('hunted') that is not in line with the semantic attraction account.

3.3.2 Monitoring Theory

Based on their findings, Van Herten et al. (2005) proposed another explanation for the presence of a P600 in the above mentioned cases. Their Monitoring Theory states the presence of a syntax-driven analysis that is working in parallel with sentence interpretations utilizing plausibility based on world-knowledge. The N400 component in this account is proposed to index the difficulty to arrive at a coherent interpretation of the sentence. If, the two streams (syntax and world-knowledge probabilities) arrive at conflicting interpretations of the sentence, a P600 modulation expresses the reanalysis process. Hence, an N400 modulation should be observed when a sentence interpretation is implausible, and a P600 modulation should be observed when the two streams are in conflict with their respective interpretation of the sentence. Although this approach is able to explain the data from Kim and Osterhout (2005) and Van Herten et al. (2005), it encounters problems to sufficiently explain some occurrences of biphasic N400-P600 patterns. For example, Hoeks et al. (2004) found a biphasic N400-P600 pattern for sentences like 'De speer heeft de atleten opgesomd' (literal: 'The javelin has the athletes summarized') on the word 'opgesomd' ('summarized'). As the sentence is syntactically sound, but semantically unusual, the Monitoring Theory should predict the presence of an N400 but no P600 modulation, instead of the observed biphasic N400-P600 pattern.

3.3.3 Continued Combinatory Analysis

Kuperberg et al. (2007) proposed a three stream account in which, as in the Monitoring Theory, one stream is based on semantic computations and another one on morphosyntactic computations. Additional to those two streams, they proposed a TR based stream that uses verb-agent-patient relations to form an interpretation. The three streams are proposed to influence each other constantly during online processing. In their study, Kuperberg et al. (2007) presented participants with sentences that are syntactically congruent, but contain predictability violations based on the TR similar to Hoeks et al. (2004) and Kim and Osterhout (2005). Sentences appeared in four conditions: a) a completely congruent baseline ('For breakfast the boys would eat toast and jam'), b) a pragmatic (contextually unexpected)violation but no TR ('For breakfast the boys would plant flowers in the garden'), c) a pragmatic sentence with an animacy TR violation ('For breakfast the eggs would eat toast and jam'), and d) a sentence with violations of both pragmatics and animacy TR ('For breakfast the eggs would plant flowers in the garden'). On the verb, they observed an N400 effect only for the pragmatic violation without a TR violation in condition b) and a P600 effect for both TR violation conditions (both not showing an N400 effect). The TR violated conditions c) and d) are argued to lead to a conflict between the syntax-driven stream, that assigns the 'eggs' the agent position, and the TR based stream, that would assign them a patient position. This conflict is assumed to be expressed by a revision induced P600 effect. Kuperberg et al. (2007) do not predict an N400 effect for these sentences, as, according to their model, the reanalysis process would block semantic processing. Only when the semantic stream is not blocked by a conflict of the other two streams, an N400 effect can be observed when pragmatics are violated as in condition b). Although their proposed account fits with their presented data, a later experiment by Kuperberg et al. (2010) found a biphasic N400-P600 pattern for conditions similar to condition d) in their 2007 experiment. In the sentence 'The journalist astonished the article before his coffee break', a similar conflict as assumed in the earlier experiment should be present where the TR stream should assign the patient role to the 'journalist', whereas the syntax-driven stream should assign the agent role. Therefore, no N400 should be observed if the semantic stream was blocked by this conflict. If interpreted in terms of their proposed account, it might be possible to assume that the semantic stream is not entirely blocked by the conflict of the other two streams, as originally proposed. However, even with a more lenient interpretation of the semantic stream blocking, this account could not accommodate the findings by Nieuwland and Van Berkum (2005). In their study, they presented participants with short stories that formed a context. These short stories consisted of multiple sentences introduced a number people and objects interacting with one-another (e.g., a woman and a toursit with a suitcase). The contexts were continued with either coherent continuations ('Next, the woman told the tourist that she thought he looked really trendy') or anomalous continuation ('Next, the woman told the suitcase that she thought he looked really trendy') in which an object ('suitcase') is given a role stereotypically not assigned to inanimate objects. The syntax-driven stream as well as the TR stream should assign the animate participant ('woman') the role of agent and the inanimate object ('suitcase') the role of patient. As this would lead to no conflict between the two streams, no P600 effect should be induced and, instead, an N400 effect should be found due to the semantically anomalous content of the sentence. In fact, however, the opposite was the case, so that the condition elicited a P600 effect, but no N400 effect.

3.3.4 Extended Argument Dependency Model

Bornkessel and Schlesewsky (2006a) proposed the extended Argument Dependency Model (eADM) as a three-phase hierarchically organized neurocognitive model of language comprehension. In the first phase, phrase structure information — or syntactic templates — are independently activated (Bornkessel and Schlesewsky, 2006b). In phase two, two streams are assumed. One stream processes prominence information for noun phrases (such information contains for example animacy and case marking) and links these nouns to processed verbs. This linking is done in terms of thematic role assignment. Additionally, this phase contains plausibility processing. Any difficulty in this stage is assumed to be expressed by an N400 modulation. In the third and final stage, the model assumes a 'generalized mapping' step. In this step, the result of the two streams are combined. A conflict of integration of the two streams with one another is assumed to be expressed by a P600 modulation. An example for such a situation can be found in the previously mentioned experiment by Hoeks et al. (2004). In the sentence 'De speer heeft de atleten geworpen' (literal: 'The javelin has the athletes thrown'), based on thematic role assignment, the situation should be interpreted as the 'athletes' being thrown by the javelin. The plausibility processing however should reverse the role of agent and patient. Finally, the current analysis is checked for 'well-formedness'. This process, for the model, is defined as a evaluation of a structure's acceptability under different environments, as for example discourse context. According to Bornkessel and Schlesewsky (2006a), this last step can also contain repairs based on the well-formedness check. These repairs are assumed to be expressed by a late positivity. They differentiate between the P600 and late positivities as they are assumed to represent different processes (generalized mapping and the check for well-formedness respectively).

Although this model is able to explain most of the previously introduced findings from studies (especially the data from Hoeks et al. (2004), Kim and Osterhout (2005), Van Herten et al. (2005) and Kuperberg et al. (2007)), an experiment by Kos et al. (2010) challenges the model. In their experiment, they manipulated thematic and grammatical role assignment in Dutch sentences. The two baseline conditions presented sentences that either contained a direct object ('Fred eet een boterham...'; literal: 'Fred eats a sandwich...') or a locative ('Fred eet in een restaurant...'; literal: 'Fred eats in a restaurant...'). The two anomalous conditions then were formed by switching the noun ('Fred eet een restaurant...'; literal: 'Fred eats a restaurant...' and 'Fred eet in een boterham...'; literal: 'Fred eats in a sandwich...' respectively). For both anomalous conditions, an increased N400 was found on the manipulated noun. This finding would not be predicted by the eADM. As 'restaurant' can be assigned the patient role in a sentence, there should be no difficulty for thematic role asignment. Also, as 'eat' and 'restaurant' can be plausibly combined, there should also be no difficulty in the plausibility processing. Hence, no N400 modulation would be predicted by the eADM. However, when the two analyses of these streams are to be combined in the 'generalized mapping' step, a conflict should arise, which should entail a P600 effect. Therefore, the findings of Kos et al. (2010) can not easily be interpreted by the eADM. Additionally, Kos et al. (2010) mention that the eADM does not provide any clear predictions for the sentences containing a locative, as the model is only designed for verb arguments. Similarly, the findings by Nieuwland and Van Berkum (2005) as well can't be easily explained by the eADM. Additionally to the P600 that was found in their experiment, the eADM should predict a N400 modulation for the anomalous continuation as a conflict should be expected for the plausibility processing as 'tell' and 'suitcase' do not stand in a obviously plausible relationship.

3.3.5 Retrieval-Integration account

The previously summarized accounts all are able to explain the findings from several studies, but fail to account for others. In reference to the up to this point reported studies, especially the findings from Nieuwland and Van Berkum (2005) seem not to be explicable by these models in their current form. In response to some of the shortcomings of the models, Brouwer et al. (2012) proposed a different approach. In their work, which also discusses the previously summarized models in greater detail, they proposed an alternative view of the meaning behind the N400 and P600 components. In their Retrieval-Integration account, the N400 component is assumed to express lexicosemantic access, or memory retrieval (Federmeier and Laszlo, 2009; Kutas and Federmeier, 2000, 2011; Van Berkum, 2009), whereas the P600 component is assumed to index integration of a word with the mental representation of the situation (Bornkessel and Schlesewsky, 2006a; Burkhardt, 2006, 2007). According to the account, biphasic N400-P600 patterns are expressing the retrieval and integration costs for each word added, such that both components are always present in language comprehension. Further differences in either the N400 or P600 (or both) observed across experimental conditions are then expressing the relative difficulty of a word's retrieval or integration with context respectively. The result of each N400-P600 cycle therefore is an updated mental representation of the communicated current situation. The account further assumes stronger effects for words carrying more meaning (or information) (Brouwer, 2014).

The Retrieval-Integration account can account for most of the findings of the earlier described studies. For example, the findings from Nieuwland and Van Berkum (2005) based on semantically anomalous continuations that elicited a P600 effect rather than the N400 effect predicted by the other models, is explainable in terms of this account. The contexts in their study introduced both referents that would either continue the context in a semantically coherent way (the 'tourist' in 'Next, the woman told the tourist that she thought he looked really trendy') or in a semantically anomalous way (the 'suitcase' in 'Next, the woman told the suitcase that she thought he looked really trendy'). As both referents were introduced, their retrieval from memory is effortless, expressed by an attenuated N400 in either condition. However, the integration with context to form a coherent representation of the situation is inhibited for the anomalous condition, as evident from the increased P600 in this condition only. Also the data showing both an N400 effect as well as a P600 effect in the same condition, as for example the data from Hoeks et al. (2004), can be explained by the Retrieval-Integration account. In comparison to their baseline condition 'De speer werd door de atleten geworpen' (lit. 'The javelin was by the athletes thrown), both semantically anomalous conditions containing an unrelated word ('De speer werd door de atleten opgesomd'; literally: 'The javelin was by the athletes summarized', and 'De speer heeft de atleten opgesomd'; literally: 'The javelin has the athletes summarized') elicited an N400 effect as well as a P600 effect, whereas the the anomalous condition containing a contextually expected word ('De speer heeft de atleten geworpen'; literally: 'The javelin has the athletes thrown') elicited only a P600 effect. These patterns are predictable under the Retrieval-Integration account. In all three anomalous conditions, the integration of the word making the anomaly evident ('thrown' and 'summarized' respectively) leads to integration difficulties with the mental representation of the situation. The two conditions that additionally elicited an N400 effect contain a word for which retrieval is inhibited, as it is not predictable from context. While the word 'thrown' still stands in a direct contextual relation to 'athletes' and 'javelin' and, hence, is easier to retrieve, the word 'summarize' stands in no relation to the context and is therefore harder to retrieve as evident by the increased N400 effect.

Regarding the data from Kos et al. (2010), however, the model potentially would predict a P600 effect for both anomalous conditions in addition to the observed N400 effects. For both 'Fred eet een restaurant...' (literal: 'Fred eats a restaurant...') and 'Fred eet in een boterham...' (literal: 'Fred eats in a sandwich...') the noun ('restaurant' and 'sandwich' respectively) should also elicit a P600 effect as the new information provided at that point should not be easily integrated into a mental representation of the situation, or, more generally speaking, the provided meaning should not be easily integrated with sentence meaning. Brouwer et al. (2012), however, mentioned that the stimuli used in the Kos et al. (2010) study appear semantically coherent until the end of the sentence. Therefore, following the averaging process, P600 effects occurring only on some critical words (such words in a context where the word itself renders the sentence anomalous) are lost due to those words that do not elicit a P600 effect (such words that are only appearing as anomalous when reaching the end of the sentence). Brouwer et al. (2012) argue that for example the sentence 'Fred eet een restaurant.' (literal: 'Fred eats a restaurant.') should elicit a P600, whereas 'Fred eet een restaurant...' (literal: 'Fred eats a restaurant...') could be continued in a way that forms a meaningful sentence as for example 'Fred eet een restaurant leeg.' (literal: 'Fred eats a restaurant empty.') In addition to the possibly on average diminished P600 effect, the lack of the effect in Kos et al. (2010) could also be partially due to component overlap between the strong N400 effect found in their data and the P600 (see Brouwer and Crocker (2017)).

This interpretation of the mechanisms underlying the N400 and P600 in combination with the aforementioned interpretation of the N200 as a PMN are the foundation of the hypotheses that will be presented in more detail in Chapter 4.3.

3.4 The P600-as-P3 Hypothesis

While the P600, as mentioned before, was interpreted as a syntactic error related distinct component (Osterhout and Holcomb, 1992; Hagoort et al., 1993) when originally discovered, further research — as described above — revealed different modalities in which a P600 can occur. Among those, commonalities between the P600 and P300 have often been reported, leading to the hypothesis that the P600 can be interpreted as a instance of the P3b (Coulson et al., 1998; Münte et al., 1998; Vissers et al., 2008; Van De Meerendonk et al., 2010; Bornkessel-Schlesewsky et al., 2011; Sassenhagen et al., 2014). While the similarities in the scalp distribution of two components are possibly indicatory for a relatedness, this is not necessarily evidential as similar topographical distributions may still involve different generators. Looking into the modalities evocing the components however, stronger connections can be drawn. As described above, the P3b is sometimes associated with context updating (Donchin, 1981; Polich, 2007). This is similar to some interpretations of the P600 as indexing integration (Brouwer et al., 2012) and also context updating (Bornkessel and Schlesewsky, 2006a; Burkhardt, 2006). In a more biological context, Nieuwenhuis et al. (2005) proposed an involvement of norepinephrine (NE) release from the Locus Coreolus (LC) in the the occurrence of the P300. This was related to cognitive (re)orientation as for example decision making, response execution or response inhibition. Similarly, Verleger et al. (2005) propose a monitoring mechanism behind the P300 that links perceived new information with reactions (and the decision making leading up to those reactions). The P600 component can be observed in similar conditions as the P300: The occurrence of a P600 followning syntactic violations could more broadly be seen as an overall violation or an unexpected event, which is, as described above, an eliciting factor of the P3b (for a more detailed review of the commonalities between the P3 and P600, see Sassenhagen et al. (2014)). Overall, both components are, at least in part, connected to task-relevance (Geyer et al., 2006) of a stimulus and violation salience (Van De Meerendonk et al., 2010).

Following this Hypothesis, it could be argued that the predicted P600 as indexing the integration mechanism under the situational integration account could instead be indexed by a P300 component or, as described above, by a P600-like P300. One reason for the apparently late P300 could be component overlap.

3.5 Component Overlap

Although ERP components are conceptually independent in terms of amplitude, latency and source, the recorded waveform measured on the surface of the scalp does not necessarily reflect the actual isolated components, but the sum and interplay of any component recorded at any site on the scalp. Therefore, when considering the ERP method, it is not possible to recover the actually involved components, their actual amplitude or actual latency from the recorded signal. A reduced effect, or even the absence of an effect, could therefore either stem from an actual lower amplitude of the component or from an overlap of two components with opposite polarity in a given time-window (Luck, 2014; Brouwer and Crocker, 2017). It was for example argued that the N400 may obscure a subsequent P600 (Delogu et al., 2019a). In contrast, component overlap of two components with the same polarity can make the two components indistinguishable from one another. Such situations are argued to be possibly responsible for findings such as those by Van Petten et al. (1999), where the reported early onset of the N400 could reasonably be an overlap between the PMN and the N400 component. Additionally, the P600-as-P3 Hypothesis as well could potentially be influenced by component overlap. It could be argued that, when accompanied by an early, strong negativity such as the N400 (and potentially PMN), the overlap of the negative component(s) with the P300 could result in an apparently late onset of the positivity.

Overall, the reviewed literature provides a base for the anticipation of the PMN, N400 and P3b/P600 components as indexing the mechanisms assumed under both the prominence account and the situational integration account.

CHAPTER 4

Speaker Gaze as a Predictive Cue for Linguistic Content

The previously reviewed works on gaze (Chapter 2 on page 7) laid out the effects of speaker-gaze on listeners' sentence comprehension as observed in behavioral studies. It was shown that gaze, as a communicative tool, is utilized by both speakers and listeners to ground and disambiguate referential expressions in situated linguistic interactions. Further, it has been shown that speaker gaze is also used to form expectations about the upcoming referents in a sentence. Also evident form those studies, however, is that the underlying mechanisms and the extent to which gaze is utilized is not well researched as to this point. Hence, the experiments presented in this chapter aimed to investigate whether the observed influence of speaker gaze on listeners comprehension stems from an increase in prominence of the gazed-at object alone (prominence account) or whether the gaze to an object entails a deeper processing of the gazed-at object with regard to the utterance (situational integration account) as laid out in Chapter 1 on page 1.

The current chapter describes two ERP experiments investigating whether gaze is utilized to anticipate upcoming linguistic content of a sentence (similar to the breakfast scenario in Chapter 1). If gaze was utilized this way, a processing beyond the increase of prominence could be deduced. Additionally, the previously described ERP method can shed light on the involved cognitive mechanisms as indexed by the ERP components discussed in Chapter 3 on page 15.

In these experiments, German native speakers were watching scenes containing three objects that either varied in size or shade, so that each scene contained a small, medium sized and large object, or light, medium shaded or dark object respectively. All objects in each item were of the same grammatical gender in German, so that no information was revealed by the articles preceding the nouns. After 3000ms, a stylized face appeared in the middle of the screen with a straight gaze toward the participant. The face then performed gaze actions timed to the auditory presented German sentence, so that the gaze was directed toward objects in the scene 800ms prior to their naming (Griffin and Bock, 2000; Kreysa, 2009). The sentence was uttered by the CereVoice TTS system's Alex voice (Version 3.2.0). The sentences were of the form *"Verglichen mit dem Auto, ist das Haus verhältnismäßig klein, denke ich"* (*'Compared to the car, the house is relatively small, I think'*). In order to minimize the influence of possible effects induced by the processing of the first noun (*'car'*) on the critical second gaze cue as well as on the subsequent noun (*'house'*), a pause of variable length was introduced after the first noun, so that the distance of the offset of the first noun to the onset of the second noun always was about 1000ms. This includes a jitter of 50ms to avoid entrainment of the alpha rhythm by the stimulation rate (Luck, 2014). The following section provides details about the creation and testing of the stimulus materials that were used in all presented experiments.

4.1 Stimulus Materials

In all presented experiments, participants were listening to synthesized sentences in which two out of three objects in a presented scene were compared as described above.

A stylized representation for both the face as well as the objects was chosen to keep the cognitive demand of recognition of the depicted objects and the face as low as possible in order to be able to draw stronger conclusions about the impact of gaze on sentence comprehension in a shared visual world. Hence, a stylized face and stylized objects that were easily manipulated in terms of size and shade and wouldn't contain to many features that might distract participants or could increase cognitive load beyond a basic degree of recognition were used.

For the auditory presentation of the sentences, a synthesized voice was used in order to acquire full control over the prosodic contour and turn-internal pauses. Additionally, this approach ensures that there are no differences in intonation patterns across sentences. The sentences across items in each experiment only differ in the two nouns representing the objects that are compared with one another and the adjective representing the comparative, while the remaining words remain constant across items. As mentioned, these constant words should be as similar as possible in order to be able to trace back any occurring effects to the changed words only. Also, reoccurring nouns and adjectives across sentences again should sound as similar as possible for the very same reason. The most straight forward approach to achieve the named goals is the use of a text-to-speech (TTS) system. The CereVoice TTS system was used to create the auditory stimuli. It provides two different German voices, the Alex voice (male) and the Gudrun voice (female). Both voices were used in Version 3.2.0.

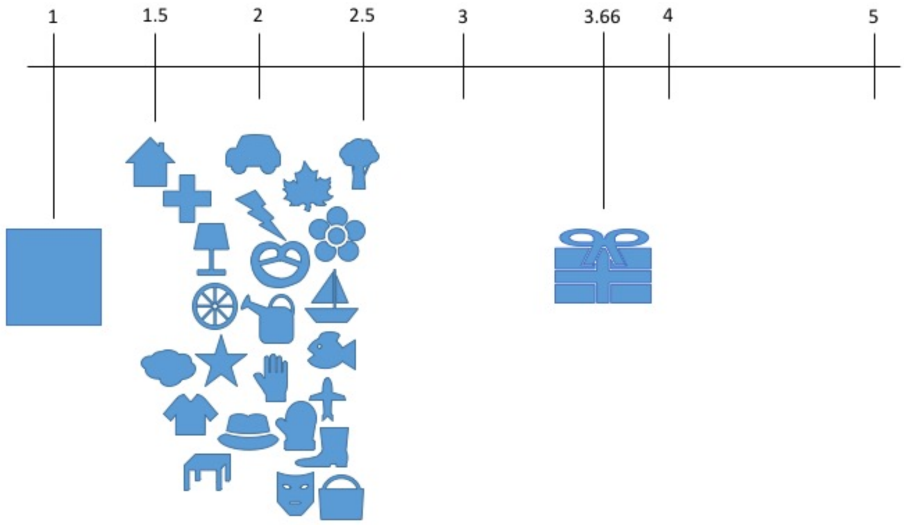


Figure 4.1: Distribution of complexity ratings of the used objects, including the objects with the highest and lowest rating.

4.1.1 Pre-test

In order to obtain easily recognizable and similarly complex visual stimuli, 60 stylized representations of objects of masculine (25), feminine (18) and neuter grammatical gender (17) respective to their naming in German were created. The pictures were presented to seven participants using Goggle Forms with the task to name the objects and indicate how complex they appear to the participant on a scale from 1 (low complexity) to 5 (high complexity). The lowest average rating given was 1 for a plain square. The highest average rating given was 3.66 for a present (see Figure 4.1 for comparison). Out of the 47 objects that were named identical by all participants, eight objects per gender were chosen for the experiment. In the selection of the objects, further, the complexity rating of the participants was used, so that only objects with a similar complexity ranging from 1.5 to 2.5 were chosen. Figure 4.1 provides an overview of the used objects. All used objects are summarized in Table 4.1 on the following page.

In order to present participants with as natural sounding utterances as possible while retaining a constant similarity between different sentences, different versions of example utterances were created that varied in speaker gender — utilizing the Gudrun (female) and Alex (male) voice respectively — as well as intonation contour and turn internal pause length. The sentences presented in the pre-test were of the form *Verglichen mit dem Haus, ist das Auto verhältnismäßig klein, denke ich* (*'Compared to the house, the car is relatively small,*

Table 4.1: Summary of the objects presented in experiment 1 with their English translation separated by grammatical gender

Masculine	Feminine	Neuter
Baum (tree)	Blume (flower)	Auto (car)
Blitz (bolt)	Brezel (pretzel)	Blatt (leaf)
Fisch (fish)	Gießkanne (watering can)	Boot (boat)
Handschuh (glove)	Hand (hand)	Flugzeug (airplane)
Hut (hat)	Lampe (lamp)	Haus (house)
Stern (star)	Maske (mask)	Kreuz (cross)
Stiefel (boot)	Tasche (bag)	Rad (wheel)
Tisch (table)	Wolke (cloud)	T-Shirt (t-shirt)

I think'), which corresponds to the sentence form as utilized in experiments one and two. A Google Form was used to collect responses of seven participants, who listen to those examples. The questionnaire can roughly be split in two types of questions.

The goal of the first type was to assess which of the two voices was overall perceived as more natural. In this set of questions, participants listened to two versions of the very same sentence only differing in the voice used to produce them. This means, that the code that was entered into the TTS system was not altered for each utterance pair. Across sentence pairs the nouns and adjective were changed to have a broad coverage of articulations of different words that appeared in the experiment itself. The nouns were taken from the list of objects decided on in the first pre-test. All four adjectives were presented to the participants across items. These being *"kleiner"* (*'smaller'*), *"größer"* (*'bigger'*), *"heller"* (*'brighter'*) *and "dunkler"* (*'darker'*). The participants' task was to rate each of the sentence for naturalness on a 5 point Likert scale ranging from 'very natural' (5) to 'very unnatural' (1). Additionally, they were asked to indicate which of the two utterances they preferred.

In the second type of questions, participants were presented with 5 different versions of the same sentence. This time, participants heard only one voice per item. However, the intonation contour and turn-internal pause length were manipulated for each of the versions. Again, participants were asked to rate each utterance on a 5 point Likert scale ranging from 'very natural' (5) to 'very unnatural' (1). Additionally, they were also asked to rank the 5 utterances from most natural to least natural.

The results of the first set of questions showed, that participants had a rather strong preference for the male voice overall. 57% of the answers named the male voice as more natural sounding, while only 31% indicate the female voice as more natural sounding. 12% of the answers fell on a tie between the two voices with both being equally natural sounding. The results for the naturalness by voice also showed this trend. The results are summarized in Table 4.2. In summary, the male voice received 59.7% of ratings stating it was natural sounding, whereas the female voice only received 45.2% ratings in the same direction.

The second set of questions as well showed that the utterances of the male voice were overall rated as more natural than the female counterpart. On average, the male voice received a rating of 3.5 whereas the female voice received a rating of 3.1. When only looking into the best performing intonation and turn-internal pause length versions for each voice, the male version was rated with an average score of 4.4 where as the female version only received a rating of 3.8. The best performing manipulation for the two voices were chosen based on the ranking given by participants as well as on the highest average score for the utterances.

As the results of both sets of questions strongly suggested the male voice to outperform the female voice in terms of naturalness, the CereVoice Alex voice was used to create the utterances for the experiments. The nouns uttered by this voice had an average length of 560ms with the shortest noun lasting for 383ms (Tisch) and the longest noun lasting for 791ms (Flugzeug).

Table 4.2: Summary of the ratings for naturalness for the two different voices.

Voice	Very Natural	Natural	Neutral	Unnatural	Very Unnatural
Female	19%	26.2%	21.4%	31%	2.4%
Male	28.7%	31%	33.3%	7%	0%

4.2 Experimental Design

This section describes the experimental design for the experiments presented in this chapter utilizing the previously described stimulus materials.

On speech onset, 1000ms after the face appeared, it retained its straight gaze but opened the mouth to evoke the impression of the face being the speaker of the sentence. The first gaze cue appeared approximately 800ms before the first noun was mentioned (Kreysa, 2009). This first gaze cue was always Congruent — toward the first named object — for all experimental trials. Also, in order to ensure the participants' attention throughout the entire sentence, the first named object in the experimental items was always the medium sized/shaded object. Since if the first mentioned object were the smallest/lightest or biggest/darkest object in the scene, it would not matter which of the other objects were named, as for both the same adjective would render the sentence true or false. An example of an experimental trial in experiment 1 can be seen in Figure 4.2, which displays the time line of a Congruent trial, containing a small house, medium sized car and a large t-shirt. If the first gaze action were directed toward the t-shirt, both remaining objects would be smaller and, hence, would no longer require the participant to pay attention to the upcoming noun in order to evaluate the sentence. The second, manipulated gaze cue appeared 800ms prior to the onset of the second noun. This gaze cue was manipulated so that it either was directed toward the consecutively referred to object in the scene (Congruent – *'house'*), toward the object that remained unmentioned throughout the course of the trial (Incongruent – *'t-shirt'*), or toward no object on the screen (Uninformative). The latter condition differed across the two experiments. In the first experiment, the gaze was directed toward the bottom of the screen where no object was present (Averted), while the gaze was redirected toward the participant in the second experiment (Mutual). In all conditions, the gaze was redirected toward the participant 400ms before the end of the sentence, and the mouth closed on the offset of the sentence. After each item, the participants were asked to indicate whether the sentence was true given the visual context they were presented with by pressing one of two buttons. Answers were recorded using a Response Pad RB-834 (Cedrus Corporation).

Assuming that the mental representation of the situated utterance is incrementally expanded, when hearing *'car'* in the example sentence above, participants should integrate this object as the base for a comparison introduced by the very first word of the sentence *'compared'*. Participants should then expect one of the two remaining objects to be consecutively named as the object *'car'* should be compared with. Based on linguistic content alone, participants should have no strong preference for either of the two remaining objects to be named, as the comparative (*'small'*) appears only after the two to-be-compared objects are named. However, preceding the mentioning of an object, the displayed face performed gaze actions toward the upcoming object so that, based on the gaze cue, participants possibly obtain evidence about which object is likely to be mentioned next. In short, if gaze cues are treated similar or equal to linguistic information in order to inform the mental representation of the situation, a gaze cue toward an object should increase the listeners expectation for the gazed-at object to be relevant for the situation and, hence, the listener should anticipate the object to be referred to next (Staudte et al., 2014).

4.3 Hypotheses

As stated above, two possible accounts regarding the influence of speaker gaze on situated comprehension were identified. The resulting hypotheses for the situated integration account — for which the gazed-at object is predicted to be processed beyond the increase of prominence of the cued position — entail three stages of language processing that are known to be indexed by distinct ERP components: Phonological matching (PMN), lexical retrieval (N400), and semantic integration and updating (P600). The proposed indexing component for the former (PMN) is taken from work by, e.g., Connolly and Phillips (1994) as discussed in Section 3.2.4 on page 25, while the proposed indexing components for the latter two mechanisms (N400 and P600) are based on the retrieval-integration account proposed by Brouwer et al. (2012) as reviewed in Section 3.3.5 on page 32. (See also Delogu et al. (2019b) for evidence and elaboration.)

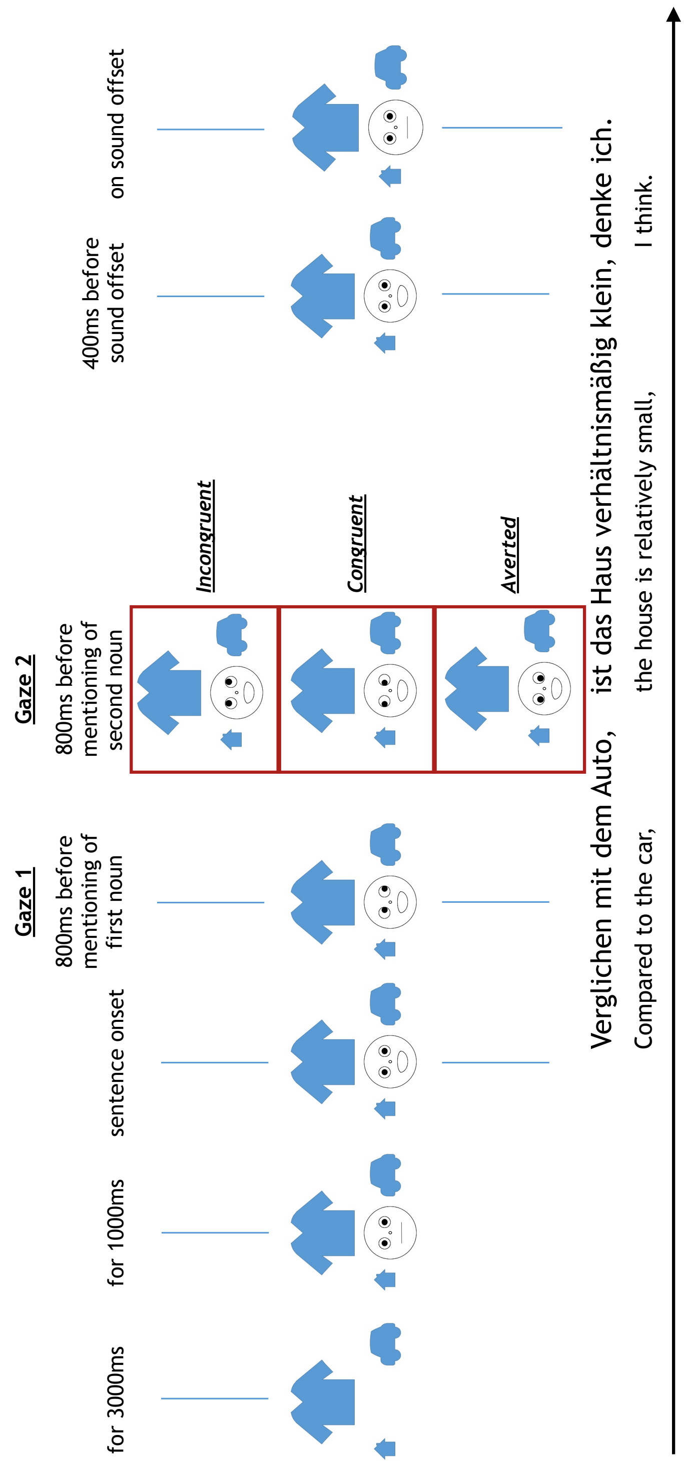


Figure 4.2: Timeline of an experimental trial in experiment 1.

Under the Prominence Account — for which the effects are proposed to stem from an increase in prominence alone — semantic integration should not be affected. Also, if no predictions about an upcoming referent are formed, it is unlikely for a phonological matching mechanism to be affected either. However, it could be argued that, if the PMN was also involved in a form of selection, such that the perceived initial phoneme is utilized to identify the actual referent, a brain response could be entailed. However, as no prior information of the upcoming referent is accessible under the PA, no difference across conditions should be detectable, as the initial phoneme of the referent conveys the same amount of information for all three conditions.

Thus, I hypothesized that the presence of an effect in the PMN and, primarily, P600 components would provide evidence for the Situational Integration Account, whereas the lack of such an effect would provide evidence for the Prominence Account. Both accounts however predict the presence of an N400 modulation.

More precisely, for the situational integration account, it is proposed here that information provided by gaze in linguistic context as well as the linguistic signal is incrementally used to inform a mental representation of the conveyed situation. Thus, any information provided — gaze as well as linguistic — is used to build a representation of the current situation. Evidence for possibly resulting expectations on the base of the current state of the mental representation of the situation (PMN and P600) could only be explained by the SIA. Further, unexpected input in turn is hypothesized to lead to an inhibited phonological matching mechanism on the phoneme level. In other words, the contextually expected word-form derived from the current state of the mental representation of the situation is matched with the actually perceived auditory signal as soon as the first phoneme is available. This matching mechanism is predicted to be expressed by an increased PMN effect (Praamstra and Stegeman, 1993; Connolly and Phillips, 1994; Hagoort and Brown, 2000; Van Den Brink et al., 2001) when the perceived phoneme does not match with the anticipated phoneme. Additionally, following the Retrieval-Integration (RI) account proposed by Brouwer et al. (2012), I further hypothesized an increased difficulty of retrieving the semantics of the referent that is expressed by an increased N400 effect (Van Petten and Kutas, 1990; Van Berkum et al., 1999; Federmeier and Laszlo, 2009; Kutas and Federmeier, 2000, 2011), when the referent differs from the gazed-at object (Incongruent condition) or when no prior information about the referent was accessible (Uninformative condition). Finally, an integration with the situation model is hypothesized to be impeded when the retrieved semantics are not unifiable with the current mental representation that was informed by the preceding gaze cue, as the semantics of the gazed-at object are not the same as the semantics of the heard referent. Such integration difficulties are expected to be expressed by an increased P600 effect (Bornkessel and Schlesewsky, 2006a; Burkhardt, 2006, 2007; Brouwer et al., 2012).

If, however, gaze was only utilized to draw the listeners attention to a certain object or region, one would not anticipate effects that indicate higher levels of processing. More precisely, while facilitatory and inhibitory effects on word retrieval are to be expected by the mere drawing of attention, this should neither lead to precise predictions of the upcoming word-form, nor should the gazed-at object be integrated into the mental representation of the situation as the anticipated upcoming referent. Hence, while an N400 effect for referents that were not gazed at could be expected, neither a PMN effect representing the phonological matching mechanism, nor a P600 effect representing the cost of the revision of the mental representation should be found. A summary of the proposed mechanisms and their predicted indexing ERP components can be found in Table 4.3.

Table 4.3: Summary of the assumed mechanisms and the corresponding indexing components.

Mechanism	Component	Base
Phonological Matching Mechanism	PMN	Praamstra and Stegeman (1993); Connolly and Phillips (1994); Hagoort and Brown (2000); Van Den Brink et al. (2001) Desroches et al. (2009)
Semantic Retrieval Mechanism	N400	Federmeier and Laszlo (2009); Kutas and Federmeier (2000, 2011); Van Berkum (2009); Brouwer et al. (2012)
Mental Representation Integration Mechanism	P300	Donchin (1981); Polich (2007)
	P600	Bornkessel and Schlesewsky (2006a); Burkhardt (2006, 2007); Brouwer et al. (2012)

Following the aforementioned hypotheses, the predictions for the two experiments are as follows: When gaze could be used to form expectations about the unfolding sentence content, there are no predicted effects in response to the gaze region itself. When encountering the gaze cue to an object, it is hypothesized under the Prominence Account that the gaze shifts the attention of the listener but does not elicit any deeper processing of the gazed at object or position and, thus, does not influence comprehension of the spoken utterance. Hence, no language related effects are anticipated. As there is no auditory input, no involvement of a phonological matching mechanism would be expected in this region.

Concerning the retrieval and integration mechanisms, while an involvement of both is predicted under the Situational Integration Account, no inhibition of either mechanism is anticipated as there is no conflicting information at this point. It could, however, also be argued that retrieval and integration that are entailed by the gaze to an object is more demanding than the lack of informative gaze.

In the noun region, the Prominence Account predicts an involvement of only a retrieval mechanism. Hence, an inhibition of that mechanism is predicted to result in an increased N400 effect. An inhibition is predicted to occur if the gaze cue preceding the mentioning of a referent drew the listeners attention away from the subsequently named object (Incongruent) and, possibly, if the referred to object still lies outside the listeners focus of attention (Averted). Under the Situational Integration Account, however, three involved mechanisms are predicted. If the preceding gaze was utilized to process the gazed-at object and to form expectations about the unfolding sentence, firstly, a phonological matching mechanism is predicted to compare the incoming phoneme with the predicted word-form. An inhibition is predicted to occur when the gazed-at object differs from the actually mentioned referent, resulting in an increased PMN response (Incongruent). Additionally, if the PMN indexes a selection process as proposed by Hagoort and Brown (2000), an increased PMN response could also be expected when no prior information was accessiable (Averted), so that an attenuated PMN response is only expected in the Congruent condition. Secondly, as under the Prominence Account, a retrieval mechanism is predicted that is involved whenever the semantic content of a word has to be retrieved: Either, because no semantics have been retrieved so far (Averted), or because the word that was retrieved is different from the actually perceived word (Incongruent). This is predicted to be expressed by an increased N400 response (in comparison to the Congruent baseline). Finally, an integration mechanism is predicted to be involved that monitors the current state of the mental representation of the situation formed on the basis of the preceding context that includes the gaze cue and updates that representation if the signal does not fit the current representation (Incongruent). Again, this inhibition is predicted to occur when the gazed-at object is not the same as the mentioned referent. This is predicted to be expressed by an increased P600 response. A summary of the described predictions for the experiments can be found in Table 4.4 on the facing page.

I hypothesized that, under both the prominence and situational integration account, Incongruent gaze should lead to higher retrieval costs of the subsequently mentioned word compared to the Congruent baseline, as the previous gaze evoked expectations for the gazed-at object rather than the actually referred-to object, which in turn are violated. This violation of expectations is predicted to be expressed by an N400 modulation (Van Petten and Kutas, 1990; Van Berkum et al., 1999; Kutas and Federmeier, 2011). Additionally, if gaze is utilized beyond the increase of the gazed-at objects prominence, as assumed by the situational integration account, as the word-form of the perceived referent does not match

Table 4.4: Summary of the predicted effects for the two proposed accounts. (PA - Prominence Account ; SIA - Situational Integration Account ; C - Congruent ; I - Incongruent ; U - Uninformative)

Region	Component	PA		SIA	
		C-I	C-U	C-I	C-U
Gaze	PMN	-	-	-	-
	N400	-	-	(+)	(+)
	P600	-	-	(+)	(+)
Noun	PMN	-	-	+	+
	N400	+	+	+	+
	P600	-	-	+	-

with the expected word-form based on the context informed by the preceding gaze cue, the phonological matching mechanism is expected to detect a mismatch. This mismatch is predicted to be expressed by an increased PMN effect (Connolly and Phillips, 1994; Hagoort and Brown, 2000; Van Den Brink et al., 2001). Furthermore, if the gazed-at object was already integrated into the mental representation of the situation, the consequently named object can not be unified with that representation and, hence, should require an update of that representation. This updating cost of the situation model is predicted to be expressed by a P600 modulation (Burkhardt, 2006, 2007; Brouwer et al., 2012).

In turn, I hypothesized that Uninformative gaze (Averted gaze) should as well entail an increased N400 compared to the Congruent baseline as either no object benefits from an increased prominence (prominence account) or because no expectations could be formed (situational integration account), rendering both remaining objects as equally likely. This, in turn, leads to a harder retrieval of the word compared to a more prominent or even anticipated word as present in the Congruent condition. Additionally, as the PMN is believed to expresses an overall matching mechanism (e.g., Desroches et al. (2009)), the initial phoneme can be utilized to select the correct target early on. Hence, the information conveyed by that phoneme is higher than in the Congruent baseline condition, which as well is predicted to be expressed by an increased PMN effect. However, Uninformative gaze is not hypothesized to require a revision of the mental representation of the situation. As no prior evidence was provided to inform the mental representation of the situation, the newly gained information provided by the linguistic signal can effortlessly be unified with the situation model. Hence, no increased P600 modulation is predicted. It could be argued, however, that even an effortless integration into the mental representation can be considered an update of that situation model. This could lead to an intermediate P600 modulation for these conditions. Yet, as only an update and no revision is required, this effect should be significantly lower than for the Incongruent condition.

4.4 Experiment 1

In the first experiment, the objects surrounding the face where placed on the horizontal and vertical axis through the eyes of the face. There always was an object left, right and above the face while the position below the face was used as the empty position the gaze was directed to in the Averted condition. Figure 4.3 on the next page shows an example screen.

Each item appeared in three conditions (Congruent (baseline) / Incongruent / Averted). In the Congruent condition, the gaze preceding the second noun was directed toward the subsequently named object (*"Haus"*). In the Incongruent condition, the gaze cue was instead directed toward the object that remained unmentioned in the sentence (*"T-Shirt"*). In the Averted condition, the gaze was directed toward the bottom of the screen where no object was present. This led to three lists using a latin square design. Additionally, versions of those lists were created that were counterbalanced for realism. Realism was defined based on the truth value of the performed utterance in the real world. For example, in the experiment, some trials contained utterances like *'compared to the car, the house is relatively small, I think'*. In the real world, such a statement would usually be false. Therefore, such 'unrealistic' statements were counterbalanced with their 'realistic' version (e.g. *'compared to the house, the car is relatively small, I think'*). This counterbalancing also led to a swap of the size of the named objects in the visual scene, resulting in a total of six lists. Each list contained 72 experimental items (24 per condition) and 72 fillers that mentioned an object other than the medium object as the first noun, and gaze patterns different from the gaze patterns in the experimental items. Both the experimental items as well as the filler items contained the same number of true and false statements relative to the visual scene. Importantly, however, the truth value of the sentence was not revealed before the naming of the adjective at the end of the sentence. Hence, neither the gaze region nor the noun region were affected.

25% of the fillers (18) contained a manipulation of the first gaze cue instead of the second gaze cue. This subset of the fillers still started with a mentioning of the medium object as the first noun in the sentence. However, the first gaze cue was always directed toward the empty position. No incongruent first gaze cue was used in order to maintain the overall reliability of the gaze cues. The remaining fillers were of the same form as the experimental items with the difference that the first mentioned object was either the small or large (light/dark) object, followed by the naming of either of the remaining two objects. The gaze patterns performed on these fillers always started with a congruent gaze, as in the experimental items, followed by another congruent gaze toward the second named object half of the time (36) and a quarter of the time by either an incongruent or averted gaze cue (9/9). This distribution of gaze patterns throughout the experiment led to an overall ratio of congruent gaze actions of 70.8% (204). Every trial contained two gaze actions, one preceding the first noun and one preceding the second noun, the total number of gaze actions throughout the course of the experiment was 288 per list/participant. Another 17.7% (51) of the gaze actions were Averted and only 11.5% (33) of the gaze actions were Incongruent. This way, the validity of the gaze cue was kept high in order to avoid that participants would start to ignore the gaze cues altogether throughout the course of the experiment.

The stimuli were presented using the E-prime software (Version 2.0.10. Psychology Software Tools, Inc.). Each participant was seated in a sound-proof, electro-magnetically shielded chamber in front of a 24" Dell U2410 LCD monitor (resolution of 1280x1024 with a refresh rate of 75 Hz). The distance between the participant and the screen was always 100cm in order to keep all objects in a 5° visual angle from the center of the screen. This was done to minimize eye-movements throughout the experiment. While the participants were prepared for the recording, they were presented with all objects that occurred throughout the experiment and their naming. The Alex voice of the CereVoice TTS was also used for the naming of the objects. After this, participants were presented with written instructions and completed six practice trials. The items were pseudo randomized for each list and presented in 7 blocks with breaks after each block. The experiment lasted approximately 45 minutes.

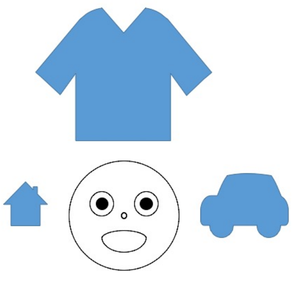


Figure 4.3: Example screen of experiment 1.

4.4.1 Participants

Forty-five right-handed native speakers of German (Mean age: 24; Age range: [18, 32]; SD: 3.39; Male: 8; Female: 37) took part in the ERP experiment. 15 participants were removed from the analysis due to their behavioral data (3) and too high numbers of eye artifacts (12). (For a concrete description of the removal see Section 4.4.2 on the following page Data Analysis.) Participants gave informed consent. All participants had normal or corrected-to-normal vision and had no hearing problems. All participants were compensated with €15 for their participation.

4.4.2 Data Analysis and Results

The EEG was recorded by 24 Ag/AgCl scalp electrodes[1] (actiCAP, BrainProducts) and amplified with a BrainAmp (BrainVision) amplifier. Electrodes were placed according to the 10-20 system (Sharbrough et al., 1991). Impedances were kept below 5kΩ. The ground electrode was placed at AFz. The signal was referenced online to the reference electrode FCz and digitized at a sampling rate of 500 Hz. The EEG files were re-referenced offline to the average of the mastoid electrodes. The horizontal electrooculogram (EOG) was monitored with two electrodes placed at the right and left outer canthi of each eye and the vertical EOG with two electrodes below both eyes paired with Fp1 and Fp2. During recording an anti-aliasing low-pass filter of 250Hz was used. The EEG data was band pass filtered offline at 0.01-40Hz in order to attenuate skin potentials and other low voltage changes as well as line noise and EMG noise (Luck, 2014). Single-participant averages were computed for a 900ms window per condition relative to the acoustical onset of the noun following the manipulated gaze cue and the manipulated gaze cue itself. All segments were aligned to a 100ms pre-stimulus baseline. The data was semi-automatically screened offline for electrode drifts, amplifier blocking, eye-movements and muscle artifacts.

Due to the nature of the task and the experimental setup containing various eye-movements performed by the displayed face, the number of eye artifacts was relatively high. Therefore, a threshold of 30% rejection rate per condition was set for participant exclusion (i.e., participants' data with more than 7 rejected trials out of 24 in one or more conditions were entirely removed). This led to the removal of 12 participants from the analysis. Overall, there were more eye-movements when the gaze cue appeared (15.3% of the trials) compared to when the corresponding noun was heard (6.9% of the trials). This is most likely due to a reflexive gaze following behavior when presented with a gaze cue. An analysis of variance (ANOVA) revealed that there were significantly more eye-movements in the noun region occurring in the Incongruent condition compared to the Congruent condition ($F(1, 44) = 7.52, p < .05, \eta^2 = 0.15$), as well as in the Averted condition compared to the Congruent condition ($F(1, 44) = 4.63, p < .05, \eta^2 = 0.095$). There was no significant difference between the number of eye-movements in the Averted condition and the Incongruent condition ($F(1, 44) = 1.08, p = .3$). The difference of both conditions to the Congruent condition can be explained by their nature: In the Incongruent condition, participants were presented with a word that contradicts the preceding gaze cue. In the Averted condition two possible targets were equally likely.

[1] This excludes the electrodes used for the electrooculogram and offline re-reference: Fp1, Fp2, T7, T8, TP9, TP10, PO9 and PO10.

Therefore, participants were possibly tempted to visually confirm the heard referent in the visual scene. A summary of the number of eye-movements per condition in the second noun region can be found in Table 4.5. Additionally, the data of 3 participants was removed due to their behavioral data. Participants' data was removed if they gave wrong answers to more than 10% of the questions. Overall, participants performed very well in the task with an average of 94.8% of correct answers. There was no difference in accuracy between conditions ($F(2, 58) = 0.96, p = .39$). After artifact rejection and participant exclusion 85% of the trials on average per participants were included in the analyses. Overall, the two criteria led to the removal of the data of 15 participants.

Table 4.5: Summary of the mean number of eye-movements in the noun region per condition in experiment 1, including the absolute values and the percentage.

	Gaze Region		Noun Region	
Condition	Mean	Percentage	Mean	Percentage
Congruent	3.8	15.8%	1.2	5.0%
Incongruent	3.8	15.8%	2	8.3%
Averted	3.4	14.2%	1.8	7.5%

The averaged data of the remaining 30 participants (Mean age: 23.7; Age range: [18, 32]; SD: 3.49; Female: 26) was exported using BrainVision Analyzer (Version 2.1) BESA export function. Two regions of interest were analyzed: The onset of the gaze cue toward the second noun, and the onset of that noun. The main focus of the analysis was put on the noun region. Analyses were performed in R by fitting Linear Mixed-Effects Models using the lme4 (Bates et al., 2015b) package (Version 1.1-10). β-Estimate, standard error and t-value and p-values are reported as well as confidence intervals. The p-values were extracted using the lmerTest package (Version 3.1-0). The confidence intervals were extracted utilizing the profile function of the stats package (Version 3.6.1). Additional to effects, model fit was used to determine the reported models by model comparison utilizing the anova function implemented in the stats package (Version 3.5.1).

4.4.2.1 Noun Region

In order to isolate the involved components and to establish the time-windows for the analyses, the approach utilizing difference waves was followed as proposed by Kappenman and Luck (2016). An Incongruent-minus-Congruent difference wave was created as well as an Averted-minus-Congruent difference wave that can be found in Figure 4.4.

Taking both difference waves into account, based on visual inspection, the time-window for the PMN lay between 150 and 300ms followed by a 300 – 450ms time-window for the N400 whereas the time-window for the P600 was established between 600 and 800ms which is consistent with previously established time-windows for the P600 (e.g., Burkhardt (2006);

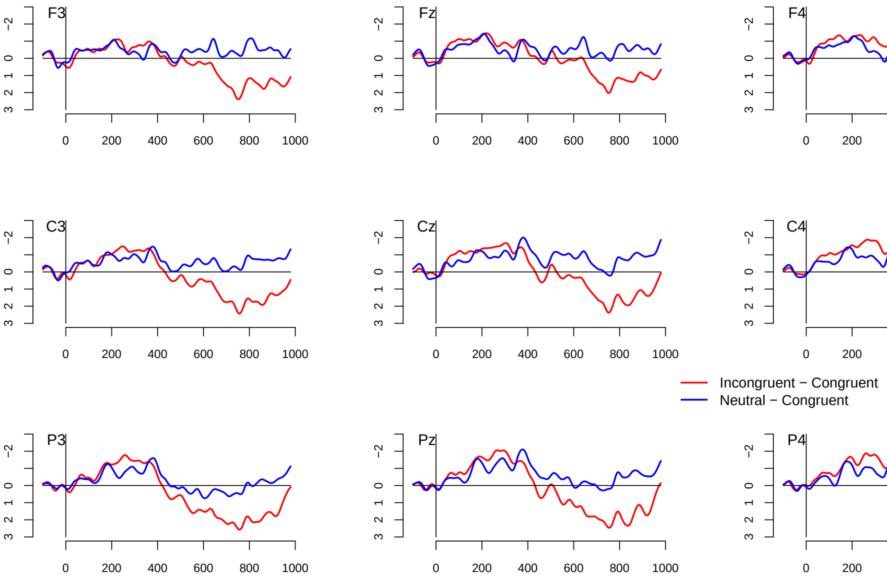


Figure 4.4: Difference waves of Incongruent-minus-Congruent (red) and Averted-minus-Congruent (blue) in the Noun region. The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

Brouwer et al. (2017)). Additionally, the time-windows were verified by running a moving time-window analysis with overlapping time-windows of 100ms length every 50ms. This analysis confirmed the time-windows indicated by the difference wave approach. Similar to other studies presenting auditory stimuli consisting of continuous speech (Connolly et al., 1990, 1992; O'Halloran et al., 1988; Hagoort and Brown, 2000), no N100-P200 complex was found, which is a usual response to the abrupt onset of auditory stimuli. The ERPs corresponding to the here reported findings can be found in Figure 4.5 on the next page.

In an initial step, models were fitted for the previously established time-windows with maximal random structure following (Barr et al., 2013). For the analysis of the noun region, Condition was included as a fixed factor with contrasts embedded in the factor itself to compare both the Incongruent and Averted conditions with the Congruent baseline (see Table 4.6 on the facing page for the contrast matrix as embedded in the factor Condition). Additionally, Longitude was added as a fixed effect in order to attest for scalp distribution of potential effects. For this factor, the electrodes were grouped into 3 Regions Of Interest (ROIs) for frontal (F3, Fz, F4, FC5, FC1, FC2, FC6), central (C3, Cz, C4, CP5, CP1, CP2, CP6) and posterior (P7, P3, Pz, P4, P8, O1, O2) distributions. Again, contrasts were directly embedded into the factor with two comparisons. Firstly, frontal against centro-parietal electrodes and, secondly, central against parietal electrodes (see Table 4.7 for the contrast

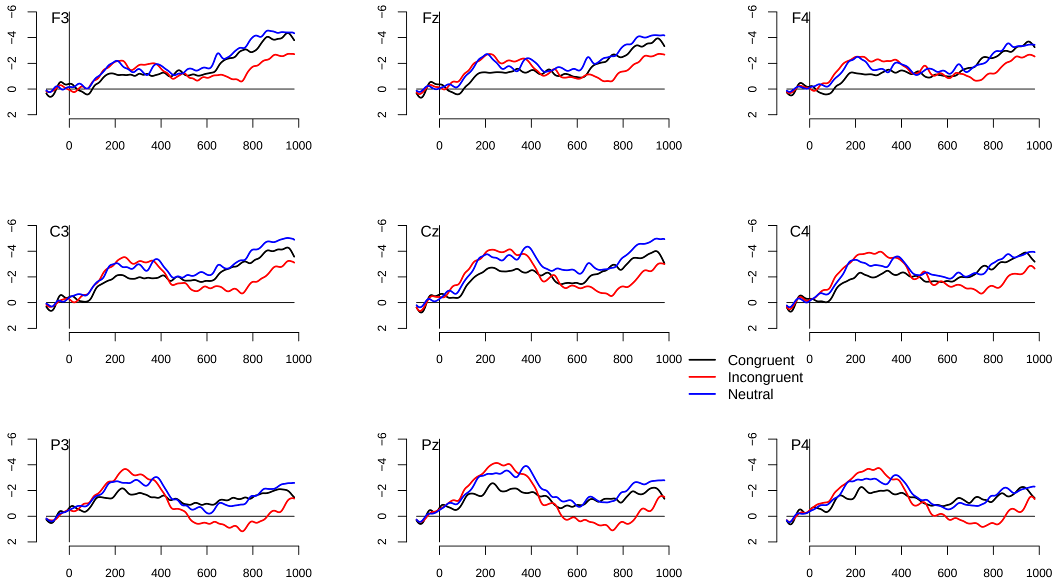


Figure 4.5: ERP time-locked to the Second Noun Onset in experiment 1 separated by the experimental conditions (Congruent (black), Incongruent (red) and Averted (blue)). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

matrix as embedded in the factor Longitude). The full random structure including the interaction between Condition and Longitude under subject led to singular fit warnings in every time-window. This possibly expresses that the models are over-fitted due to a random effects structure that is too complex to be supported by the data.

Table 4.6: Contrast matrix as embedded in the factor Condition in experiment 1.

Condition	c1	c2
Congruent	-1/3	-1/3
Incongruent	2/3	-1/3
Averted	-1/3	2/3

In order to determine how to simplify the random structure, a parsimonious mixed models approach was followed as described by Bates et al. (2015a), where factors or interactions between factors are dropped from the random structure according to the variance in the data they can account for. These contributions can be extracted utilizing a Principal Components Analysis (PCA) of the random-effects variance-covariance estimates from the mixed-effects model. Following the proposed approach to arrive at a parsimonious model, in a first step the zero-correlation model is computed. Additionally,

Table 4.7: Contrast matrix as embedded in the factor Longitude in experiment 1.

Longitude	l1	l2
Frontal	-2/3	0
Central	1/3	-1/2
Parietal	1/3	1/2

contrasts were no longer entered embedded in the respective factors, but entered separately. As this approach may result in different random structures for each of the fitted models and time-windows, each models R code is added in the respective sections. An example of the reduction process can be found in Appendix D on page 161.

PMN (150 – 300ms)

The zero-correlation parameter model was identified. Hence, correlations were reintroduced. This, however, led to a degenerate model. The PCA revealed three unidentified components. To address this issue, the correlations between the interactions were removed. The PCA of the resulting model, however, still showed one unidentified component. By removing the correlations of the contrast 'l2', the model did no longer show unidentified components. While this model could be reported, the parsimonious approach by Bates et al. (2015a) suggests that the model with the least parameters not leading to a significant decrease in model-fit should be preferred. As two components only explain roughly one percent of the variance each (1.09% and 1.01% respectively), removing the components with the least variance from the random structure is supported. Removing the two interaction contrasts 'c1:l2' and 'c2:l2' did not significantly influence model-fit. The code of the resulting final model was as follows:

```
final.PMN.model = lmer(eeg ~ 1 + (c1 + c2)*(l1 + l2) +
                    (1 + c1 + c2 + l1 | subject) +
                    (0 + l2 | subject) +
                    (0 + c1:l1 | subject) +
                    (0 + c2:l1 | subject),
                  REML=FALSE, data=EXP1.PMN.Data,
                  control = lmerControl(calc.derivs=FALSE),
                  na.action = na.omit)
```

The linear Mixed-Effects Model fitted to the time-window between 150 and 300ms as described above showed that both the Incongruent and Averted condition contained a significantly larger negativity compared to the Congruent baseline $((\beta = -1.45, SE = 0.42, t = -3.5, p < .01, CI = [-2.2966; -0.6091])$ and $(\beta =$

$-1.03, SE = 0.34, t = -3.1, p < .05, CI = [-1.7147; -0.3464]$) respectively). However, there was no significant interaction between either comparison and the Longitude factor, suggesting a global distribution of the effects. A summarized model output can be found in Table 4.8. The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure A.4, Figure A.5 (both on page 134) and Figure A.6 on page 135 respectively.

Table 4.8: Fixed Effects for the PMN time-window in experiment 1.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-2.2914	0.1739	-13.173	***
c1	-1.4528	0.4163	-3.490	**
c2	-1.0306	0.3376	-3.053	**
l1	-0.4207	0.1223	-3.440	**
l2	0.8443	0.1099	7.680	***
c1:l1	-0.2299	0.3302	-0.696	
c1:l2	0.3446	0.2070	1.665	
c2:l1	0.0652	0.2745	0.238	
c2:l2	0.3986	0.2070	1.926	

N400 (300 – 450ms)

Following the same approach of model reduction for the PMN time-window as described above, the code of the resulting final model for the N400 time-window was as follows:

```
final.N400.model = lmer(eeg ~ 1 + (c1 + c2)*(l1 + l2) +
                   (1 + c1 + c2 + l2 | subject) +
                   (0 + l1 | subject) +
                   (0 + c1:l1 | subject) +
                   (0 + c2:l1 | subject),
                 REML=FALSE, data=EXP1.N400.Data,
                 control = lmerControl(calc.derivs=FALSE),
                 na.action = na.omit)
```

The linear Mixed-Effects Model fitted for the time-window between 300 and 450ms also showed that both the Incongruent and Averted condition contain a significantly larger negativity compared to the baseline (($\beta = -0.84, SE = 0.38, t = -2.2, p < .05, CI = [-1.6123; -0.0685]$) and ($\beta = -0.91, SE = 0.45, t = -2.03, p <$

$.05, CI = [-1.8288; -0.0012]$) respectively). The significant interactions of 'l2' with both Congruency comparisons suggests a more fronto-central distribution of the effects for both conditions (($\beta = 0.53, SE = 0.21, t = 2.5, p < .05, CI = [0.1160; 0.9398]$) and ($\beta = 0.48, SE = 0.21, t = 2.26, p < .05, CI = [0.0633; 0.8870]$) respectively). The entire model output can be found in Table 4.9. The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure A.7, Figure A.8 (both on page 136) and Figure A.9 on page 137 respectively.

Table 4.9: Fixed Effects for N400 time-window in experiment 1.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.70749	0.28428	-6.006	***
c1	-0.84039	0.38128	-2.204	*
c2	-0.91499	0.45140	-2.027	*
l1	-0.10979	0.16375	-0.670	
l2	1.00466	0.13862	7.247	***
c1:l1	0.21553	0.34065	0.633	
c1:l2	0.52788	0.21009	2.513	*
c2:l1	0.06987	0.32814	0.213	
c2:l2	0.47514	0.21009	2.262	*

In order to assure the validity of a split between the PMN and N400 time-window, additionally, a linear Mixed-Effects Model was fitted for the time-window from 250-350 ms, overarching the 'gap' between the two effects. Following the same approach of model reduction as for the previous time-windows, the code of the resulting final model for this time-window was as follows:

```
final.gap.model = lmer(eeg ~ 1 + (c1 + c2)*(l1 + l2) +
                    (1 + c1 + c2 + l2 | subject) +
                    (0 + l1 | subject) +
                    (0 + c1:l1 | subject) +
                    (0 + c1:l2 | subject) +
                    (0 + c2:l1 | subject) +
                    (0 + c2:l2 | subject),
                 REML=FALSE, data=EXP1.gap.Data,
                 control = lmerControl(calc.derivs=FALSE),
                 na.action = na.omit)
```

There was a significantly stronger negativity for only the Incongruent condition compared to the Congruent baseline ($\beta = -1.24, SE = 0.46, t = -2.7, p < .05, CI = [-2.1757; -0.3003]$). However, neither the comparison of the Averted condition with the Congruent baseline, nor any of the interactions were significant. This supports the claim that the PMN and N400 effects are to be regarded as separate effects. The corresponding model output can be found in Table 4.10.

Table 4.10: Fixed Effects for 250-350ms time-window in experiment 1.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-2.10199	0.23514	-8.939	***
c1	-1.23799	0.46320	-2.673	*
c2	-0.80922	0.42122	-1.921	
l1	-0.41928	0.15058	-2.784	**
l2	0.85686	0.12425	6.896	***
c1:l1	-0.15560	0.34056	-0.457	
c1:l2	0.42496	0.21972	1.934	
c2:l1	-0.07927	0.30681	-0.258	
c2:l2	0.40977	0.23265	1.761	

P600 (600 – 800ms)

Following the same approach of model reduction as for the previous time-windows, the code of the resulting final model for the P600 time-window was as follows:

```
final.P600.model = lmer(eeg ~ 1 + (c1 + c2)*(l1 + l2) +
                    (1 + c1 + c2 + l2 | subject) +
                    (0 + l1 | subject) +
                    (0 + c1:l1 | subject) +
                    (0 + c1:l2 | subject) +
                    (0 + c2:l1 | subject) +
                    (0 + c2:l2 | subject),
                  REML=FALSE, data=EXP1.P600.Data,
                  control = lmerControl(calc.derivs=FALSE),
                  na.action = na.omit)
```

The linear Mixed-Effects Model fitted to the time-window between 600 and 800ms revealed an effect for only the Incongruent condition compared to the Congruent baseline ($\beta = 1.07, SE = 0.48, t = 2.2, p < .05, CI = [0.0866; 2.0490]$), that is strongest in centro-parietal sites ($\beta = 0.85, SE = 0.42, t = 2.0, p < .05, CI = [-0.0011; 1.6952]$).

Additionally, the Averted condition, although not showing a global effect ($\beta = -0.42, SE = 0.49, t = -0.83, p > .05$), showed a stronger frontal negativity compared to the Congruent baseline ($\beta = 0.75, SE = 0.35, t = 2.1, p < .05, CI = [0.0335; 1.4571]$) that could also be observed in central sites ($\beta = 0.65, SE = 0.25, t = 2.6, p < .01, CI = [0.1673; 1.1302]$). The entire model output can be found in Table 4.11. The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure A.10, Figure A.11 (both on page 138) and Figure A.12 on page 139 respectively.

Table 4.11: Fixed Effects for P600 time-window in experiment 1.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.0685	0.4179	-2.557	*
c1	1.0678	0.4847	2.203	*
c2	-0.4165	0.4994	-0.834	
l1	1.2065	0.2876	4.195	***
l2	1.5849	0.2254	7.032	***
c1:l1	0.8470	0.4194	2.020	*
c1:l2	0.4210	0.2456	1.714	
c2:l1	0.7453	0.3520	2.117	*
c2:l2	0.6487	0.2456	2.642	**

Table 4.12 summarizes the results of the models for the reported time-windows.

Table 4.12: Summary of the model results in experiment 1. Significance indicates results according to the lmerTest package.
($-$ - $p > .05$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Comparison	PMN	gap	N400	P600
c1	$**$	$*$	$*$	$*$
c2	$**$	$-$	$*$	$-$
c1:l1	$-$	$-$	$-$	$-$
c1:l2	$-$	$-$	$*$	$-$
c2:l1	$-$	$-$	$-$	$*$
c2:l2	$-$	$-$	$*$	$**$

4.4.2.2 Gaze Cue preceding the Noun

As the results of the analyses of the noun region showed clear effects of the congruity of the preceding gaze cue, it could be argued that the utilization of the gaze cue — as for example the integration of the gazed-at referent — should be observable in the gaze region itself. However, the split by condition as used in the noun region is not as meaningful on the gaze: As participants have no prior information on which referent is going to be gazed at or subsequently named, Incongruent gaze as well as Congruent gaze are, at the point of their occurrence, indistinguishable. Only the subsequently named referent renders the gaze to an object to be Congruent or Incongruent. Therefore, in the gaze region, the Congruent and Incongruent conditions were collapsed and relabled as Informative, as a gaze-cued object possibly provides information about the continuation of the sentence. In contrast, the Averted condition, with the gaze being directed to the bottom of the screen away from the objects present in the scene, is labeled as Uninformative as no information about the curse of the sentence can be derived from it. Hence, for the analysis of the gaze region, Informativity of the gaze cue was included as a fixed factor with contrasts embedded in the factor itself to compare Informative, object-directed gaze (Congruent/Incongruent) and Uninformative (Averted) gaze with one another (see Table 4.13 for the contrast matrix as embedded in the factor Informativity).

Table 4.13: Contrast matrix as embedded in the factor Gaze Informativity in experiment 1.

Informativity	i1
Informative (object-directed)	-1/2
Uninformative (averted)	1/2

Similar to the approach in the analyses of the noun region, initially, the difference-waves approach (Kappenman and Luck, 2016) was utilized in order to identify the time-windows of possible effects. An Uninformative-minus-Informative difference wave was created that can be found in Figure 4.6, and supplemented with a moving time-window analysis. When combining the two approaches a fronto-central positivity lasting from 300 to 800ms was revealed. A model fitted to this time window confirmed this ($\beta = 0.93, SE = 0.39, t = 2.3, p < .05, CI = [0.1254; 1.7358]$). Further, the model showed that the effect is more fronto-centrally distributed ($\beta = -0.83, SE = 0.33, t = -2.5, p < .05, CI = [-1.4787; -0.1717]$).

However, as the presented scene arranged the objects on the horizontal and vertical axis across the eyes of the face, and as Uninformative gaze always is directed downward, it is possible that this positioning confounds the reported effects. Therefore, additionally, the analysis of gaze Informativity was followed by an analysis of gaze Direction. This factor was created in a way so that first vertical and horizontal gaze cues were compared with one-another, followed by comparisons within those two levels. In more detail, this means a comparison between leftward and rightward gaze cues as well as upward and downward gaze cues. The latter comparison, hence, is a subset of the previous analysis of

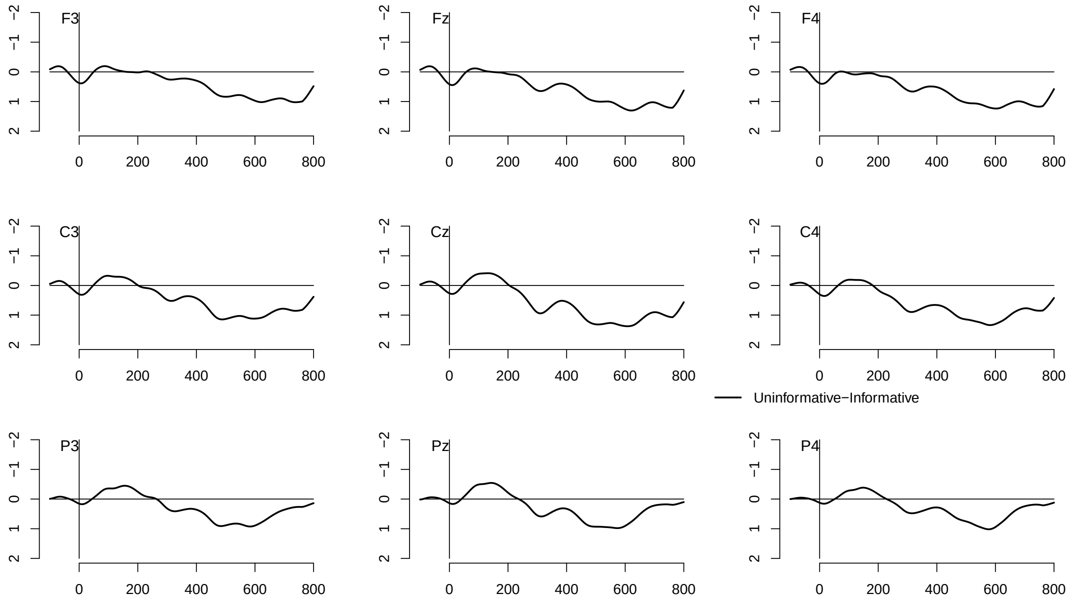


Figure 4.6: Difference waves of Uninformative-minus-Informative gaze cues (black) in the Gaze region. The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

gaze Informativity as it compares the Uninformative gaze cues (Averted/Downward) with a subset of the Informative gaze cues that directed the gaze to an object above the face.

Again, in a first step a Vertical-minus-Horizontal difference wave was created as well as Left-minus-Right and Up-minus-Down difference waves that can be found in Figure 4.7, in order to check for possible effects due to gaze Direction. This was again supplemented with a moving time-window analysis. In combination, these two approaches reveal a more positive deflection for Vertical gaze cues compared to Horizontal gaze cues starting at 250ms. Additionally, a fronto-central positivity was found for Downward gaze cues compared to Upward gaze cues between 400 and 700ms. This positivity falls into a sub-time-window of the positivity found for the Uninformative gaze cues compared to the Informative gaze cues (300-800ms). There was no indication for a difference between Leftward and Rightward gaze cues.

A linear Mixed-Effects Models for the previously established time-windows was fitted following the same approach as for the other described models. The analysis included Direction as a fixed factor with nested contrasts embedded in the factor itself to compare Vertical with Horizontal gaze cues as well as to compare within the two levels (Upward compared to Downward and Leftward compared to Rightward gaze cues). Table 4.14 on the facing page shows the embedded contrast matrix.

Additionally, Longitude was added as a fixed effect in order to attest for scalp

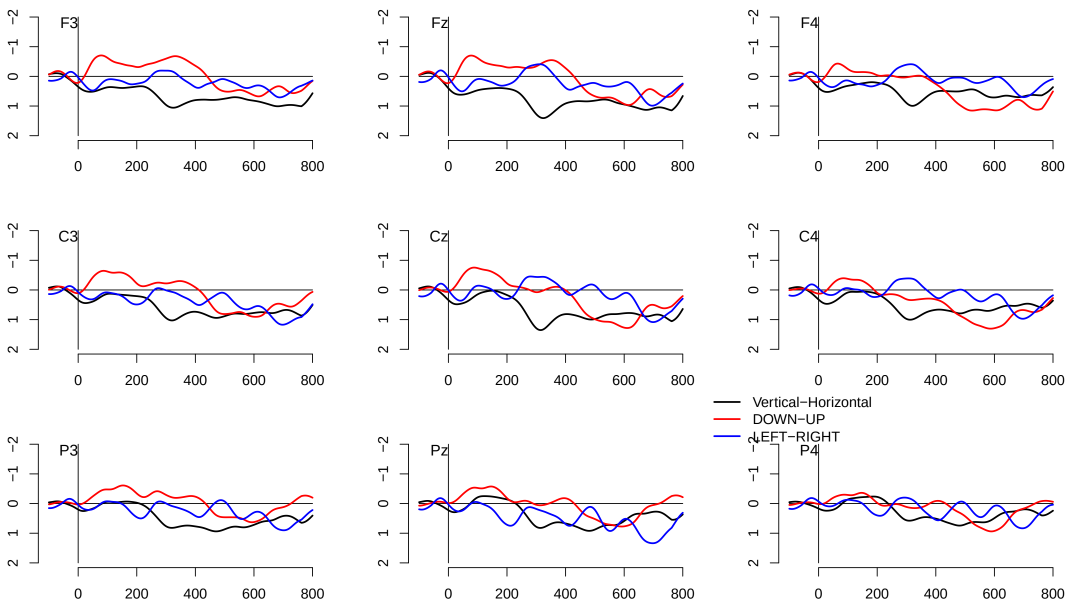


Figure 4.7: Difference waves of Vertical-minus-Horizontal (black), Downward-minus-Upward (red), and Leftward-minus-Rightward (blue) Gaze Cues in experiment 1. The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

Table 4.14: Contrast matrix as embedded in the factor Gaze Direction in experiment 1.

Direction	d1	d2	d3
UP	-1/2	-1/2	0
DOWN	-1/2	1/2	0
LEFT	1/2	0	-1/2
RIGHT	1/2	0	1/2

distribution of potential effects. The coding of this factor was identical to its utilization in the noun region (3 ROIs for frontal (F3, Fz, F4, FC5, FC1, FC2, FC6), central (C3, Cz, C4, CP5, CP1, CP2, CP6) and posterior (P7, P3, Pz, P4, P8, O1, O2) distributions with embedded contrasts as shown in Table 4.7 on page 54). The full random structure including the interaction between Direction and Longitude under subject led to singular fit warnings in both time-windows. The same approach running PCAs to simplify the random structure as for the noun region was utilized. The R code used for the final model was as follows:

```
final.GazeDirection.model = lmer(eeg ~ 1 + (d1 + d2 + d3)*(l1 + l2) +
                (1 + d1 + d2 + d3 + l2 | subject) +
                (0 + l1 | subject) +
                (0 + d1:l1 | subject) +
                (0 + d1:l2 | subject) +
                (0 + d2:l1 | subject) +
                (0 + d3:l1 | subject),
              REML=FALSE, data=EXP1.GazeDirection.Data,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
```

The linear Mixed-Effects Model fitted to the time window from 250-800ms revealed a long lasting globally distributed positivity for vertical gaze cues compared to horizontal gaze cues ($\beta = -0.99, SE = 0.32, t = -3.1, p < .01, CI = [-1.6337; -0.3427]$) (see Figure 4.8 for comparison), that was strongest in fronto-central electrodes ($\beta = 0.51, SE = 0.21, t = 2.4, p < .05, CI = [0.0805; 0.9420]$). Additionally, there was an also fronto-centrally distributed significantly stronger positivity for downward as compared to upward gaze cues ($\beta = 0.71, SE = 0.32, t = 2.2, p < .05, CI = [0.0693; 1.3408]$).

Table 4.15: Fixed Effects for Gaze Direction in experiment 1.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	0.6331	0.2645	2.394	*
d1	-0.9884	0.3185	-3.104	**
d2	-0.3404	0.5357	-0.636	
d3	-0.5039	0.4326	-1.165	
l1	-0.6517	0.1891	-3.447	**
l2	-0.8271	0.1425	-5.805	***
d1:l1	0.2921	0.2554	1.144	
d1:l2	0.5113	0.2138	2.392	*
d2:l1	0.4778	0.4307	1.109	
d2:l2	0.7050	0.3242	2.174	*
d3:l1	-0.2374	0.2576	-0.922	
d3:l2	0.0108	0.2647	0.041	

The linear Mixed-Effects Model fitted to the time window from 400-700ms had the same random structure as the model for the longer time-window reported above. It showed the same fronto-central distributed positivity for Downward gaze cues compared to Upward gaze cues ($\beta = 0.8, SE = 0.35, t = 2.3, p < .05, CI = [-1.6574; -0.3229]$) (see Figure 4.9 for comparison) that was also reported for the longer time-window. The global positivity

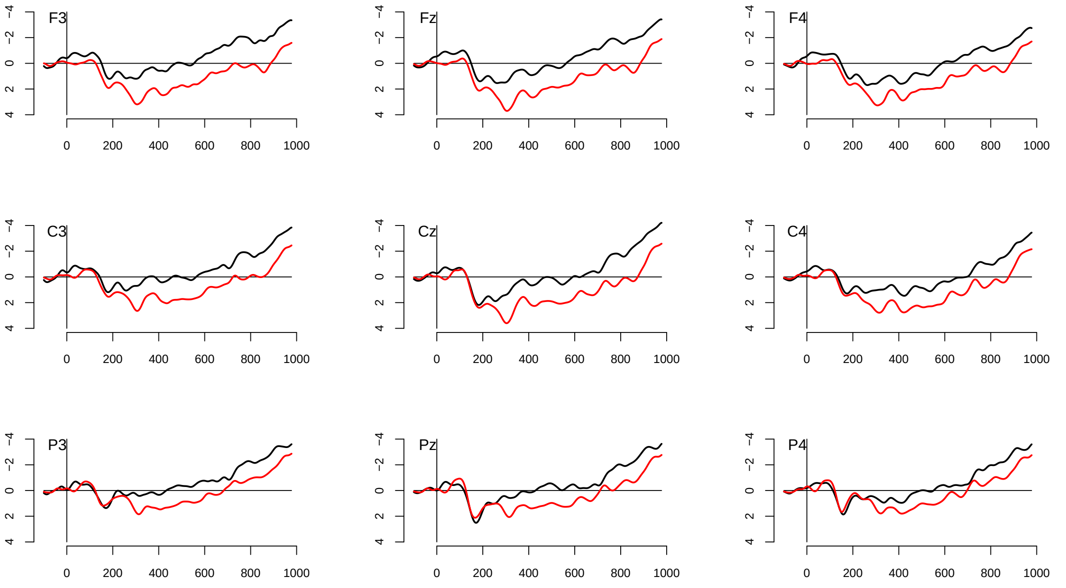


Figure 4.8: ERP time-locked to the gaze cue Onset separated by Horizontal (black) and Vertical (red) gaze cues in experiment 1. The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

for Vertical compared to Horizontal gaze cues (d1) can also be found in this time-window ($\beta = -0.99, SE = 0.33, t = -3.0, p < .01, CI = [0.1354; 1.4683]$). In this time-window, the effect, however, is not stronger in any specific scalp region, indicating a global distribution.

Although the analysis of the 400-700ms time-window suggests that there is an effect of gaze Informativity, the distribution of the objects on the screen also led to a clear confound introduced by the long-lasting (250-800ms) effect found for Vertical gaze cues as compared to Horizontal gaze cues. As there was no case where the Uninformative (Averted) condition contained a gaze cue other than Downward, the results should — to at least some extent — be considered with caution.

4.4.3 Preliminary Discussion

4.4.3.1 Noun Region

Research from Koornneef and Van Berkum (2006) and Van Berkum et al. (2007) suggests that comprehenders generate expectations about the unfolding sentence based on the previously gathered information that they integrated in a situation model (Zwaan and Radvansky, 1998). Various studies further suggest that not only linguistic information is used to form such expectations about upcoming sentence content but also visual

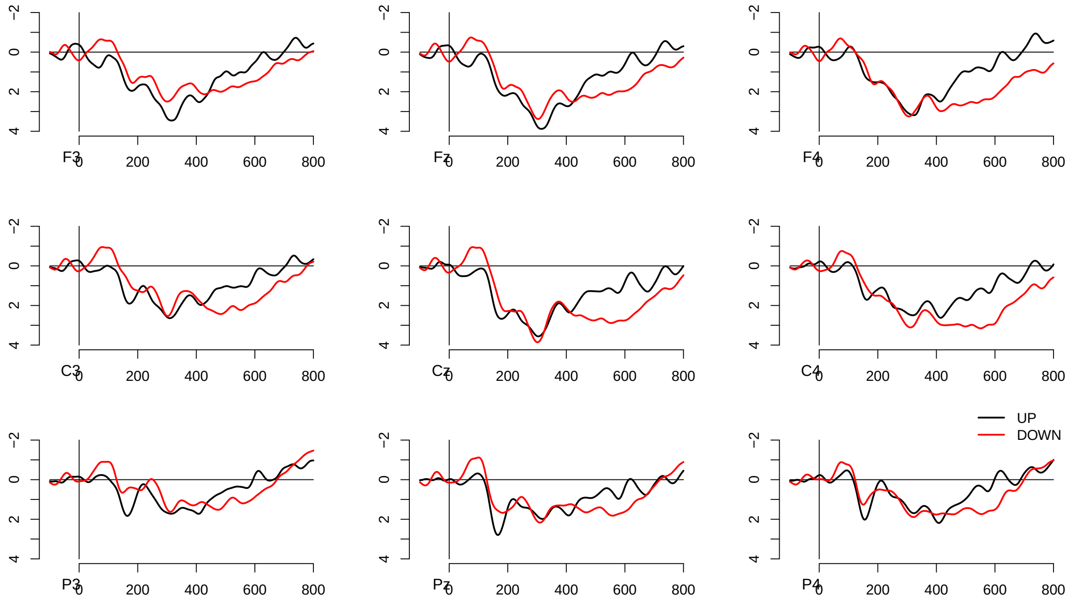


Figure 4.9: ERP time-locked to the gaze cue Onset separated by Upward (black) and Downward (red) gaze cues in experiment 1. The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

information provided by the combination of the provided scene and gaze cues (Staudte and Crocker, 2010, 2011; Staudte et al., 2014; Ferreira et al., 2013). It is therefore reasonable to hypothesize that this visual information also contributes to constructing the situation model. The earlier peak (150 - 300ms) is interpreted as a PMN reflecting an auditory matching process that is driven by the amount of information the incoming phoneme conveys. This results in an attenuated PMN for the Congruent gaze condition only as no new information is provided beyond that of the gaze cue. In both the Averted and Incongruent condition however, the phoneme provides additional information. In the former case by supporting only one of two possible referents, and in the latter case by mismatching with the expected word form due to the highly lexically specific expectations the gaze cues elicit. The N400 (300 - 450ms) is interpreted to be reflecting a word's retrieval cost, influenced by how strongly supported or expected a word is given a visual context, such as situated gaze. This is in line with the effects found for the Incongruent and Averted condition in comparison to the Congruent condition as both conditions entail higher retrieval cost: Either because no word was retrieved so far (Averted) or because the already retrieved word does not fit the perceived word, requiring the retrieval of the actual referent (Incongruent). Finally, the P600 is interpreted to be reflecting the cost of revising the situational model formed on prior contextual information. This is only the case in the Incongruent condition, where the gazed at object was integrated in the mental representation, which leads to the necessity to revise that representation. Taken together, the results provide support for the situational integration account over the prominence account. While the effects in the N400 time-window could equally be explained by both accounts, the other time-windows provide evidence for both the prediction of a specific word form (PMN) and the integration of the gazed-at object in the mental situational representation (P600).

4.4.3.2 Gaze Region

Based on the confounding positioning of the objects, a discussion of the effects found in this region should be taken with caution. The positioning of the objects on the vertical and horizontal axis through the eyes of the displayed face showed clear differences in the ERP wave forms, so that underlying effects of object-directed informative gaze in comparison to averted uninformative gaze might be obscured or even falsely induced. Although the comparison of upward (Informative) and downward (Uninformative) gaze cues showed an effect within the vertical axis, it is not entirely clear whether upward and downward gaze cues induce different effects independent of the presence or absence of an object in this position. Hence, in order to more reliably investigate the effects of gaze informativity, a follow-up experiment was conducted that addresses this shortcoming. Additionally, the follow-up experiment aimed to ensure that the results in the noun region from Experiment 1 were replicable and robust. The precise description of the changes made for the follow-up experiment will be discussed in greater detail in the following chapter.

4.5 Experiment 2

In this follow-up experiment, the positioning of the objects relative to the face was adjusted. As the cross-wise positioning in Experiment 1 (up, down, left and right of the face) led to significantly different ERP responses when participants were presented with the gaze cues, the objects were positioned diagonally in this experiment. All objects in Experiment 2 were positioned 30° above and below the horizontal axis through the eyes of the face (see Figure 4.10 on the facing page for comparison). The new positioning of the objects also required an increased distance between the participant and the screen of 114cm in order to keep all objects in a 5° visual angle from the center of the screen. In line with the change in object positioning, the empty position also rotated through the four possible positions in Experiment 2 instead of being always below the face as in Experiment 1. The same objects as in Experiment 1 were used with one change: As the onsets for the words Stern (star) and Stiefel (boot) were similar in the first experiment, the star was exchanged with the equally well-performing object Mond (moon). 'Well-performing' is defined as similar complexity ratings in combination with all participants naming the object identically in the pre-test. All objects used in Experiment 2 are summarized in Table 4.16.

Table 4.16: Summary of the adjusted objects presented in Experiment 2 with their English translation seperated by gramatical gender

Masculine	Feminine	Neuter
Baum (tree)	Blume (flower)	Auto (car)
Blitz (bolt)	Brezel (pretzel)	Blatt (leaf)
Fisch (fish)	Gießkanne (watering can)	Boot (boat)
Handschuh (glove)	Hand (hand)	Flugzeug (airplane)
Hut (hat)	Lampe (lamp)	Haus (house)
Mond (moon)	Maske (mask)	Kreuz (cross)
Stiefel (boot)	Tasche (bag)	Rad (wheel)
Tisch (table)	Wolke (cloud)	T-Shirt (t-shirt)

The sentences presented to the participants were of the same form as in Experiment 1 (e.g., *"Verglichen mit dem Haus, ist das Auto verhältnismäßig klein, denke ich"; 'Compared to the house, the car is relatively small, I think'*). Figure 4.10 on the facing page shows a time-line of a Congruent trial in Experiment 2.

It is also perhaps debatable how 'uninformative' a gaze cue toward a position is, even if that position does not contain an object (Averted). In order to address this concern, an additional version of an 'uninformative' gaze cue was introduced to the three conditions as present in the first experiment. Different from the previous Averted gaze cue, this version moved the eyes of the face back to the straight gaze position instead of directing gaze to the empty position. The new gaze cue redirected straight toward

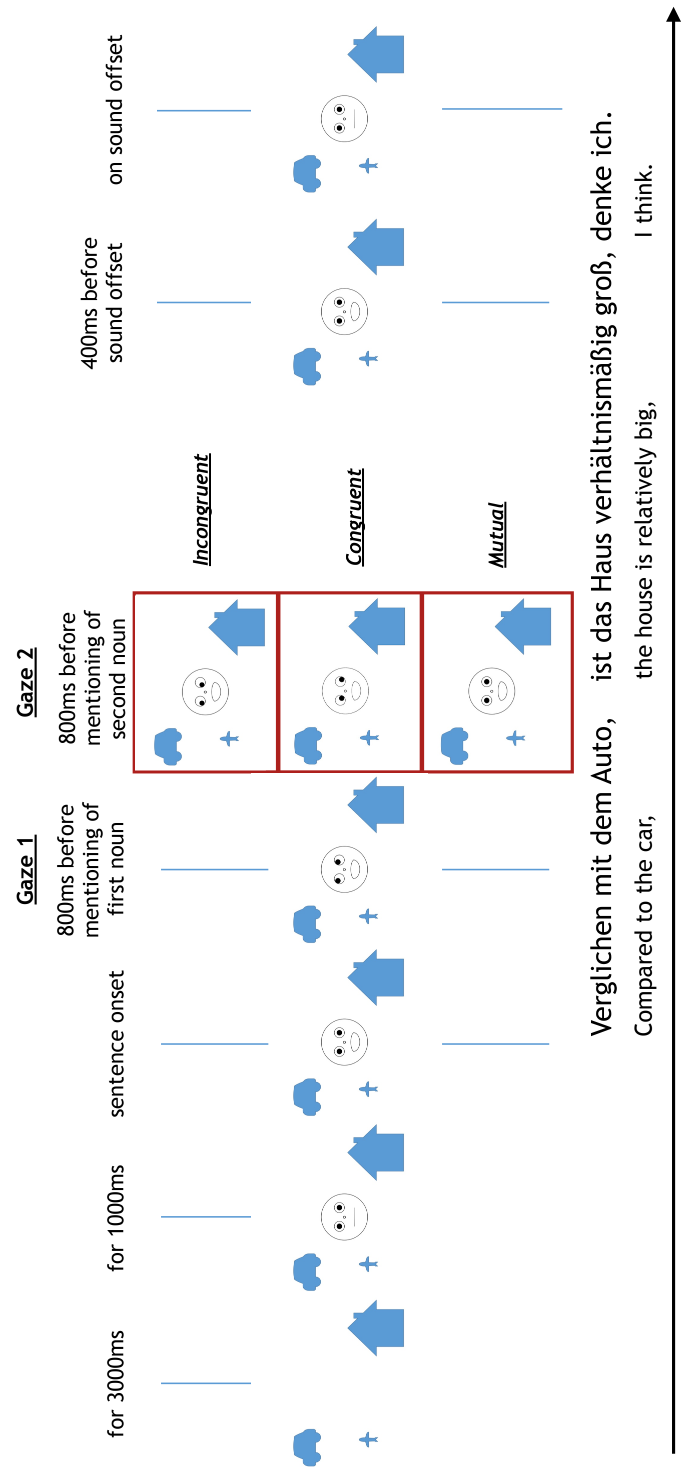


Figure 4.10: Timeline of an experimental trial in Experiment 2.

the listener will be referred to as Mutual. Both Averted and Mutual gaze are considered to be 'uninformative', in contrast with object-directed gaze (Congruent/Incongruent). The Averted gaze cue from Experiment 1 with the gaze being directed to the empty position was demoted to a control condition that was added as a filler type. In order to still achieve comparable data, the number of this filler type was matched with the number of items per experimental condition. Each item in Experiment 2 appeared in three conditions (Congruent / Incongruent / Mutual) with an additional filler type that provides comparability between the two experiments (Averted). This led to three lists using a Latin square design. Additionally, versions of those lists were created that were counterbalanced for the truth value of a sentence. This means that for the scene as displayed in Figure 4.10 on the previous page, two versions of the sentence were existent: 1) *'compared to the house, the car is relatively small, I think'* and 2) *'compared to the house, the car is relatively big, I think'*. This led to a total of six lists. As no effect of realism was found in the first experiment, this was no longer counterbalanced across lists in the second experiment. It should be noted though that the number of 'realistic' and 'unrealistic' sentences within a list was still balanced.

Each list contained 126 experimental items (42 per condition) and 126 fillers. 42 fillers were created identical to the Averted condition in Experiment 1 to retain comparability. This means that the first gaze was always congruent toward the object named first in the sentence followed by a gaze toward the empty position 800ms before the mentioning of the second noun. In the remaining fillers (84) the first gaze cue was manipulated followed by a Congruent second gaze cue. In these fillers, the first gaze cue was either directed toward the empty position (42) or toward the object that remained unmentioned throughout the continuation of the sentence (42). A summary of the overall distribution of gaze patterns can be seen in Table 4.17.

Table 4.17: Summary of the distribution of gaze patterns in Experiment 2 split by occurrence in experimental and filler trials

Experimental		
First gaze cue	Second gaze cue	No. of occurrences per list
Congruent	Congruent	42
Congruent	Incongruent	42
Congruent	Mutual	42
Fillers		
First gaze cue	Second gaze cue	No. of occurrences per list
Congruent	Averted	42
Averted	Congruent	42
Incongruent	Congruent	42

This distribution of gaze actions additionally led to a slight adjustment of the reliability of the gaze cue. The overall percentage of Congruent gaze actions was lowered from 70.8% to 58.3%, whereas the Incongruent and uninformative gaze actions – represented by Mutual and Averted gaze actions – both were increased (from 11.5% to 16.6% and from 17.7% to 25% respectively). These adjustments were used to test for the robustness of the effects found in the first experiment.

In line with the increased number of trials per participant, the experiment lasted approximately 60 minutes. The items were pseudo randomized for each list and presented in 5 blocks with breaks after each block. The participants task remained unchanged from the task in Experiment 1.

I hypothesized that the modulations in the PMN, N400 and P600 time-windows found in the first experiment would be replicated. Specifically, a stronger modulation of the PMN and N400 was expected in the Mutual and Incongruent conditions compared to the Congruent condition related to the expectability of the noun given the visual context. Additionally, a P600 was predicted in the Incongruent condition related to the necessity to revise the mental model formed by utilizing the visual input.

4.5.1 Participants

Forty-four right-handed native speakers of German, who did not participate in Experiment 1 (Mean age: 24.6; Age range: [18, 35]; SD: 3.65; Female: 34), took part in the ERP experiment. 14 participants were removed from the analysis due to their behavioral data (4), technical errors (3) and too high numbers of eye artifacts (7). (For a concrete description of the removal see Section 4.5.2 Data Analysis.) Participants gave informed consent. All participants had normal or corrected-to-normal vision and had no hearing problems. All participants were compensated with €15 for their participation.

4.5.2 Data Analysis and Results

The technical setup and EEG recording sites were the same as in Experiment 1 (see Section 4.4.2 on page 50 for comparison).

The 30% threshold for the rejection rate per condition was kept for participant exclusion due to eye-movements and other artifacts. This led to the removal of 7 participants from the analysis. In this experiment, participants performed less eye movements in the noun region on average compared to the previous experiment. The number of eye movements in the gaze region preceding the target noun however remained approximately the same (see table 4.18 on the next page for a summary of the eye movements in this experiment and 4.5 on page 51 for the respective Table for Experiment 1). With the reduced number of eye movements in the noun region, the significant difference of the Incongruent and Mutual condition compared to the Congruent condition found in Experiment 1 disappeared

($(F(1, 41) = 1.8, p = .18)$ and $(F(1, 41) = 0.18, p = .67)$. It is possible that this hints toward an overall easier monitoring of the objects in the diagonal positioning compared to the cross-wise positioning in Experiment 1. Additionally, the data of 4 participants was removed due to their behavioral data with more than 10% of wrong answers to the question. Again, remaining participants performed very well in the task with an average of 97.4% of correct answers. As in Experiment 1, there was no difference in accuracy between conditions $(F(2, 58) = 1.98, p = .15)$. Another 3 participants had to be removed due to technical errors. Overall, the three criteria led to the removal of the data of 14 participants. After artifact rejection and participant exclusion 94.3% of the trials on average per participants were included in the analyses. The analysis of the data of the remaining 30 participants (Mean age: 24.3; Age range: [19, 33]; SD: 3.2; Female: 24) was conducted in the same way as in Experiment 1 (see section 4.4.2 on page 50 for comparison).

Table 4.18: Summary of the mean number of eye-movements in the gaze and the noun region per condition in Experiment 2, including the absolute values and the percentage.

	Gaze Region		Noun Region	
Condition	Mean	Percentage	Mean	Percentage
Congruent	7.6	15.8%	1.8	3.7%
Incongruent	7.3	15.2%	2.4	5.0%
Mutual	7.4	15.4%	1.9	4.0%
Averted	8.9	18.5%	1.7	3.5%

4.5.2.1 Noun Region

Linear Mixed-Effects Models were fitted for the differences between the three experimental conditions (Congruent, Incongruent, Mutual) as a single 3-level factor time locked to the onset of the second noun, for which the preceding gaze cue was manipulated. For the analysis, the same time-windows were used as established in the first experiment. The embedded contrast matrix resembles the matrix as used in Experiment 1 with a swap of the Averted condition (Experiment 1) with the Mutual condition (Experiment 2). More precisely, the contrasts encoded a comparison of the Incongruent condition with the Congruent baseline and a comparison of the Mutual condition with the Congruent baseline (for comparison see Table 4.6 on page 53). Additionally, separate models were fitted containing only a 2-level condition factor for the comparison between Congruent and Averted gaze in order to allow for a direct comparison between the two experiments. The embedded contrast matrix can be found in Table 4.19 on the next page, while the corresponding ERPs can be found in Figure 4.11 on the facing page.

It is to be kept in mind that the Mutual condition stands for a straight gaze toward the listener, whereas the Averted condition stands for a gaze toward the empty position as it

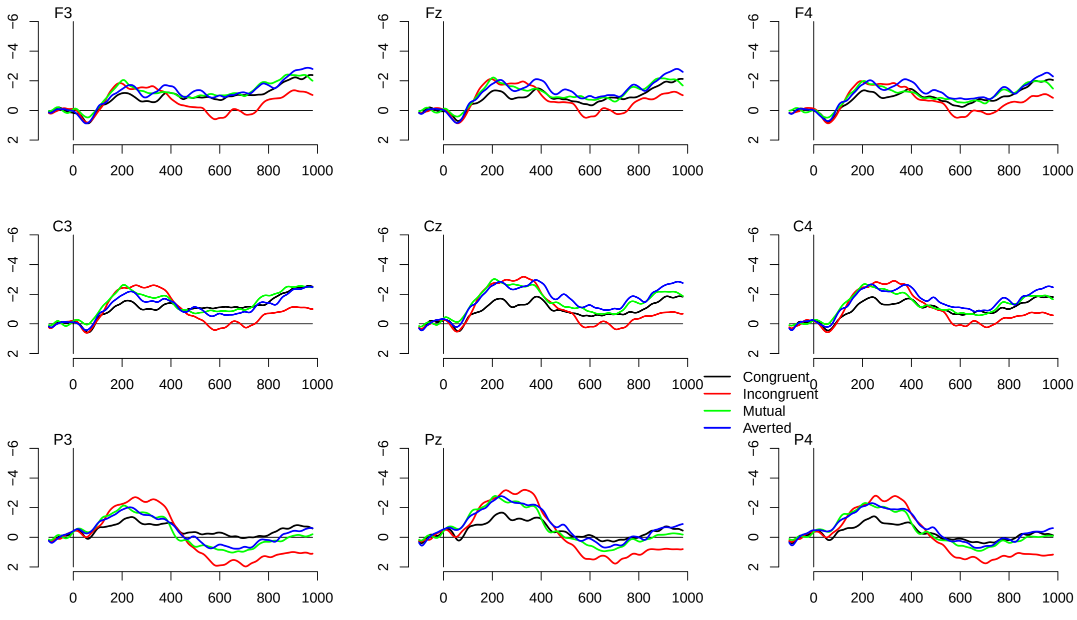


Figure 4.11: ERPs time-locked to the Second Noun Onset in Experiment 2 separated by the Experimental Conditions (Congruent (black), Incongruent (red) and Mutual (green)) and the Averted control (blue). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

Table 4.19: Contrast matrix as embedded in the factor Condition in the Control comparison in Experiment 2.

Control	co1
Congruent	-1/2
Averted	1/2

was the case in the first experiment. Additionally, Longitude was added as a fixed effect in order to attest for scalp distribution of potential effects. For this factor, the electrodes were again grouped into 3 Regions Of Interest (ROIs) for frontal (F3, Fz, F4, FC5, FC1, FC2, FC6), central (C3, Cz, C4, CP5, CP1, CP2, CP6) and posterior (P7, P3, Pz, P4, P8, O1, O2) distributions. The embedded contrast matrix was the same as in Experiment 1 (see Table 4.7 on page 54 for comparison).

The models were fitted in the same manner as in Experiment 1: Initially a model with maximal random structure following (Barr et al., 2013) was fitted and followed by the parsimonious mixed models approach as described by (Bates et al., 2015a) utilizing PCAs to reduce the complexity of the models. A full description of the approach can be found in

Subsection 4.4.2.1 on page 51. The R code used for each fitted model will be reported in the respective section.

PMN (150 – 300ms)

Following the same approach of model reduction as described for the first experiment, the code of the resulting final model for the PMN time-window was as follows:

```
final.PMN.model = lmer(eeg ~ 1 + (c1 + c2)*(l1 + l2) +
                 (1 + c1 + c2 + l1 + l2 + c1:l1 + c2:l1 | subject) +
                 (0 + c1:l2 | subject) +
                 (0 + c2:l2 | subject),
               REML=FALSE, data=EXP2.PMN.Data,
               control = lmerControl(calc.derivs=FALSE),
               na.action = na.omit)
```

In the PMN time window from 150 – 300ms, a significantly increased negativity was found for both the Incongruent and Mutual condition compared to the Congruent baseline (($\beta = -1.04, SE = 0.25, t = -4.1, p < .001, CI = [-2.2966; -0.6091]$) and ($\beta = -0.79, SE = 0.30, t = -2.68, p < .05, CI = [-1.7147; -0.3464]$) respectively). The entire model output is summarized in table 4.20. The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure B.1, Figure B.2 (both on page 142) and Figure B.3 on page 143 respectively.

Table 4.20: Fixed Effects for PMN time-window in Experiment 2.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.65922	0.14587	-11.375	***
c1	-1.03587	0.25182	-4.113	***
c2	-0.79492	0.29649	-2.681	*
l1	-0.09142	0.17716	-0.516	
l2	0.57939	0.11706	4.950	***
c1:l1	-0.40131	0.23287	-1.723	.
c1:l2	0.03937	0.14366	0.274	
c2:l1	-0.10486	0.21418	-0.490	
c2:l2	0.27078	0.13633	1.986	.

For the comparability of the second with the first experiment, a linear Mixed-Effects Model was fitted for this time window that compares only the Congruent condition with the Averted filler type, which was comparable to the Uninformative/Averted condition in

Experiment 1. The corresponding R code of the final model was as follows:

```
final.PMN.Cont.model = lmer(eeg ~ 1 + co1 *(l1 + l2) +
                (1 + co1 + l2 | subject) +
                (0 + l1 | subject) +
                (0 + co1:l1 | subject),
              REML=FALSE, data=EXP2.PMN.Cont.Data,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
```

As in the first experiment — and similar to this experiments Mutual condition — the Averted filler type contained a significantly larger PMN modulation than the Congruent baseline ($\beta = -0.69, SE = 0.29, t = -2.4, p < .05, CI = [-1.7878; -0.9877]$). The entire model output is summarized in table 4.21. The corresponding plot of means and the topographical scalp maps can be found in Figure B.2 on page 142 and Figure B.3 on page 143 respectively.

Table 4.21: Fixed Effects for PMN time-window (control) in Experiment 2.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.38774	0.19762	-7.022	***
co1	-0.69290	0.28929	-2.395	*
l1	-0.00843	0.06432	-0.131	
l2	0.52410	0.07427	7.057	***
co1:l1	-0.16704	0.12863	-1.299	
co1:l2	0.10349	0.14853	0.697	

N400 (300 – 450ms)

Following the same approach of model reduction as for the previous time-window, the code of the resulting final model for the N400 time-window was as follows:

```
final.N400.model = lmer(eeg ~ 1 + (c1 + c2)*(l1 + l2) +
                (1 + c1 + c2 + l1 + l2 + c1:l1 + c2:l1 | subject) +
                (0 + c1:l2 | subject) +
                (0 + c2:l2 | subject),
              REML=FALSE, data=EXP2.N400.Data,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
```

In the N400 time-window between 300 and 450ms the Incongruent condition is significantly more negative than the Congruent baseline ($\beta = -0.76, SE = 0.3, t = -2.5, p < .05, CI = [-1.3797; -0.1416]$). This effect was stronger pronounced in centro-parietal regions ($\beta = -0.59, SE = 0.28, t = -2.1, p < .05, CI = [-1.1536; -0.0278]$). However, unlike the Averted condition in Experiment 1, the Mutual condition in this experiment was not significantly different from the Congruent baseline ($\beta = -0.30, SE = 0.30, t = -0.99, p > .05$). The entire model output is summarized in Table 4.22. The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure B.4, Figure B.5 (both on page 144) and Figure B.6 on page 145 respectively.

Table 4.22: Fixed Effects for N400 time-window in Experiment 2.
(. - $p < 0.1$, $*$ - $p < 0.05$, $**$ - $p < 0.01$, $***$ - $p < 0.001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.25141	0.24476	-5.113	***
c1	-0.76065	0.30579	-2.487	*
c2	-0.29515	0.29942	-0.986	
l1	0.04655	0.21491	0.217	
l2	0.76782	0.13825	5.554	***
c1:l1	-0.59072	0.27806	-2.124	*
c1:l2	0.03540	0.16660	0.212	
c2:l1	-0.08562	0.27278	-0.314	
c2:l2	0.21382	0.18150	1.178	

The corresponding R code of the final model for the control comparison in the N400 time-window was as follows:

```
final.N400.Cont.model = lmer(eeg ~ 1 + co1 *(l1 + l2) +
                (1 + co1 + l1 + co1:l1 | subject) +
                (0 + l2 | subject) +
                (0 + co1:l2 | subject),
           REML=FALSE, data=EXP2.N400.Cont.Data,
           control = lmerControl(calc.derivs=FALSE),
           na.action = na.omit)
```

Unlike the Mutual condition, the Averted filler type did elicit a significantly greater negativity compared to the Congruent baseline ($\beta = -0.71, SE = 0.34, t = -2.1, p < .05, CI = [-1.5109; 0.0863]$), similar to that found for the Averted condition in the first experiment. The entire model output is summarized in table 4.23. The corresponding plot of means and the topographical scalp maps can be found in Figure B.5 on page 144 and Figure B.6 on page 145 respectively.

Table 4.23: Fixed Effects for N400 time-window (control) in Experiment 2.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.1986	0.2442	-4.907	***
co1	-0.7123	0.3377	-2.110	*
l1	0.1144	0.1878	0.609	
l2	0.7521	0.1246	6.038	***
co1:l1	0.0575	0.2348	0.245	
co1:l2	0.2745	0.1769	1.552	

P600 (600 – 800ms)

Lastly, following the same approach of model reduction as for the previous time-window, the code of the resulting final model for the P600 time-window was as follows:

```
final.P600.model = lmer(eeg ~ 1 + (c1 + c2)*(l1 + l2) +
                        (1 + c1 + c2 + l2 + c1:l1 + c2:l1 | subject) +
                        (0 + l1 | subject),
                      REML=FALSE, data=EXP2.P600.Data,
                      control = lmerControl(calc.derivs=FALSE),
                      na.action = na.omit)
```

In the P600 time-window (600 – 800ms), there was a significantly larger positivity only for the Incongruent condition compared to the Congruent baseline ($\beta = 1.06, SE = 0.31, t = 3.48, p < .01, CI = [0.4438; 1.6817]$). The Mutual condition showed a centro-parietal positivity ($\beta = 0.59, SE = 0.29, t = 2.0, p < .05, CI = [0.0001; 1.1741]$) that was strongest in posterior electrodes ($\beta = 0.48, SE = 0.2, t = 2.6, p < .01, CI = [0.1156; 0.8488]$). Notably, the Averted counterpart in Experiment 1 instead expressed a frontal negativity (for comparison see Table 4.11 on page 58 and its description). The entire model output is summarized in table 4.24 on the following page. The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure B.7, Figure B.8 (both on page 146) and Figure B.9 on page 147 respectively.

Table 4.24: Fixed Effects for P600 time-window in Experiment 2.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $* * *$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-0.04971	0.27841	-0.179	
c1	1.06279	0.30574	3.476	**
c2	0.16809	0.35460	0.474	
l1	1.04494	0.22478	4.649	***
l2	1.16873	0.14918	7.835	***
c1:l1	0.29888	0.32294	0.926	
c1:l2	0.29170	18700	1.560	
c2:l1	0.58714	0.28997	2.025	*
c2:l2	0.48222	0.18700	2.579	**

The corresponding R code of the final model for the control comparison in the P600 time-window was as follows:

```
final.P600.Cont.model = lmer(eeg ~ 1 + co1 *(l1 + l2) +
                (1 + co1 + co1:l1 | subject) +
                (0 + l2 | subject) +
                (0 + co1:l2 | subject),
           REML=FALSE, data=EXP2.P600.Cont.Data,
           control = lmerControl(calc.derivs=FALSE),
           na.action = na.omit)
```

As its Mutual counterpart, the Averted filler type showed no significant main effect in this time-window ($\beta = 0.01, SE = 0.33, t = 0.04, p > .05$). However, it showed a similar centro-parietal positivity as found in the Mutual condition ($\beta = 0.5, SE = 0.23, t = 2.2, p < .05, CI = [0.0376; 0.9711]$) that is more positive in posterior electrodes ($\beta = 0.39, SE = 0.20, t = 2.0, p = .055, CI = [-0.0048; 0.7903]$).[2] The entire model output is summarized in Table 4.25 on the next page. The corresponding plot of means and the topographical scalp maps can be found in Figure B.8 on page 146 and Figure B.9 on page 147 respectively.

4.5.2.2 Gaze Cue preceding the Second Noun

The changed positioning of the objects with a fully counterbalanced rotation of the empty position enabled a more thorough analysis of the gaze region. For this region, a factor called Informativity was included with three levels: Informative (object-directed) gaze,

[2]While the p-value only shows a marginally significant effect ($p = 0.055$), the t-value equals 2.00 which is considered the threshold for a significant effect.

Table 4.25: Fixed Effects for P600 time-window (control) in Experiment 2.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-0.45297	0.28681	-1.579	
co1	0.01407	0.33407	0.042	
l1	1.00177	0.23556	4.253	***
l2	1.10714	0.13045	8.487	***
co1:l1	0.50436	0.23058	2.187	*
co1:l2	0.39278	0.19637	2.000	.

Table 4.26: Summary of the model results in Experiment 2. Significance indicates results according to the lmerTest package.
($-$ - $p > .1$, . - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Comparison	PMN	N400	P600
c1	$***$	$*$	$**$
c2	$*$	$-$	$-$
co1	$*$	$*$	$-$
c1:l1	$-$	$*$	$-$
c1:l2	$-$	$-$	$-$
c2:l1	$-$	$-$	$*$
c2:l2	$-$	$-$	$**$
co1:l1	$-$	$-$	$*$
co1:l2	$-$	$-$	.

Mutual gaze and Averted gaze. Again, contrasts were directly embedded into the factor (see Table 4.27). The first contrast compared Informative with Uninformative (both Mutual and Averted gaze) gaze cues, whereas the second compared the two types of Uninformative gaze cues (Mutual/Averted) to account for possible differences in their perception or processing. Also, the Longitude factor was added with the same coding as in previous analyses (see Table 4.7 on page 54).

Table 4.27: Contrast matrix as embedded in the factor Informativity in the Gaze region in Experiment 2.

Informativity	i1	i2
Object-directed	-2/3	0
Mutual	1/3	-1/2
Averted	1/3	1/2

The approach utilizing difference waves and a moving time-window analysis revealed a significant difference between Informative (object-directed) gaze and Uninformative gaze in the time-window between 300 and 500ms. The difference wave can be found in 4.12. It should be noted that there was no indication for a difference between the two Uninformative gaze cues (Mutual/Averted).

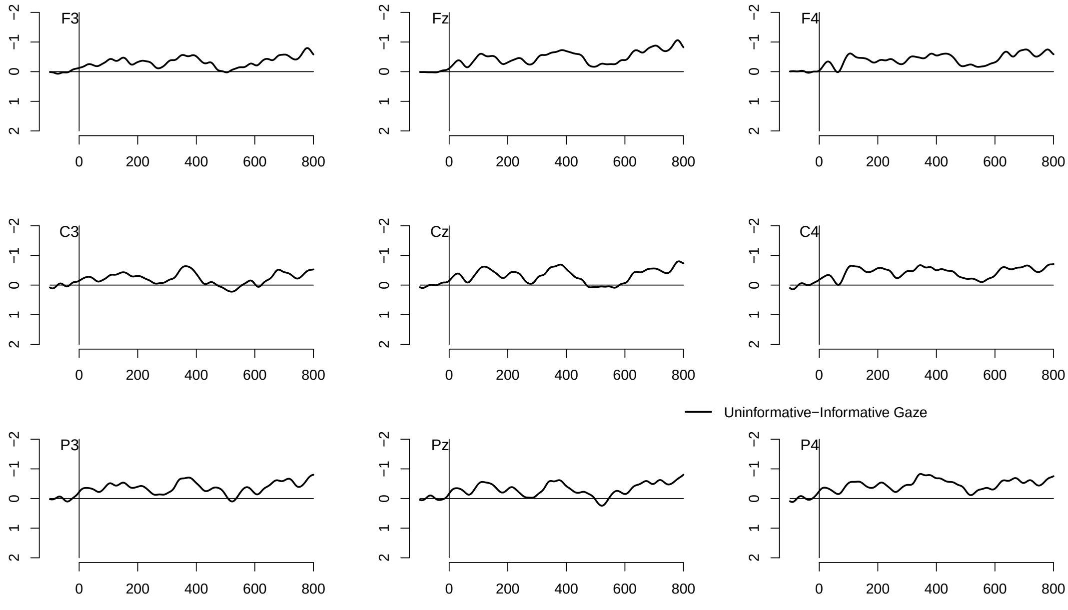


Figure 4.12: Difference wave of Uninformative-minus-Informative (black) gaze cues in Experiment 2. The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only.

A linear Mixed-Effects Model was fitted for this time-window in the same way as in previous analyses by reducing the maximal model utilizing PCAs. The R code used for the final model was as follows:

```
final.Gaze.model = lmer(eeg ~ 1 + (i1 + i2)*(l1 + l2) +
                   (1 + i1 + i2 + l1 + l2 + i1:l1 + i2:l1 | subject) +
                   (0 + i1 :l2 | subject) +
                   (0 + i2 :l2 | subject),
                 REML=FALSE, data=EXP2.Gaze.Data,
                 control = lmerControl(calc.derivs=FALSE),
                 na.action = na.omit)
```

The Uninformative gaze cues showed a significantly more negative, globally distributed deflection compared to the Informative gaze cues in this time-window ($\beta = -0.38, SE =$

$0.16, t = -2.44, p < .05, CI = [-0.6917; -0.0638]$). The entire model output is summarized in Table 4.28.

Table 4.28: Fixed Effects for 300 – 500ms time-window in the Gaze region in Experiment 2. (. - $p < 0.1$, $*$ - $p < 0.05$, $**$ - $p < 0.01$, $***$ - $p < 0.001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	1.23552	0.25558	4.834	***
i1	-0.37771	0.15508	-2.436	*
i2	0.12446	0.26390	0.472	
l1	-0.36745	0.22080	-1.664	
l2	-0.50325	0.15968	-3.152	**
i1:l1	0.01335	0.17770	0.075	
i1:l2	-0.04650	0.12163	-0.382	
i2:l1	-0.16244	0.22088	-0.735	
i2:l2	-0.21763	0.15068	-1.444	

4.5.3 Preliminary Discussion

4.5.3.1 Noun Region

The findings in the second experiment largely replicated findings from the first experiment. This holds especially for the PMN region (150 – 300ms) and the P600 region (600 – 900ms): For both time-windows the results from the second experiment replicated the findings from the first experiment. While the comparison of the Incongruent and Congruent condition also replicated the findings from Experiment 1, the results in the N400 time-window (300 – 450ms) deviated from the findings in the first experiment for the two 'uninformative' conditions (Mutual and Averted) indicating differences in the perception of these gaze cues. The difference found for these two conditions provides further support for the claim that the PMN and N400 are indexing different processes.

4.5.3.2 Gaze Region

The negativity found in the gaze region for Uninformative gaze cues compared to Informative gaze cues toward objects can possibly hint toward the process of prediction making. When the gaze cue is directed toward one specific object, only one prediction has to be made, leading to a reduced N400-like effect. However, when the gaze cue is either directed to the empty position or back to the listener, none of the objects on the screen can be anticipated based on the gaze cue. This could in turn lead to distributed expectations for either so far unnamed objects on the screen entailing a higher effort. Alternatively, the

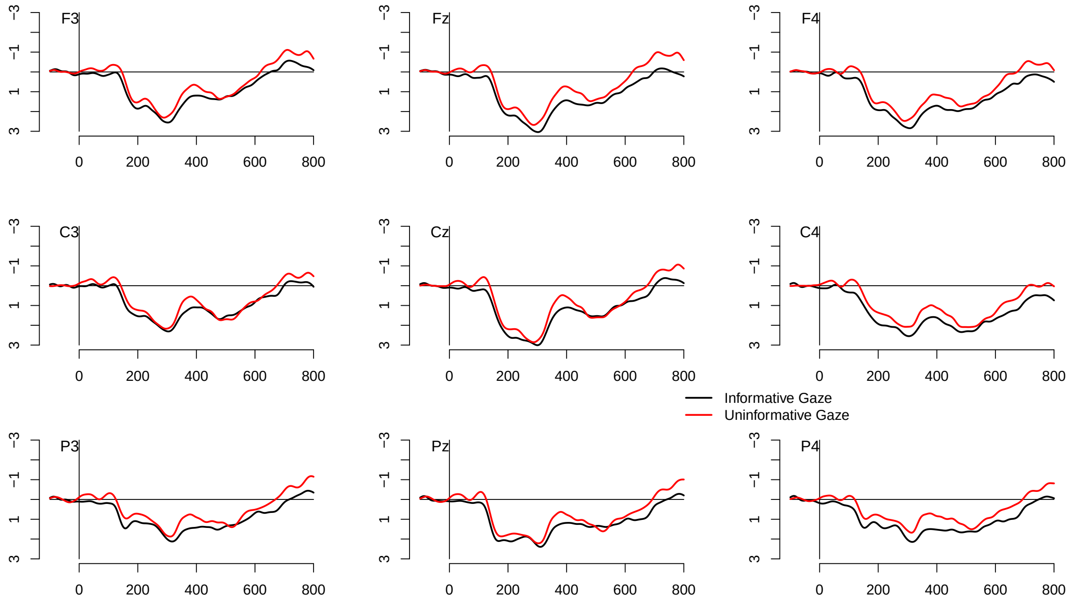


Figure 4.13: ERP time-locked to the gaze cue Onset separated by gaze toward an object (black) and Mutual (red) gaze cues in Experiment 2. The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only.

difference in the N400-like effect in this region could express a different form of prediction violation. Throughout the course of the experiment, participants are much more often exposed to gaze cues that are directed toward an object. If all gaze actions are taken into account, three out of four gaze cues are directed toward an object whereas only one fourth of the gaze cues are not directed toward an object. Therefore, it could be argued that participants have higher expectations for the gaze cue to be directed toward an object, such that the (relatively infrequent) absence of such a cue results in the observed negativity. It should be noted that the effects as reported for the first experiment when comparing upward (Informative) with downward (Uninformative) gaze cues in the time-window between 400 and 700ms were not replicated. However, as discussed earlier, the comparison in the first experiment could only be analyzed on the subset of gaze cues on the vertical axis.[3] It is therefore reasonable to assume that the analysis in the first experiment suffered from data sparsity. Additionally, it is also possible that upward and downward gaze cues are perceived and processed differently. In Experiment 2, the gaze cues are much more reliably comparable as, firstly, the angle of each position in relation to the horizontal axis through the eyes of the face was the same (30°) and, secondly, because the empty position

[3]For a complete description see Section 4.4.2.2 on page 58 Gaze Cue preceding the Second Noun.

rotated through all possible positions across the experiment.

4.6 Discussion

The results from the presented experiments suggest that the gaze cue preceding the critical second noun is used to predict a word form, which is then in turn immeditately integrated into the situation model as reflecting the speakers referential intentions. This results in clear expectations regarding how the sentence will continue. Three ERP components were identified that reveal different aspects of these processes, indexing an auditory matching mechanism (PMN), word retrieval (N400) and the integration into, as well as the revision of, a mental representation of the situation (P600). The latter provides stong evidence that gaze toward objects leads to the immediate integration of the gazed-at object into the mental representation, thus going beyond a simple increase of the objects prominence. In other words, the presented data supports the Situational Integration Account over the Prominence Account. In the following, each of the components will be discussed separately.

4.6.1 PMN

The early negative component between 150 and 300ms can be plausibly interpreted as a Phonological Matching/Mapping Negativity (PMN) as described by, e.g., Connolly and Phillips (1994). Similar results have been found by Hagoort and Brown (2000). They explain this early effect peaking at around 250ms as a mismatch between the expected word form given a context and the actual activated word candidates given the speech signal listeners perceive and additionally suggest that the 'effect might reflect the lexical selection process that occurs at the interface of lexical form and contextual meaning' (p. 1528).

Importantly, while the above mentioned studies established the predictive context based on linguistic information alone, in this study, the expectations were established by speaker gaze toward an object present in the visual scene. The linguistic context alone supports no preference for either of the valid referents.

In these experiments, the objects appearing together all had names that began with different phonemes,[4] which is an important factor to elicit a PMN (Connolly and Phillips, 1994; Hagoort and Brown, 2000; D'Arcy et al., 2004). If the initial phoneme of the input matches the onset of the predicted word (i.e. named the gazed-at object), this phoneme provides little new information and is therefore easily processed (Congruent). If however the phoneme provides more information, either contradicting the prediction of a specific

[4]With the exception of the pair 'Stiefel' – 'Stern' in Experiment 1, which was present in three experimental trials. In two of those three cases one of the two objects was the medium sized object, making it the first gazed-at object in those sentences independent of the condition. Thereby, they are no longer a valid target for the second gaze cue or naming as the second noun in the sentence in any condition. 'Stern' was replaced with 'Mond' in Experiment 2.

noun (Incongruent), or by helping to reduce the set of possible nouns to a single target (Averted/Mutual), a PMN modulation is elicited. The Incongruent gaze cue leads to expectations about the upcoming referent, including its word form and, thus, the initial phoneme of that word. When perceiving the initial phoneme of the actual referent, this leads to a mismatch. Hence, the initial phoneme of the actual referent provided new information that was required to retrieve the corresponding word. The Averted and Mutual gaze cues do not provide any information about the upcoming word. Therefore, lacking an early visual point of disambiguation (VPoD), the earliest possibility to identify and select the actual target is provided by the first phoneme of the actual noun as the linguistic point of disambiguation (LPoD). This in turn increases the information load conveyed by this phoneme. In sum, the PMN is interpreted to reflect the processing of information provided by the phoneme, given gaze-driven word-form expectations.

4.6.2 N400

The N400 effect has been reported to reflect semantic retrieval effort based on expectations arising from contextual information (Van Berkum, 2009; Kutas and Federmeier, 2011; Schumacher, 2012). In the presented experiments, the gaze cue preceding the second noun leads to expectations for the upcoming noun. In the Congruent condition, those (matched) expectations lead to facilitated retrieval of the referent or could even be interpreted to indicate a — based on the gaze cue — already retrieved meaning, as revealed by an attenuated N400 amplitude. In the Averted and Mutual conditions, participants have two possible upcoming nouns activated. The noun alone is used to identify the referent, resulting in a significantly larger N400 effect compared to the Congruent condition. In the Incongruent condition, the noun is not consistent with the expectations formed using the gaze cue. This increases the retrieval cost of the noun, as manifest by the significantly larger N400 effect, compared to the Congruent condition. In both experiments, the Incongruent condition displays a single negative movement in central and parietal regions compared to the two peaks observed in the Averted condition. However, in frontal electrodes, the two peaks are visually distinguishable in both conditions. This was also shown by the analysis of the time-window between the PMN and N400 effect, where only the Incongruent condition showed a significant difference from the Congruent condition. This could be interpreted as a stronger N400 effect in the Incongruent condition that interacts/overlaps with the PMN.

The findings in the N400 time-window in the second experiment replicate the results from the first experiment for the three conditions that were also present in the first experiment (Congruent, Incongruent and Averted), as summarized in Table 4.26 on page 77. However, the added Mutual condition — utilizing a straight gaze back to the participant instead of being directed toward the empty position (Averted) — shows no significant difference from the Congruent baseline condition unlike its Averted counterpart.

The significant difference in the PMN time-window followed by a lack of difference in the N400 time-window only for the Mutual condition provides further evidence that the two peaks (PMN and N400) are indeed separable, expressing two distinct processes. In both the Mutual and Averted conditions, the auditory input can be utilized to identify the target out of the remaining objects early on as evident from the PMN effect found for both conditions. However, the reduced N400 effect following the listener-directed Mutual gaze compared to the Averted gaze suggests that the two different gaze cues might introduce qualitatively different expectations. Although the precise nature of the difference between the two Uninformative cues in relation to the Congruent baseline remains to be investigated, two possible explanations are proposed: Firstly, it could be argued that every position-oriented gaze action is interpreted as being meaningful throughout the experiment based on the higher number of object-oriented gaze actions (75%). As a consequence, even gaze to an empty position could bind the listeners' attention — pulling their attention away from the objects provided in the scene — hindering word retrieval for any object outside of the attentional focus.

Alternatively, it may be the case that Mutual gaze expresses a higher amount of certainty about the upcoming word compared to an Averted gaze cue toward an empty position, which might rather imply some degree of uncertainty and might even open the space of suitable candidates beyond those objects present on the screen. The speaker's gaze away from the interlocutor or away from discourse relevant objects is often described as a disengagement from the environment and used to facilitate remembering or, more generally, to lower the cognitive load of the speaker (Doherty-Sneddon et al., 2002; Glenberg et al., 1998). Such a gaze behavior is often understood to display uncertainty or disfluency (Griffin, 2004). Work from Swerts and Krahmer (2005) has shown that interlocutors pick up on such cues that display uncertainty and interpret them in relation to the so called Feeling of Anothers Knowing (FOAK). In their study they showed that participants presented with videos of speakers displaying such cues of uncertainty rated those sentences with a lower FOAK score than videos in which those cues were not displayed. If the averted gaze cue in this experiment is interpreted along those lines, it is possible that this Averted condition leads to different expectations or predictions than the Mutual condition utilizing a straight gaze toward the participant. It is possible that word retrieval for a small set of expectable objects (Averted and Mutual conditions) is benefiting from a higher FOAK (Mutual).

4.6.3 P600

The update of the mental situation model following the violation of the comprehenders expectations can elicit a P600 effect (Van Berkum et al., 2007; Burkhardt, 2006, 2007). Following those accounts, the findings in the P600 region are interpreted as revision/integration costs of the situation model. In both the Congruent and Incongruent gaze condition participants can exploit the gaze cue toward an object to integrate the identified referent into their situation model, and establish expectations for the upcoming noun. In both the Congruent and Averted/Mutual condition, there is no violation of expectations: Either the expected referent was named (Congruent), or no expectations have been formed (Averted/Mutual). In the Incongruent condition however, the violation of the expectations leads to the necessity to revise the situational model by replacing the expected referent in the model with the actually named referent that had been excluded based on the visual information. This could not be explained under the Prominence Account, as the gazed-at object would not be integrated into the mental representation of the situation and, hence, no revision of that representation would be required when hearing the referent. Thus, the effect found in the P600 time-window for the Incongruent condition can only reasonably be explained under the Situational Integration Account.

4.6.4 Summary

Experiments 1 and 2 provide strong support for the utilization of speaker's gaze by interlocutors in forming expectations of the unfolding sentence. The findings suggest that gaze is used to form expectations about the upcoming referent, resulting in increased retrieval costs when gaze is uninformative or misleading, as indicated by a stronger N400 modulation in those cases. The attenuated PMN for Congruent gaze preceding the N400 time-window further suggests that predictions are not only formed on a conceptual level but also about the concrete lexical form when a single object is gazed at. The additional findings in the second experiment regarding the Mutual gaze cue as displayed by a straight gaze to the participant provides further evidence to distinguish between the two processes. It is important to note, however, that the reported results also suggest a strong interplay between these two components. The relatively short lived N400 (300 – 450ms) following the PMN suggests that the retrieval of the full word benefits from the phonological matching as indexed by the PMN. It is speculated that the presence or absence of the PMN might have an effect on the strength of the N400. This, however, requires further investigation.

In the P600 time-window, the results suggest that the visual scene, utterance, and gaze are used to form a mental representation of the discourse. This is consistent with the view that gaze is interpreted as conveying referential intentions (Staudte and Crocker, 2011). The first gaze action in each experimental trial correctly provided evidence about the upcoming referent. In case of a following Incongruent gaze, participants were led to believe that the gazed-at object actually would be the upcoming noun and integrate this into the mental representation of the situation, eliminating the remaining objects in the scene as likely referents. The upcoming noun however forces the participant to reintegrate the formerly dismissed object into the mental representation. This in turn is then reflected by a P600 modulation representing the (re-)integration difficulty. No such difference is induced in the Uninformative conditions (Mutual/Averted) as either upcoming referent is still possible and, hence, does not require a revision of the situation model. While the N400 for both the Incongruent and Averted conditions was consistent with both the Prominence and Situational Integration Accounts, the observation of a P600 only in the Incongruent condition was predicted by the Situational Integration Account alone.

Results from the second experiment largely replicated the results from the first experiment. When comparing only the three conditions that were present in both experiments, the second experiment shows similar patterns in the PMN, N400 and P600 time-windows. This demonstrates the robustness of the observed effects to variation in object position and gaze cue validity. Further, two different types of Uninformative gaze cues were used in the experiments, either being directed toward an empty position on the screen (Averted) or back toward the participant (Mutual). The results showed a significantly increased N400 modulation in the noun region for the Averted gaze condition compared to Congruent condition that is absent in the Mutual condition, while both Uninformative gaze versions replicate the effects in every other time-window.

Taken together, the reported findings are consistent with the Rretrieval-Integration Account (Brouwer et al., 2017), such that retrieval difficulty (N400) is observed for the Incongruent and Averted conditions, while integration difficulty (P600) is found only when revision of the situation model is necessary in the Incongruent condition. Extending the Retrieval-Integration Account (RIA), however, is the PMN modulation that is not included in the account. While the RIA only considered the N400 and P600 components to express retrieval and integration mechanisms respectively, the provided results provide evidence that, in situated spoken interactions, a phonological matching mechanism, as indexed by the PMN, is part of the retrieval process by identifying the word for which the semantics need to be retrieved.

4.6.5 Conclusion

Experiments 1 and 2 reveal a robust and replicable influence of speech-related gaze cues on a range of underlying cognitive processes, including auditory word processing, lexical retrieval, and integration with sentence meaning, as expressed by an PMN, N400 and P600 effect respectively. The distinct PMN and N400 components suggest that gaze elicits predictions on word form level which are matched with the incoming phonological information whereas the N400 indicates a broader expectation-driven retrieval mechanism. The P600 results indicate that listeners utilize speakers' gaze above and beyond any increase in prominence, such that the information provided by gaze is used to update the situation model even in advance of hearing the gazed-at referent. The presented findings therefore provided support for the Situational Integration Account rather than for the Prominence Account.

While the up to this point presented experiments provide evidence for the utilization of speaker gaze as a predictive cue of upcoming linguistic content, it can be argued that gaze could possibly be used even further to also confirm predictions about an unfolding utterance based on the linguistic context provided. This will further be investigated in the following chapter.

CHAPTER 5

Linguistic Content as a Predictive Cue for Speaker Gaze

The previous experiments showed the effects of speaker gaze on sentence comprehension when it can be utilized to anticipate the continuation of an unfolding utterance. The results provided strong support for a mental representation of the situation that is formed and adjusted incrementally utilizing the unfolding utterance as well as the gaze cues as soon as they are available. The two experiments presented in Chapter 4 on page 37 provided evidence for the effect of speaker gaze preceding the mentioning of a referent on mechanisms of language comprehension when hearing that referent. Three such mechanisms where identified in those ERP experiments: a) a phonological matching mechanism that utilizes the perceived initial phoneme of the referent to match it with the anticipated word form based on the preceding gaze. This was indexed by a PMN effect with a greater negative deflection in response to phonemes that conveyed new information (Uninformative and Incongruent conditions). b) a semantic retrieval mechanism that was inhibited when gaze was either Uninformative or Incongruent, requiring the retrieval of the perceived referent. This was indexed by an N400 effect. c) a mechanism of integration with the mental representation of the situation that was inhibited when the perceived referent was not the previously gazed-at object (Incongruent condition). As the gazed-at object was immediately integrated in the mental representation, hearing a referent different than the gazed-at object required a revision of that representation. This revision of the mental representation was expressed by a P600 effect. Taken together, these results provided strong support for the proposed Situational Integration Account.

While these results showed that speaker gaze is utilized to form expectations about the upcoming referents in an utterance, the question arises whether gaze could reversly also be utilized to confirm expectations formed on the preceding linguistic content. To address this question, the experiment presented in this chapter investigates the processing of speaker gaze as a confirming cue for expectations formed on linguistic context. To recall, the presentation of the sentences in the previously presented experiments did not provide any information about the upcoming referent prior to the gaze cue, such that the gaze cue was the first piece of information that could be utilized to anticipate the upcoming referent in a sentence. The results provided strong support for the view that gaze is indeed treated as a part of the communicative signal. Hence, it should be possible for gaze cues to also elicit a neurophysiological response when they are unexpected given the prior linguistic context.

In this chapter the roles of the linguistic context and gaze are reversed, so that the linguistic information is providing grounds for expectations about the upcoming referent that follows a subsequent gaze cue. The gaze cue can thereby possibly be utilized to either confirm or disconfirm those linguistic expectations before the naming of the referent. In order to investigate whether gaze is utilized to confirm expectations about upcoming linguistic content of a sentence, an ERP study utilizing the same visual scenes as presented in experiment 2 (see Figure 4.10 on page 67) but with an alternation of the provided utterance was conducted.

5.1 Experimental Design

The third experiment, while utilizing the same visual setup, presented a sentence which, based on the linguistic content, could be used to form expectations about the continuation of the sentence before a gaze cue toward the object was presented. The gaze cue thus could be used to confirm those expectations, but was not required to understand the sentence or to anticipate the continuation of the sentence. The stimuli were created using the same objects and screen types as in experiment 2 with a diagonal placement of the objects around the face. Again, the CereVoice TTS system's Alex voice (Version 3.2.0) was used to synthesize the sentences that were presented to the participants.

In order to create a scenario in which the sentences support the forming of expectations based on the linguistic content, the sentences were changed from their **Gaze-Referent-Comparative** structure as utilized in the previous experiments to be of the form **Comparative-Gaze-Referent** as in *"Das Auto ist kleiner**<Comparative>** als das **<Gaze>** abgebildete Haus**<Referent>**, denke ich"* (*'The car is smaller**<Comparative>** than the **<Gaze>** displayed house**<Referent>**, I think'*). This sentence replaced the sentence *"Verglichen mit dem Auto, **<Gaze>** ist das Haus**<Referent>** verhältnismäßig groß**<Comparative>**, denke ich"* (*'Compared to the car, **<Gaze>** the house**<Referent>** is relatively big**<Comparative>**, I think'*) in previous experiments while retaining the same semantic meaning in relation to the same visual scene as depicted in Figure 5.1 on page 90. In contrast to the previous experiments, the linguistic content of the sentence supports linguistic expectations for an upcoming referent early on in the sentence, namely, on the mentioning of the comparative. Provided with the related visual scene, the comparative (*'kleiner/smaller'* in the example above) supports expectations for *'house'* to be the upcomming referent, as it is the object that is bigger than the car. This expectation can in principal be formed both before the gaze toward the object and its subsequent naming. With this change, the role of the comparative also shifts: In the first two experiments, the comparative was only useful to determine whether the sentence was true or false, whereas in this experiment the comparative serves as predictive linguistic information to anticipate the continuation of the sentence.

The gaze cue preceding the mentioning of the second object was manipulated such that it was either present or absent. If, however, the gaze cue was present, it was always congruently directed toward the actual referred to object. Therefore, in this experiment, any occurring gaze cue was always congruent and never incongruent in respect to the subsequently named referent. However, 50% of the experimental items contained false statements and, thus, gaze toward objects that — although being subsequently mentioned — were not supported by the linguistic content of the sentence. In the previously mentioned example related to the scene depicted in Figure 5.1 on the following page, this would mean that such a false statement was *"Das Auto ist größer als das abgebildete Haus, denke ich"* (*'The car is bigger than the displayed house, I think'*), which, in relation to to the provided scene, is an incorrect statement. Still, the gaze in this condition was directed toward the house and, therefore, was congruent in respect to the following referent. Overall, this manipulation led to a 2x2 design with the factors Gaze Presence (Present/Absent) and Linguistic Expectability of the referent in respect to the provided scene (Expected/Unexpected). It is to be noted that, in this experimental setup, linguistic expectability and the truth value of the sentence are directly related. Expectability is defined as the mentioning of the referent that leads to a true statement.

In total, four lists were created following a Latin square design. Each list contained 128 experimental items (32 per condition) and 124 fillers. Fillers were used to raise the number of true statements above chance and to also manipulate the presence/absence of the gaze cue preceding the mentioning of the first object. If the gaze cue toward the first referent was absent, the gaze cue preceding the mentioning of the second object was always present in order to retain at least one gaze action per item.

The stimuli were presented using the E-prime software (Version 2.0.10. Psychology Software Tools, Inc.). Each participant was seated in a sound-proof, electro-magnetically shielded chamber in front of a 24" Dell U2410 LCD monitor (resolution of 1280x1024 with a refresh rate of 75 Hz). The distance between the participant and the screen was always 114 cm in order to keep all objects in a 5° visual angle from the center of the screen. This is the same setup as used in experiment 2 due to the reused distribution of objects. While the participants were prepared for the recording, they were presented with all objects that occurred throughout the experiment and their naming. The Alex voice of the CereVoice TTS

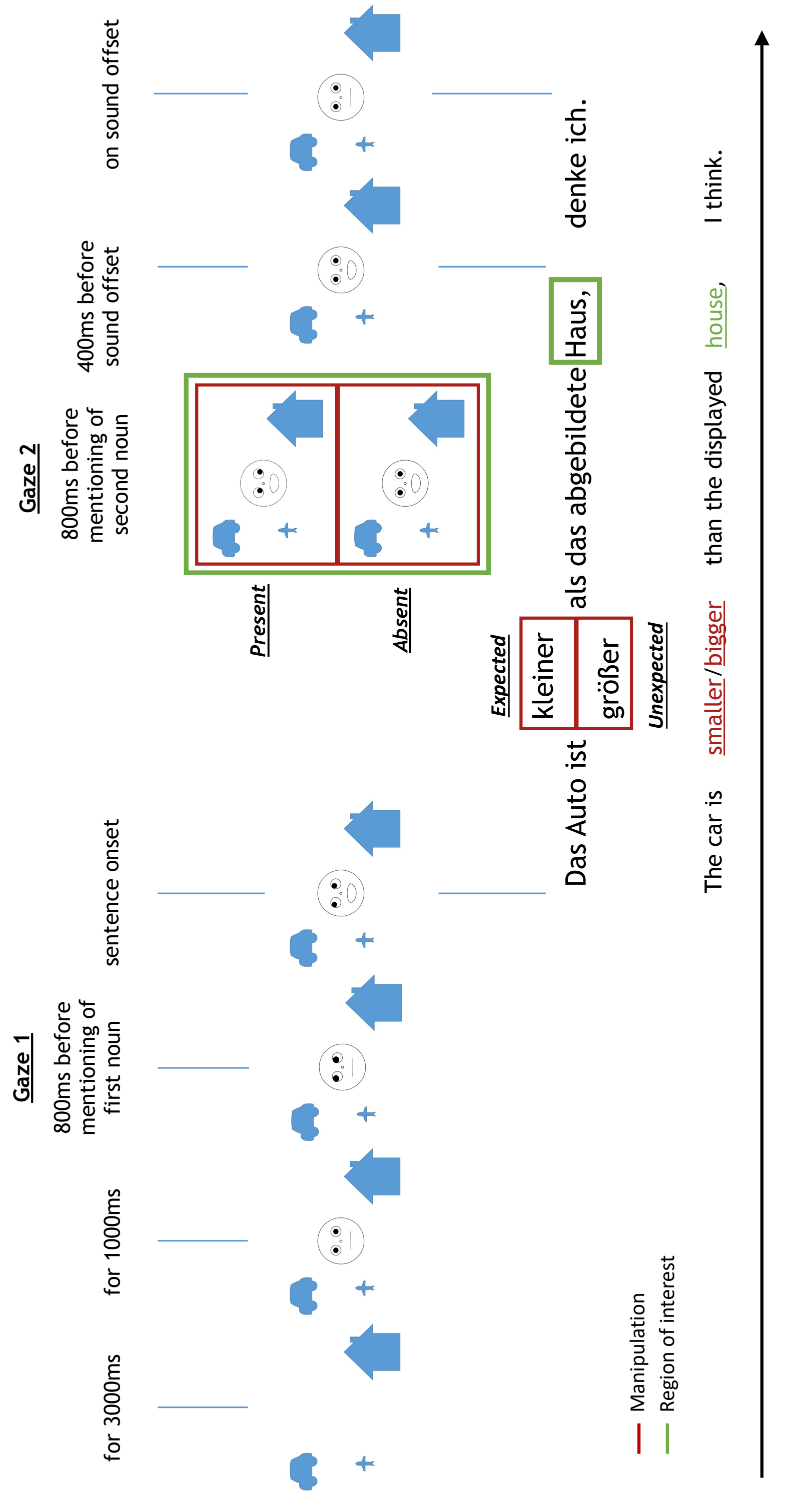


Figure 5.1: Timeline of an experimental trial in experiment 3.

was also used for the naming of the objects. After this, participants were presented with written instructions and completed six practice trials. The items were pseudo randomized for each list and presented in 5 blocks with breaks after each block. After each item, the participants were asked to indicate whether the sentence was true given the visual scene they were presented with by pressing one of two buttons. Answers were recorded using a Response Pad RB-834 (Cedrus Corporation). The experiment lasted approximately 60 minutes.

5.2 Hypotheses

When gaze could be used to evaluate expectations formed on the preceding linguistic context of a sentence, there are two regions of interest: The gaze cue itself, and the subsequent noun. Under the Prominence Account, again no effects are predicted on the gaze cue. Under the Situational Integration Account, however, the involvement of both the retrieval mechanism as well as the integration mechanism are predicted. If gaze was utilized to evaluate the expectations formed on the basis of the preceding linguistic context, a mismatch with those predictions (Present Unexpected gaze) should entail the retrieval of the gazed at objects semantics (N400) and the necessity to update the mental representation of the situation (P600). As there is no auditory input during the gaze cue, no involvement of a phonological matching mechanism is predicted.

In the noun region, under the Prominence Account, all three proposed mechanisms are predicted. As expectations about the unfolding sentence could be formed on the basis of the preceding linguistic content of the sentence, these expectations should be eliciting the involvement of the phonological matching mechanism as well as the retrieval and integration mechanisms. It could be argued that comprehension still could benefit from a gaze to the not-anticipated referent if this referent is consequently named: If the gaze to an object draws the listeners attention to that object, word retrieval might benefit from the allocated focus. Under the Situational Integration Account, effects in the noun region following the gaze cue are predicted to be dependent on the presence or absence of the gaze cue. If a gaze cue to the subsequently mentioned referent was present, no effects are predicted in the noun region as the evaluation of the expectations was done on the preceding gaze. If however gaze was absent, the noun region provides the first grounds to evaluate the expectations and, hence, should express effects relative to the inhibition of the involved mechanisms. In other words, if the mentioned referent is not the obeject expected (*'house'* after hearing *'größer'*) based on the preceding linguistic context (Absent Unexpected gaze condition), the phonological matching mechanism (PMN) as well as the retrieval (N400) and integration (P600) mechanisms should be impaired. A summary of the described predictions for both accounts can be found in Table 5.1 on the next page.

Table 5.1: Summary of the predicted effects for the two proposed accounts for the Unexpected conditions relative to the Expected conditions. (PA - Prominence Account ; SIA - Situational Integration Account)

Region	Component	PA	SIA
Gaze	PMN	-	-
	N400	-	+
	P600	-	+
Noun (Gaze absent)	PMN	+	+
	N400	+	+
	P600	+	+
Noun (Gaze present)	PMN	+	-
	N400	(+)	-
	P600	+	-

5.3 Experiment 3

This experiment aimed to investigate whether gaze is not only used to incrementally form expectations — as evident from the previous experiments — but whether gaze is also utilized to evaluate expectations that where formed based on preceding linguistic information and, if so, how this evaluation impacts subsequent word processing of the actual referent.

As the previous experiments provided evidence for the Situational Integration Account over the Prominence Account, the hypotheses for this experiment are according to that account, so that gaze was hypothesized a) to be used to verify expectations formed on the linguistic context as soon as it is provided and b) to ease word processing of the subsequent referent. More precisely, it was hypothesized that, if gaze is provided, the gazed-at object is used to confirm the expectations that were formed based on the preceding linguistic context and that a violation of those expectations leads a) to predictability induced N400 effect Van Petten and Kutas (1990, 1991) and b) to revision costs of the mental representation of the situation indexed by a P600 effect. Consequently, in the absence of gaze, the same effects that were present in the first two experiments were hypothesized to appear on the referent. Explicitly, a phonological matching mechanisms indexed by a PMN, word-retrieval mechanisms indexed by an N400, and integration with a mental representation of the situation indexed by a P600. If gaze, as shown in Chapter 4.4.3.2 on page 65, is utilized to incrementally inform the mental representation of the situation and to influence expectations, expectations formed on the grounds of linguistic context should lead to similar effects on the subsequent words.

5.3.1 Participants

Thirty-four right-handed native speakers of German (Mean age: 24; Age range: [19, 31]; SD: 3.11; Female: 26) took part in the ERP experiment. 10 participants were removed from the analysis due to technical errors (3) and too high numbers of eye artifacts (7). For a concrete description of the removal see the following section. Participants gave informed consent. All participants had normal or corrected-to-normal vision and had no hearing problems. All participants were compensated with €20 for their participation.

5.3.2 Data Analysis and Results

The technical setup was the same as in the previous experiments, so that the EEG was recorded by 24 Ag/AgCl[1] scalp electrodes (actiCAP, BrainProducts) and amplified with a BrainAmp (BrainVision) amplifier. Electrodes were placed according to the 10-20 system (Sharbrough et al., 1991). Impedances were kept below 5kΩ. The ground electrode was placed at AFz. The signal was referenced online to the reference electrode FCz and digitized at a sampling rate of 500 Hz. The EEG files were re-referenced offline to the average of the mastoid electrodes. The horizontal electrooculogram (EOG) was monitored with two electrodes placed at the right and left outer canthi of each eye and the vertical EOG with two electrodes below both eyes paired with Fp1 and Fp2. During recording an anti-aliasing low-pass filter of 250Hz was used. The EEG data was band pass filtered offline at 0.01-40Hz in order to attenuate skin potentials and other low voltage changes as well as line noise and EMG noise (Luck, 2014). Single-participant averages were computed for a 900ms window per condition relative to the acoustical onset of the noun following the manipulated gaze cue and the manipulated gaze cue itself. All segments were aligned to a 100ms pre-stimulus baseline. The data was semi-automatically screened offline for electrode drifts, amplifier blocking, eye-movements and muscle artifacts.

Different from the previous experiments, the recorded data was processed in Matlab (MATLAB, 2017), utilizing the EEGLAB plugin (Delorme and Makeig, 2004) in combination with the ERPLAB extension (Lopez-Calderon and Luck, 2014). This change was mainly done in order to extract additional item information instead of a mere per participant average.

Three participants had to be removed due to technical errors. Further 7 participants were removed due to the 30% threshold for the rejection rate per condition for participant exclusion due to eye-movements and other artifacts. We chose a 30% threshold because of the nature of the presented stimuli: As we presented Overall, the remaining participants

[1]This excludes the electrodes used for the electrooculogram and offline re-reference: Fp1, Fp2, T7, T8, TP9, TP10, PO9 and PO10.

performed 5.2% eye movements on average. There was no significant difference for eye movements across the experimental conditions when taking both regions of interest (gaze and noun region) together ($(F(3,69) = .25, p = .84)$). Again, participants performed very well in the task with an average of 97.5% of correct answers. As in the previous experiments, there was no difference in accuracy between conditions ($F(3,69) = 1.93, p = .15$). This criterion did not lead to the removal of any participant. Overall, the three criteria led to the removal of the data of 10 participants. After artifact rejection and participant exclusion 94.8% of the trials on average per participants were included in the analyses.

The averaged data of the remaining 24 participants (Mean age: 24.3; Age range: [19, 33]; SD: 3.2; Female: 24) was exported from MATLAB in a format processable by R. Two regions of interest were analyzed: The onset of the gaze cue toward the second noun, and the onset of said noun. Analyses were performed in R by fitting Linear Mixed-Effects Models using the lme4 (Bates et al., 2015b) package (Version 1.1-10). β-Estimate, standard error and t-value are reported as well as the p value for significant effects and confidence intervals. The p values were extracted utilizing the lmerTest package (Version 3.0-1). The confidence intervals were extracted utilizing the profile function of the stats package (Version 3.6.1). Additional to effects, model fit was used to determine the reported models by model comparison utilizing the anova function implemented in the stats package (Version 3.5.1).

As in the previous analyses, in a primary step, models with maximal random structure were fitted following (Barr et al., 2013). Based on the results from the previous experiments and the manipulation in place as well as evident from visual inspection of the data, a main effect of Gaze Presence was to be anticipated. As the noun region is immediately following the gaze region, the consequential potential differences this could entail for the baseline of the noun region led to a different approach of coding the contrasts for this experiment. All four conditions, namely Present gaze - Expected continuation (PE), Present gaze - Unexpected continuation (PU), Absent gaze - Expected continuation (AE) and Absent gaze - Unexpected continuation (AU), were encoded in a single factor. The embedded contrast structure was coded so that Expectability was nested under Gaze Presence. Therefore, the factor included comparisons firstly for the presence of gaze followed by a comparison of the expectability of the continuation under each of the two Gaze Presence factor levels. The embedded contrast matrix can be seen in Table 5.2 on the next page.

Additionally, Longitude was again added as a fixed effect in order to attest for scalp distribution of potential effects. For this factor, as in the previous analyses, the electrodes were grouped into 3 ROIs for frontal (F3, Fz, F4, FC5, FC1, FC2, FC6), central (C3, Cz, C4, CP5, CP1, CP2, CP6) and posterior (P7, P3, Pz, P4, P8, O1, O2) distributions. Again, contrasts were directly embedded into the factor with two comparisons. Firstly, frontal against centro-parietal electrodes and, secondly, central against parietal electrodes (see Table 4.7 on page 54 for the contrast matrix as embedded in the factor Longitude). As

Table 5.2: Contrast matrix as embedded in the factor Condition in experiment 3.
P - Present gaze , A - Absent gaze , E - Expected continuation , U - Unexpected continuation

Condition	c1	c2	c3
PE	-1/2	-1/2	0
PU	-1/2	1/2	0
AE	1/2	0	-1/2
AU	1/2	0	1/2

in the previous experiments, the full random structure including the interaction between Condition and Longitude under subject and, additionally, item led to singular fit warnings in every time-window.

The same model reduction method for the random structure was used as described for the previous experiments, namely the parsimonious mixed models approach as described by (Bates et al., 2015a) utilizing PCAs. The resulting random structure used in the final models is reported for each model.

5.3.2.1 Gaze Cue preceding the Second Noun

Based on the difference waves (see Figure 5.2 on the next page) and the moving time-window approaches, two time-windows in the gaze region were identified showing effects to the manipulations. The first time-window spanning from 250-800ms for the comparison of Present with Absent gaze cues and the second time-window lasting from 250-450ms for the comparison of Expected with Unexpected gaze cues when gaze was Present. In the following, the first time-window will be refered to as the P300 time-window (250-800ms) and the second time-window as the N400 time-window (250-450ms).

P300 (250 – 800ms)

In the P300 time-window, there was a main effect of Gaze Presence ($\beta = -1.7, SE = 0.33, t = -5.1, p < .001, CI = [-2.3684; -0.9539]$) with a more positive deflection for Present gaze as compared to absent gaze as can be seen in Figure 5.3 on page 97. (For a summary of the model results, see Table 5.3 on page 97.) The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure C.4, Figure C.5 (both on page 152) and Figure C.6 on page 153 respectively. The R code of the final model was as follows:

```
final.Gaze.P3.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
                (1 + c1 + c2 + c3 + l1 + l2 | subject) +
                (0 + c1:l1 | subject) +
                (0 + c1:l2 | subject) +
                (0 + c2:l1 | subject) +
                (0 + c3:l1 | subject) +
                (1 + c1 + c2 + c3 + l1 | item) +
                (0 + l2 | item) +
                (0 + c1:l1 | item) +
                (0 + c2:l1 | item) +
                (0 + c3:l1 | item),
              REML=FALSE, data=EXP3.Gaze.P3.Data,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
```

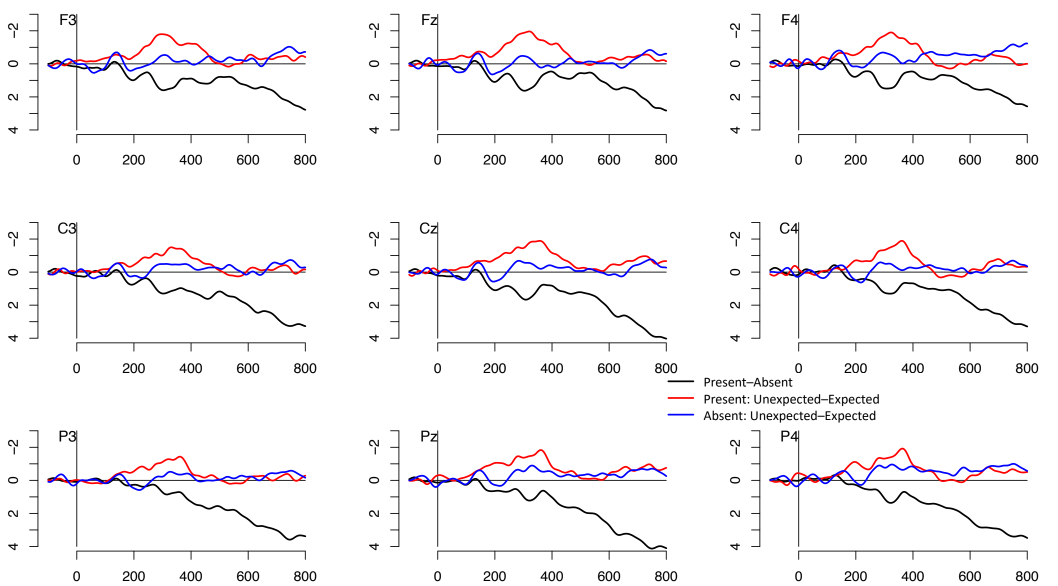


Figure 5.2: Difference waves of Present-minus-Absent gaze cues (black), Unexpected-minus-Expected under Present gaze cues (red) and Unexpected-minus-Expected under Absent gaze cues (blue) in experiment 3. The data presented shows the averaged responses for the three topographic regions of interest Frontal (F7, F3, Fz, F4, F8, FC5, and FC6), Central (C3, Cz, C4, CP5, CP1, CP2, and CP6), and Parietal (P7, P3, Pz, P4, P8, O1, and O2) filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

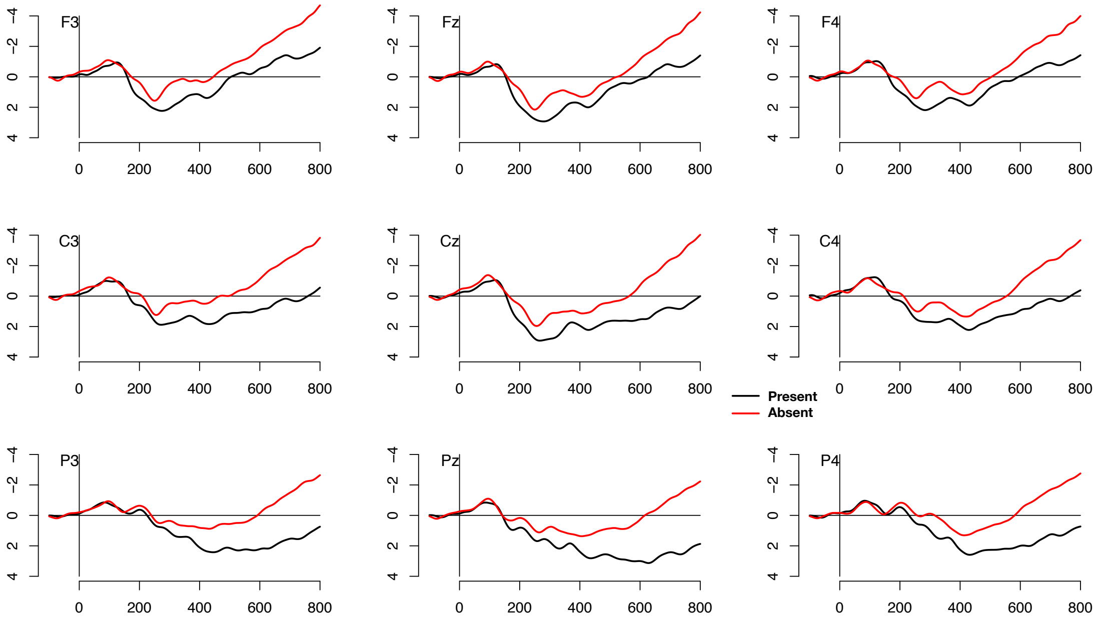


Figure 5.3: ERP time-locked to the Gaze Cue Onset in Experiment 3 separated by Gaze Presence (Present: black, Absent: red). The data presented shows all electrodes used in the analysis (F3, Fz, F4, FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1 and O2) filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

Table 5.3: Fixed Effects for P300 time-window in the Gaze region in experiemnt 3. (. - $p < 0.1$, $*$ - $p < 0.05$, $**$ - $p < 0.01$, $* * *$ - $p < 0.001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	0.22017	0.32183	0.684	
c1	-1.70320	0.33296	-5.115	***
c2	-0.56046	0.40668	-1.378	
c3	-0.35322	0.36368	-0.971	
l1	0.54351	0.18282	2.973	**
l2	-0.13531	0.15963	-0.848	
c1:l1	-0.33218	0.24392	-1.362	
c1:l2	0.09708	0.16250	0.597	
c2:l1	0.18504	0.36154	0.512	
c2:l2	0.29165	0.19525	1.494	
c3:l1	-0.03037	0.29115	-0.104	
c3:l2	-0.05974	0.19475	-0.307	

N400 (250 – 450ms)

The N400 time-window again showed a main effect of Gaze Presence with a stronger positive deflection for Present gaze ($\beta = -1.05, SE = 0.31, t = -3.4, p < .01, CI = [-1.6385; -0.4691]$). Additional to this main effect, however, an effect of Expectability was found when gaze was Present ($\beta = -1.06, SE = 0.43, t = -2.5, p < .05, CI = [-2.0653; -0.3115]$) with a more negative deflection for Unexpected gaze cues as can be seen in Figure 5.4 on the facing page. This effect was strongest in fronto-central electrodes ($\beta = 0.45, SE = 0.19, t = 2.5, p < .05, CI = [-0.0232; 0.8911]$). (For a summary of the model results, see Table 5.4 on the next page.) The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure C.1, Figure C.2 (both on page 150) and Figure C.3 on page 151 respectively. The R code of the final model for this time-window was as follows:

```
final.Gaze.N4.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
                          (1 + c1 + c2 + c3 + l1 + l2 | subject) +
                          (0 + c1:l1 | subject) +
                          (0 + c1:l2 | subject) +
                          (0 + c2:l1 | subject) +
                          (1 + c1 + c2 + c3 + l1 | item) +
                          (0 + l2 | item) +
                          (0 + c1:l1 | item) +
                          (0 + c2:l1 | item) +
                          (0 + c3:l1 | item),
                        REML=FALSE, data=EXP3.Gaze.N4.Data,
                        control = lmerControl(calc.derivs=FALSE),
                        na.action = na.omit)
```

5.3.2.2 Noun region

For the analyses of the noun following the gaze cue, the same time-windows were used that were established in the previous experiments. More precisely, models were fitted for the PMN time-window (150-300ms), the N400 time-window (300-450ms), and the P600 time-window (600-800ms). Figure 5.5 on page 100 shows the ERPs in the noun region when gaze was Present, whereas Figure 5.6 on page 101 shows the ERPs when gaze was Absent. The split was made to ensure that the effects are not directly compared, which, based on the differences in the baseline, is not informative. (For comparison, see Figure 5.7 on page 102, which displays the ERPs spanning over both ROIs.)

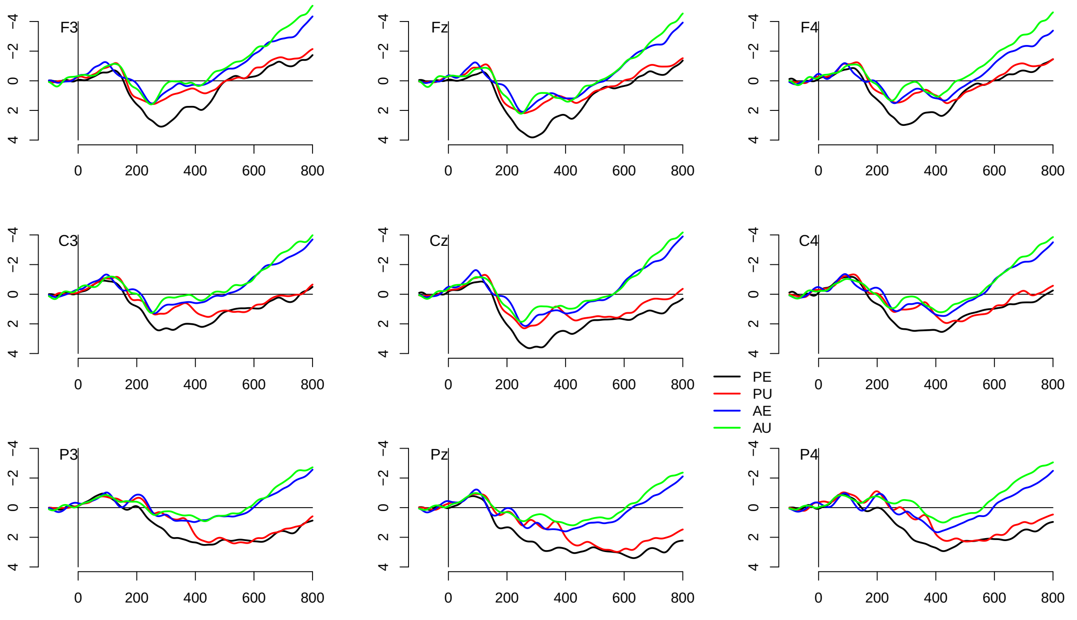


Figure 5.4: ERP time-locked to the Gaze Cue Onset in Experiment 3 separated by the 4 experimental conditions (Present-Expected: black, Present-Unexpected: red, Absent-Expected: blue and Absent-Unexpected: green). The data presented shows all electrodes used in the analysis (F3, Fz, F4, FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1 and O2) filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

Table 5.4: Fixed Effects for N400 time-window in the Gaze region in experiment 3.
(. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	0.97857	0.34574	2.830	**
c1	-1.04827	0.30800	-3.403	**
c2	-1.06369	0.42890	-2.480	*
c3	-0.37541	0.36370	-1.032	
l1	-0.42562	0.19377	-2.197	*
l2	-0.84948	0.19654	-4.322	***
c1:l1	0.05419	0.21796	0.249	
c1:l2	0.08392	0.16112	0.521	
c2:l1	0.38143	0.30462	1.252	
c2:l2	0.44953	0.19302	2.525	*
c3:l1	-0.14784	0.28182	-0.525	
c3:l2	-0.03479	0.19253	-0.181	

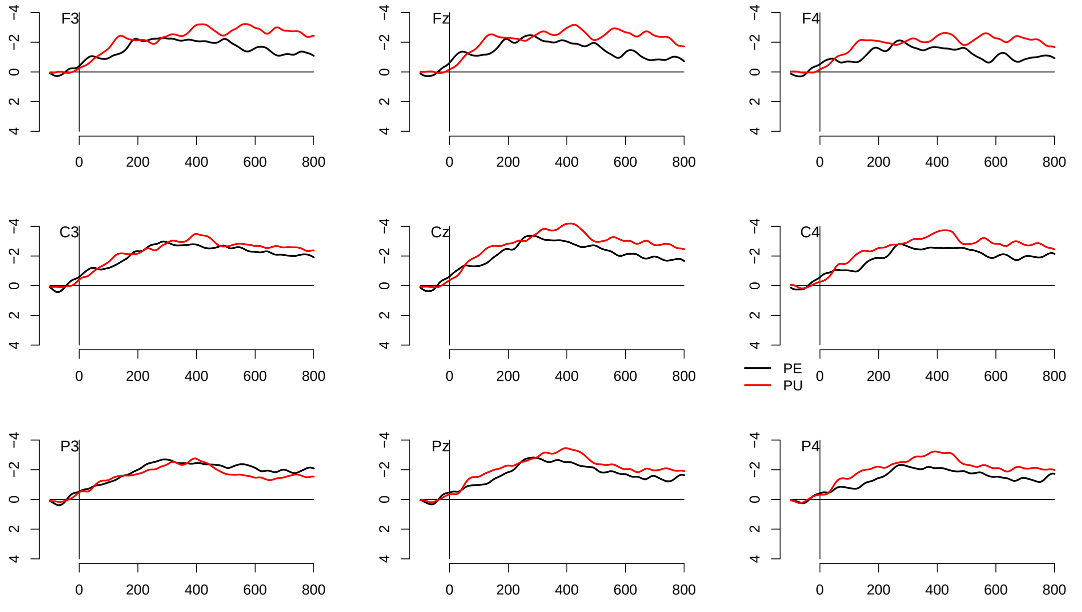


Figure 5.5: ERP time-locked to the Noun Onset in Experiment 3 separated by Expectability under Present Gaze (Present-Expected: black, Present-Unexpected: red.) The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

PMN (150 – 300ms)

In the PMN time-window, there was a main effect of Gaze Presence ($\beta = 0.68, SE = 0.33, t = 2.0, p < .05, CI = [-0.0085; 1.4437]$) with a more positive deflection for preceding Present gaze cues. This effect can most likely be attributed to the preceding gaze region, which showed a more positive movement of the ERPs in the P300 time-window, which overlaps with the baseline for the noun region. Above and beyond this effect however, the comparison of Expected and Unexpected continuations, which differ in the expectability of the gaze and consequent noun, led to a significant difference only when the preceding gaze toward the object was absent (conditions AE and AU). In this time window, unexpected nouns led to a more negative deflection when gaze was absent ($\beta = -1.36, SE = 0.50, t = -2.7, p < .01, CI = [-2.4024; -0.5467]$) as can be seen in Figure 5.6 on the next page. As reported above, no such differences where found in the noun region when preceding gaze was present as can be seen in Figure 5.5. (For a summary of the model results, see Table 5.5 on page 103.) The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure C.7, Figure C.8 (both on page 154) and Figure C.9 on page 155 respectively. The R code of the final model for this time-window was as follows:

```
final.Noun.PMN.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
                          (1 + c1 + c2 + c3 + l1 | subject) +
                          (0 + l2 | subject) +
                          (0 + c1:l1 | subject) +
                          (0 + c1:l2 | subject) +
                          (0 + c2:l1 | subject) +
                          (0 + c3:l1 | subject) +
                          (0 + c3:l2 | subject) +
                          (1 + c1 + c2 + c3 + l1 | item) +
                          (0 + l2 | item) +
                          (0 + c1:l1 | item) +
                          (0 + c2:l1 | item) +
                          (0 + c3:l1 | item),
                       REML=FALSE, data=EXP3.Noun.PMN.Data,
                       control = lmerControl(calc.derivs=FALSE),
                       na.action = na.omit)
```

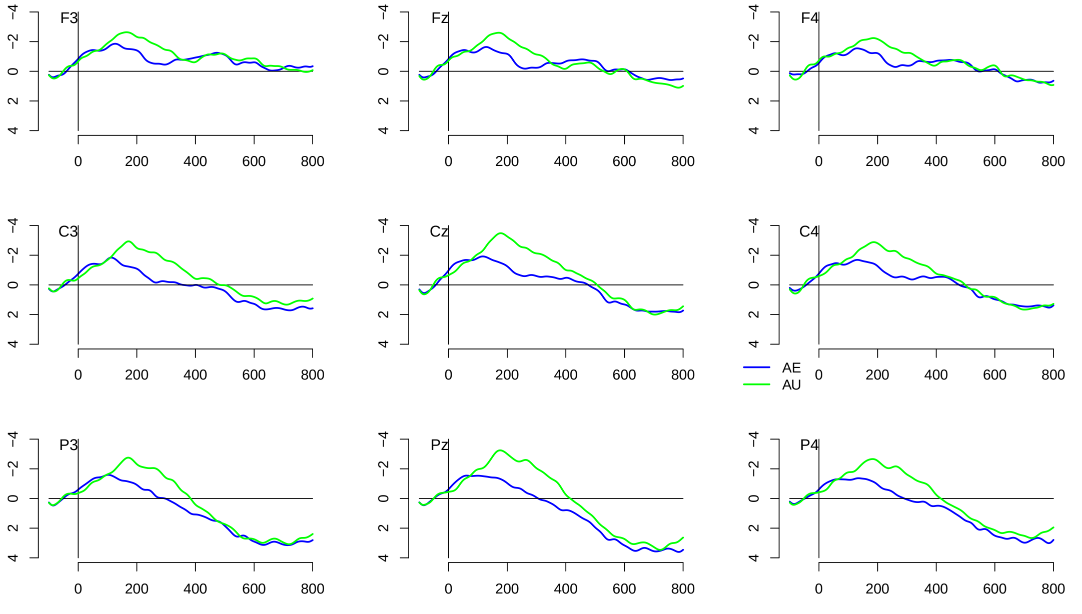


Figure 5.6: ERP time-locked to the Noun Onset in Experiment 3 separated by Expectability under Absent Gaze. (Absent-Expected: blue and Absent-Unexpected: green.) The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

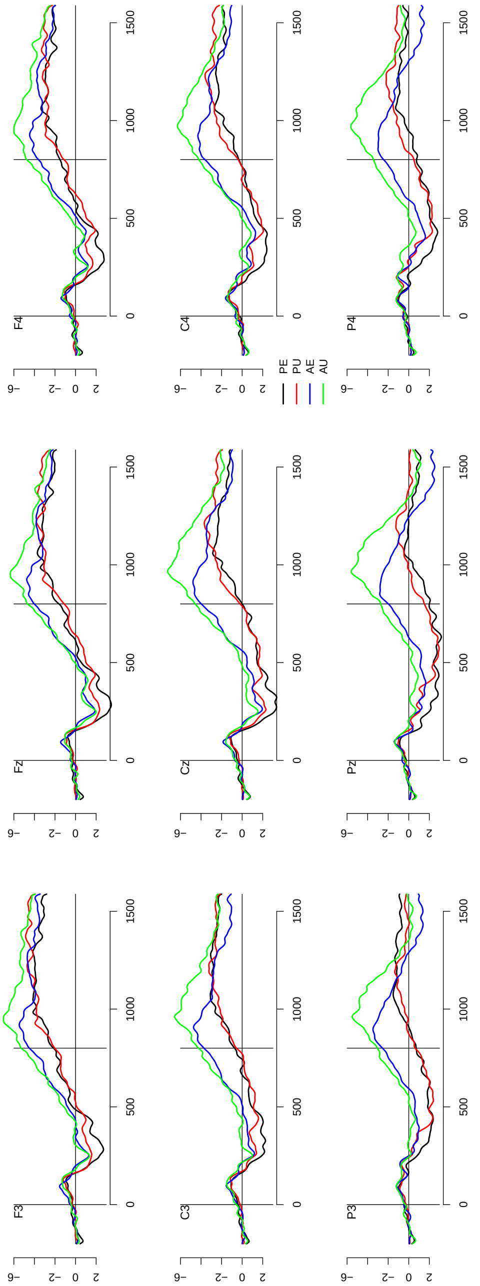


Figure 5.7: ERP time-locked to the Gaze Cue Onset in Experiment 3 separated by the 4 experimental conditions, spanning both regions of interest. (Present-Expected (black), Present-Unexpected (red), Absent-Expected (blue) and Absent-Unexpected (green). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4 filtered at 20Hz for presentation purposes only. (Negativity is plotted upward.)

Table 5.5: Fixed Effects for PMN time-window in the Noun region in experiment 3.
(. - $p < 0.1$, $*$ - $p < 0.05$, $**$ - $p < 0.01$, $* * *$ - $p < 0.001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.703833	0.275483	-6.185	***
c1	0.676556	0.332749	2.033	*
c2	-0.049757	0.437639	-0.114	
c3	-1.357151	0.495168	-2.741	**
l1	0.016384	0.185817	0.088	
l2	0.557507	0.111996	4.978	***
c1:l1	0.000697	0.251408	0.003	
c1:l2	-0.164593	0.153343	-1.073	
c2:l1	-0.014542	0.311906	-0.047	
c2:l2	0.084014	0.188377	0.446	
c3:l1	-0.383718	0.303006	-1.266	
c3:l2	0.315112	0.218423	1.443	

N400 (300 – 450ms)

In the N400 time-window, there again was a similar effect of Gaze Presence ($\beta = 2.18, SE = 0.54, t = 4.1, p < .001, CI = [1.0192; 3.4835]$) to the preceding time-window. This effect was strongest in centro-parietal regions ($\beta = 0.77, SE = 0.34, t = 2.2, p < .05, CI = [0.0852; 1.4853]$). The comparison of Expected and Unexpected continuations when gaze was Absent shows a negative deflection for Unexpected continuations only in centro-parietal regions ($\beta = -0.70, SE = 0.34, t = -2.2, p < .05, CI = [-1.3555; 0.0221]$). (For comparison, see Figure 5.6 on page 101.) A summary of the model results can be found Table 5.6 on the next page. The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure C.10, Figure C.11 (both on page 156) and Figure C.12 on page 157 respectively. The R code of the final model for this time-window was as follows:

```
final.Noun.N400.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
                          (1 + c1 + c2 + c3 + l1 + l2 | subject) +
                          (0 + c1:l1 | subject) +
                          (0 + c1:l2 | subject) +
                          (0 + c2:l1 | subject) +
                          (0 + c2:l2 | subject) +
                          (0 + c3:l1 | subject) +
                          (0 + c3:l2 | subject) +
                          (1 + c1 + c2 + c3 + l1 + c1:l1 + c2:l1 + c3:l1 | item),
                        REML=FALSE, data=EXP3.Noun.N400.Data,
                        control = lmerControl(calc.derivs=FALSE),
                        na.action = na.omit)
```

Table 5.6: Fixed Effects for N400 time-window in the Noun region in experiment 3. (. - $p < .1$, $*$ - $p < .05$, $**$ - $p < .01$, $***$ - $p < .001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-1.35214	0.37409	-3.614	***
c1	2.18637	0.53518	4.085	***
c2	-0.52315	0.47827	-1.094	
c3	-0.59339	0.54488	-1.089	
l1	0.20288	0.22387	0.906	
l2	0.71760	0.12416	5.780	***
c1:l1	0.76503	0.34281	2.232	*
c1:l2	-0.05357	0.26494	-0.202	
c2:l1	0.13393	0.36788	0.364	
c2:l2	0.33044	0.21316	1.550	
c3:l1	-0.69905	0.32493	-2.151	*
c3:l2	-0.00844	0.24454	-0.035	

P600 (600 – 800ms)

As for the previous time-windows, the P600 time-window also shows a main effect of Gaze Presence ($\beta = 3.43, SE = 0.57, t = 6.0, p < .001, CI = [1.9730; 4.7429]$) which is more prominent in centro-parietal regions ($\beta = 1.63, SE = 0.48, t = 3.4, p < .01, CI = [0.8295; 2.5942]$). In this time-window, no effect of Expectability was found. (For a summary of the model results, see Table 5.7 on the facing page.) The corresponding model plot, plot of means and the topographical scalp maps can be found in Figure C.13, Figure C.14 (both on page 158) and Figure C.15 on page 159 respectively. The R code of the final model for this time-window was as follows:

```
final.Noun.P600.model = lmer(eeg ~ 1 + (c1 + c2 + c3)*(l1 + l2) +
                (1 + c1 + c2 + c3 + l2 + c1:l2 + c3:l2 | subject) +
                (0 + l1 | subject) +
                (0 + c1:l1 | subject) +
                (0 + c2:l1 | subject) +
                (0 + c3:l1 | subject) +
                (1 + c1 + c2 + c3 + l1 + c1:l1 + c2:l1 + c3:l1 | item),
            REML=FALSE, data=EXP3.Noun.P600.Data,
            control = lmerControl(calc.derivs=FALSE),
            na.action = na.omit)
```

Table 5.7: Fixed Effects for P600 time-window in the Noun region in experiment 3. (. - $p < 0.1$, $*$ - $p < 0.05$, $**$ - $p < 0.01$, $***$ - $p < 0.001$)

Fixed Effect	Estimate	Std. Error	t value	sig.
(Intercept)	-0.16471	0.38735	-0.425	
c1	3.43266	0.57132	6.008	***
c2	-0.48065	0.54607	-0.880	
c3	-0.20360	0.51067	-0.399	
l1	0.76357	0.26404	2.892	**
l2	0.55883	0.23440	2.384	*
c1:l1	1.62503	0.48165	3.374	**
c1:l2	-0.02816	0.27660	-0.102	
c2:l1	0.75172	0.39708	1.893	.
c2:l2	0.47607	0.24724	1.926	.
c3:l1	-0.31782	0.35783	-0.888	
c3:l2	-0.06857	0.30284	-0.226	

5.4 Discussion

Unlike in the previous experiments, the gaze cue and the subsequent noun fill similar roles in this experiment. Both can potentially be utilized to confirm or disconfirm expectations based on the preceding linguistic context. While the noun is always present to fill this role, gaze can either be present or absent. Thereby, the first bit of information rendering the sentence either true or false potentially shifts between these two positions. In the following, the findings for the ERP components will be summarized and interpreted according to the underlying mechanisms they are indexing.

5.4.1 PMN

To recall: an effect in the PMN time-window was only found in the noun region when preceding gaze was Absent. In line with the interpretation of the PMN effect as an informativity driven phonological matching mechanism, visual information as provided by the gaze cue should and does not show this effect as supported by both the difference wave approach and the moving time-window analysis. In the noun region, however, the PMN modulation was observed, similar to the previous experiments. In line with the previous interpretation of the PMN, it can be argued that, in the present experiment as well, this modulation expresses the processing of an acoustically perceived phoneme that was unexpected given the context. This is further supported by a lack of a PMN effect in the noun region when preceding gaze was present. The gaze could already be used to disconfirm the previously formed expectations. The effects on the noun then can be explained in two subtly different ways: An informative gaze cue could overwrite the necessity of linguistic continuations entirely, so that the naming of the referent is no longer required to understand the utterance. Alternatively, an informative gaze cue could lead to a shift in expectations to hearing the noun corresponding to the gazed at object. In other words, the gaze cue could be utilized to disconfirm the expectations formed based on the preceding linguistic content, entailing the retrieval of the gazed-at object as well as the revision of the mental representation of the situation. Additionally, as it was argued before that expectations about the upcoming referent are derived from the current state of the mental representation of the situation, the adjustment of that representation should also entail adjustments to the expectations about the upcoming referent. This would support the view of a more fluid system of information integration that is constantly monitored and updated.

5.4.2 N400

In both the gaze region as well as the noun region, a distinct negativity was found for unexpected gaze and nouns respectively. Importantly, the N400 in the noun region was only present when no preceding gaze was provided. Even though the time-window for the N400 in the gaze region differs slightly from the time-window in the noun region (250 – 450ms and 300 – 450ms respectively), it is similar to other reported time-windows of N400s in response to pictures (e.g., 270 – 420ms (Nigam et al., 1992). Again in line with the previous experiments, the N400 effect in the Unexpected conditions can be interpreted to be induced by expectation violations. As for the PMN time-window, the N400 effect in the noun region is only present when preceding gaze was absent. Similar to the explanation in the PMN time-window, this can be explained by a shift of expectations based on the preceding disconfirmation that was already possible in the gaze region when gaze was available. Hence, following the previous interpretation, as the gaze cue was utilized to update the mental representation of the situation, the expectations derived from the current state of that representation shifted to the gazed-at object rather than the referent supported by the preceding linguistic content.

5.4.3 Positive Deflections

The early starting positive deflection that only occurs when gaze is present can be interpreted as the integration of the perceived information with the mental representation (Donchin, 1981). The positive deflection found in the gaze region can reasonably be interpreted as a P300 and, more precisely, a P3b. Although, the P300 time-windows in the literature often differ based on the modalities of the respective experiments,[2] these time-windows are often similar to the here reported time-window of 250 – 800ms.

In the previous experiments, a positive deflection could be seen on the noun when the mental representation of the situation needed to be updated. This was only the case in the Incongruent conditions in those experiments. Unlike in the previous experiments however, here the positive deflection does not only occur for information that requires an update of the model — as would be the case for the Unexpected conditions — but generally when the subsequently named object was identified. This was the case for both the Expected and Unexpected conditions. Similar to this finding in the gaze region, a positive deflection in the noun region is only present when preceding gaze was unavailable. Again, this effect can be attributed to the integration in the mental representation of the situation. As both positive deflections (P300 on the gaze cue and P600 on the noun) are interpreted to index similar mechanisms of meaning integration/update, these findings are in line with the P600-as-P3 hypothesis. However, similar to the corresponding effect in the gaze region, the P600 in the noun region occurred in both Expectability conditions under Absent gaze. While the effects in the earlier time-windows (PMN and N400) are in line with the predictions for gaze to be utilized to evaluate expectations based on the preceding linguistic content of an utterance, the findings in the positive deflections are not explicable by the hypothesized updating of a mental representation alone as was deducted from the results from the first two experiments. This difference form the previous experiments will be discussed in more detail in the General Discussion (Chapter 6).

5.4.4 Summary

When only regarding the noun region with an absent preceding gaze cue, this experiment replicates findings from the previous experiments for the early components. The PMN effect related to the matching of the acoustic signal with expectations about the upcoming word was followed by an expectation driven N400 effect. The difference in scalp distribution

[2]For example, Comerchero and Polich (1999) report differences in the presentation of auditory (250 – 450ms) and visual (350 – 600ms) stimuli.

found in the PMN and N400 time-windows adds additional evidence for the distinction of the two effects. While the PMN showed a more globally distributed effect, the N400 was centroparietally distributed, which is a typical distribution for the classical N400 effect.

Most intriguing, however, is the shift of effects between the gaze and noun region. When gaze was present, effects were only found in the gaze region. The subsequent noun region then did not show any significant differences that where previously found. This suggests that gaze in Experiment 3 was interpreted as confiriming or disconfirming linguistic expectancy, much as the noun did in Experiments 1 and 2. When gaze was absent, however, the effects established in the previous experiments were present in the corresponding time-windows on the noun. This suggests that gaze is incrementally and immediately used to confirm expectations and to inform the mental representation of the situation, so that the representation can be adjusted and used to generate revised expectations for the subsequent referent. In other words, the results show that gaze is interpreted as part of the communicative signal that is rapidly utilized to inform the mental representation of the situation in an incremental manner, much as the speech signal.

5.4.5 Conclusion

Following the evidence for the utilization of speaker gaze by listeners to anticipate the mentioning of a specific referent provided in the first two experiments in Chapter 4 on page 37, the experiment in this chapter additionally provided evidence that linguistic context — by virtue of expectations for a specific referent — can similarly result in expectations for a specific gaze cue. This further underlines that gaze is used beyond the shift of attention or the increase of a gazed-at object's or position's prominence and, hence, supports the Situational Integration Account over the Promince Account. The following chapter will discuss the results of all presented experiments in context and further elaborate on the differences found in the postivie deflections across the three experiments.

CHAPTER 6

General Discussion

The ERP experiments presented in this thesis were aimed to investigate the underlying mechanisms involved in the integration of gaze into the situation model as reflecting speaker intentions resulting in expectations, based on the evidence from behavioral data in the literature that suggests that speech aligned speaker gaze facilitates comprehension. The mechanisms indexed by the reported ERP components were investigated in relation to two proposed accounts that may explain the reported behavioral effects. The Prominence Account assumes that improvements and inhibitions of sentence comprehension stem from the allocation of attention to the gazed at position where comprehension benefits from the increased prominence of the attended object, when it is subsequently named and is inhibited if attention is drawn away from the object.

The Situational Integration Account assumes a processing of the gaze-at object beyond an increase in prominence, including its semantic retrieval as well as its integration in the mental representation of the situation. The latter further is assumed to entail anticipation for the gazed-at object to be referred to next. The account predicts the involvement of three underlying mechanisms of language comprehension: A phonological matching mechanism is assumed that matches the perceived signal with the expected word-form. Further, semantic retrieval is predicted to be inhibited if a referent other than the expected one is encountered. The Situational Integration Account further assumes an integration mechanism, which utilizes the perceived input incrementally so that the gazed-at object is already integrated with a mental representation of the situation. Hence, this mechanism is inhibited when the subsequent referent is different from the expected one, eliciting the necessity to revise the mental representation. (For a full description of the two account see Chapter 4.3 on page 42).

The findings presented for the first two experiments are summarized in Table 6.1 on the next page. They suggest that gaze is used to form expectations about the upcoming referent: An attenuated PMN for Congruent gaze preceding the N400 time-window suggests that expectations are not only formed on a conceptual level but also about the concrete lexical form when a single object is highlighted. An N400 effect when gaze is uninformative or misleading indicated an increased cost of semantic retrieval. The additional findings in the second experiment regarding the Mutual gaze cue — as displayed by a straight gaze to the participant — provides further evidence to distinguish between the two processes indexed by the PMN and N400. It is important to note, however, that the results also suggest a strong interplay between these two components. The relatively short lived N400 (300 – 450ms) following the PMN indicates that the retrieval of the full word benefits from the phonological matching as indexed by the PMN. This is a reasonable assumption following Hagoort and Brown (2000), who argue the PMN expresses a lexical selection process. It could be argued that some properties of the referent can be retrieved at this earlier phonological stage relieving the mechanism of word retrieval. Hence, it can be speculated that the presence or absence of the PMN might have an effect on the strength/duration of the N400. This, however, requires further investigation. This could possibly be done by replacing the auditory presentation with visual word presentation. This would eliminate the possibility of rapid phonological matching proposed to evoke the PMN effect. The difficulty of this approach, however, would be to keep the stimuli as similar as possible. As a centrally presented face performing gaze actions to the objects was utilized in the presented experiments, a central word-by-word presentation, usually utilized for research on written word comprehension, might be problematic.

Table 6.1: Summary of the effects in the noun region for experiments one and two. (C - Congruent ; I - Incongruent ; A - Averted ; M - Mutual)

Experiment	Component	C-I	C-A	C-M
Experiment 1	PMN	+	+	
	N400	+	+	
	P600	+	-	
Experiment 2	PMN	+	+	+
	N400	+	+	-
	P600	+	-	-

In the P600 time-window, the results suggest that the visual scene and speech signal as well as speaker gaze are used to construct a mental representation of the situation. This is consistent with the view that gaze is interpreted as conveying referential intentions (Staudte and Crocker, 2011). In the first two experiments presented in this thesis, the first gaze action in each experimental trial correctly provided evidence about the upcoming referent. In case of a following Incongruent gaze, participants were led to believe that the gazed at object actually would be the upcoming noun, eliminating the remaining objects in the scene as likely referents. The upcoming noun however forces the participant to revise the representation such that the referent is not the one gazed at, but rather the spoken one. This substantial updating of the situational representation is then reflected by a P600 modulation representing the (re-)integration difficulty. No such difference is induced in the Uninformative conditions as either upcoming referent is still possible and, hence, does not require a revision of the situation model. While the N400 for both the Incongruent and Averted conditions was consistent with both the Prominence and Situational Integration Accounts, the observation of a P600 only in the Incongruent condition was predicted by the Situational Integration Account alone.

The results of the third experiment, however, provided support for an additional function of the P600. The results of Experiment 3 are summarized in Table 6.2. Here, a P600 was found in response to both expected and unexpected continuations in the noun region. Thus, the previously assumed connection to only a revision of the mental representation when assumptions are disconfirmed seems to not entirely capture the actual indexed mechanism(s). It could be argued that the P600 also indexes a more broad monitoring mechanism of the mental representation that governs the revision of the mental representation as well as its evaluation in order to form (task specific) decisions.

Table 6.2: Summary of the effects for experiment 3 for the Unexpected conditions relative to the Expected conditions.
P - Present gaze , A - Absent gaze , E - Expected continuation , U - Unexpected continuation

Region	Component	P-A	PE-PU	AE-AU
Gaze	P300	+	-	-
	N400	-	+	-
Noun	PMN	-	-	+
	N400	-	-	+
	P600	+	-	-

Similarly, effects of the proposed monitoring of, and integration with, the mental representation of the situation can also be found on the gaze cue, which are manifested as a P3b effect instead. The P3b as indexing mental integration was also reported by, e.g., Donchin (1981) and is strongly resembling the interpretation of the P600 as by, e.g., Burkhardt (2007) and Brouwer et al. (2017). This is also in line with the P600-as-P3 hypothesis and, further, the LC/NE-P3 (e.g., Sassenhagen et al. (2014)). Additional support for this interpretation of the P3b on the gaze cue was found in the subsequent noun region. Only when gaze was absent, similar effects as found in the previous experiments were present on the noun (PMN, N400 and P600). When a gaze cue was preceding the mentioning of a referent, no significant effects of these components were found on the noun itself. This strongly suggests that the corresponding referent — even if not supported by the linguistic context — was already retrieved and integrated with the mental representation. As argued in Chapter 5 on page 87, expectations about the upcoming referent are proposed to be derived from the current state of the mental representation of the situation. Hence, the revision of the mental representation of the situation should also entail updated expectations about the upcoming referent. This supports that gaze is not only understood as part of the communicative signal, but also that the utilization of speaker gaze is a rapid and incremental process that occurs simultaneously with the speech perception to the extent that one influences the other.

Another result concerning the N400 component found in Experiment 2 was the significant modulation in the noun region for the Averted gaze condition compared to the Congruent condition that is absent for the Mutual condition, while both Uninformative gaze versions replicate the effects in every other time-window. Beyond supporting the claim of the distinctiveness of the PMN and N400 component, this raises the question about the underlying nature of the different perceptions of the two Uninformative gaze cues, which remains open for further investigation.

Overall, the presented results strongly suggest that gaze does not only support language understanding but can be an integral part of situated communication, even to the extent that gaze can be utilized in similar fashion as words to communicate meaning in context. As such, the results strongly support the assumptions made under the Situational Integration Account rather than the Prominence Account: If gaze was simply increasing an objects or positions prominence, neither should there be effects related to the mental representation of the situation, nor should the effects on the subsequent referent disappear. While the gaze to an object under the prominence account might benefit the phonological matching mechanism (PMN) and word retrieval (N400), the integration with the mental representation of the situation should still be necessary on hearing the referent and postpone decision making to this point. This however is not evident from the reported findings, supporting the Situational Integration Account.

In the following, each of the identified underlying mechanisms of language comprehension and their indexing components will be discussed separately in more detail.

6.1 Phonological Matching Mechanism / PMN

The early negative component found in the noun regions across experiments between 150 and 300ms can be plausibly interpreted as a Phonological Matching Negativity (PMN) (e.g., Connolly and Phillips (1994); Hagoort and Brown (2000), as described in Section 4.6 on page 81. The effects found in the PMN time-window across experiments suggest that the PMN is influenced by the amount of information conveyed by the perceived phoneme so that either a mismatch as well as the support of the selection of a referent lead to increased PMN responses as, in both cases, the phoneme provides more information than if an expected phoneme is perceived. This is in line with the interpretation of the PMN as reflecting 'the lexical selection process that occurs at the interface of lexical form and contextual meaning' (Hagoort and Brown (2000), p. 1528). This interpretation of this early component as a PMN is further supported by findings in the third experiment where the component was only present for auditory input (noun region) but not for visual input (gaze region) although both were utilized to confirm expectations formed on the preceding linguistic context. The experiments presented in this thesis further extended the findings of previous research by showing that the expectations necessary to evoke this component can not only be derived from linguistic context, but also from visual context as, e.g., speaker gaze. (For further elaboration on the PMN component and its interpretation see Section 3.2.4 on page 25 and Section 4.6 on page 81.)

While the results of the presented experiments indicate anticipation of a specific word form, based on the interpretation of the PMN to depend on the amount of information conveyed by a phoneme, a somewhat different explanation is possible concerning the nature of the expectations: It is possible that, rather than a specific word form, mainly semantic expectations are formed. It is possible that those expectations then activate suitable lexical candidates that are matched with the auditory input, especially in a strongly constraint context as provided in the experiments. An increased PMN response could then be expected when the phoneme doesnt match either of the activated candidates, requiring a revision of the expectations. It would, however, also be reasonable to assume that anticipated semantic content for which multiple lexical candidates exist in the mental lexicon (e.g.: synonyms), would elicit an increased PMN response even when an activated candidate fits the perceived phoneme as the actually uttered candidate has to be selected from the set of possible candidates, possibly even influenced by word frequency, etc. The objects utilized in the presented experiments where controlled so that their naming was unanimous amongst participants in the pre-test. Uncertainty about the noun used to refer to the gazed-at object was therefore most likely rather small. Thus, in the Expected/Congruent conditions, a specific word form could be anticipated because, for the anticipated semantic content, no alternative referent was available. Thus, it stands to reason that an increased PMN modulation could possibly also be elicited in response to nouns for which the number of competing referents is larger, resulting in higher uncertainty about the actual referent, hence, increasing the amount of information conveyed by the phoneme.

Alternatively, it could be argued that as all possible candidates for the naming of anticipated semantic content are activated, no increased PMN response is elicited for any of those referents. Hence, as long as a perceived referent was activated based on the expected semantic content, no effect in the PMN would be anticipated. While in the results of the presented experiments an increased PMN response was also found for the Uninformative conditions — which could be a result of the activation of any possible referent for either of the two remaining objects on the screen — the PMN could still be interpreted in the same way. While under this interpretation an attenuated PMN would be anticipated for a set of possible referents to the same semantic content, the Uninformative conditions require the selection of not only a specific referent, but also of the correct semantic content (here the correct object). This interpretation of a more semantics-driven dependence of the PMN would also support the interpretation of the interplay of the PMN and the N400: If the PMN expresses the process of selection of the correct semantic content based on the physical properties of the input, a reduced N400 (in both latency or amplitude) seems reasonable considering the N400 to index semantic retrieval. I.e.: An increased PMN response is anticipated when the perceived phoneme does not match with any of the activated candidates for the expected semantic content. Thus, the adjustment of the expectations to another semantic content which has candidates that fit the physical form of the perceived word are required. It could be argued that this adjustment further entails some early form of semantic retrieval that benefits the full retrieval as indexed by the N400 component. Either interpretation of the PMN, however, requires further investigation. E.g.: It would be possible to utilize objects for which multiple referents are available in order to investigate whether the PMN is sensitive to different referents for the same semantic content.

In short, the presented results show that, when presented with speech, listeners use the phonological signal provided to match it with the expected word form or lexical candidates for an expected semantic content, and further that those expectations can be derived from different sources such as visual context (gaze in experiments one and two (Chapter 4 on page 37)) or linguistic context (experiment 3 (Chapter 5 on page 87)).

6.2 Semantic Retrieval Mechanism / N400

The presented experiments support the interpretation of the N400 to reflect word meaning retrieval effort based on expectations arising from contextual information (Van Berkum, 2009; Kutas and Federmeier, 2011; Schumacher, 2012). As with the PMN, the presented findings provided evidence that this contextual information does not necessarily have to be purely linguistic in nature but can also be derived from different sources such as speaker gaze. While experiments one and two enabled a utilization of speaker gaze to form predictions about the upcoming referent, in experiment 3 contextual expectations for a specific referent resulted in a corresponding expectation for gaze cues preceding that referent. In the first two experiments, the N400 modulation found in the noun region following the gaze cue showed that the gaze cue was in fact understood to reflect the speaker's referential intention and that a misleading gaze cue led to inhibited word retrieval.

The interpretation of the N400 to reflect expectation-driven retrieval difficulties (Van Berkum, 2009; Kutas and Federmeier, 2011; Schumacher, 2012) is also consistent with the findings in the gaze region in the third experiment. The gaze to an object that was not supported by the linguistic context led to a higher N400 amplitude than such that is supported. This does not only support the previous findings suggesting the involvement of the described mechanisms, but further enriches it: As the N400 effect found in the gaze region suggests that the meaning of a gazed-at object is retrieved already at this point, it can be argued that the attenuated N400 in the noun region following object-directed gaze — also in the previous experiments — does not index an easier word retrieval but instead that the meaning of the word was already retrieved utilizing the gaze cue. An increased N400 effect, in turn, then indexes the necessity to retrieve another meaning than the already retrieved one.

Further, there was a difference in the N400 between the two Uninformative conditions in Experiment 2. While the gaze toward the empty position in the scene (Averted) evoked an N400 effect in the subsequent noun region similar to the effect found in the first experiment, the gaze redirected to the listener (Mutual) showed an attenuated N400. This could possibly be explained by either a distraction of the listener in the Averted condition that draws the attention to the empty position or, alternatively, in lines of the Feeling Of Another's Knowing (FOAK), where a mutual gaze is claimed to convey a higher amount of certainty about the content of the utterance. (For a more in-depth dexcription see Section 4.6 on page 81). Also supporting and related to this interpretation, the work reviewed in Chapter 2.3 on page 10 reported an increased salience of gaze cued positions when they were preceded by mutual gaze (Meltzoff et al., 2010; Böckler et al., 2011). While, in the presented experiments, no gaze cue to a position followed this mutual gaze, those studies underline the effect of mutual compared to averted gaze.

6.2.1 Monitoring Mechanism / P3b/P600

As described in Section 4.6 on page 81, the findings of the first two experiments support the interpretation of the P600 as indexing a mechanism of integration with sentence meaning as represented by a mental model of the situation (Van Berkum et al., 2007; Burkhardt, 2006, 2007) that is inhibited when the perceived signal does not fit the current representation of the situation and, hence, requires a revision of the model.

The results from the third experiment additionally suggest that the P3b as found on the gaze cue and the P600 as found on the noun, index similar effects in their respective regions. This would suggest that the P600 possibly is a member of the P300 family, as has been argued by, e.g., Coulson et al. (1998) and Sassenhagen et al. (2014). However, unlike in the first two presented experiments, the effects were not only present when a revision of the model was necessary (Incongruent/Unexpected conditions), but also for the Expected conditions. This difference across experiments as well as the relationship of the P3b and

P600 components and the resulting implications will be discussed in greater detail in the following section.

6.2.2 Differences in the Positive Deflections

The results from the first two experiments show a P600 effect only in the Incongruent condition, which is in line with the proposed interpretation of the component to express integration difficulties with the mental representation of the situation as also proposed by Van Berkum et al. (2007); Burkhardt (2006, 2007) and Brouwer et al. (2012). In the third experiment, however, regardless of Expectability, the P600 at the noun (when gaze was absent) as well as the P3b on the gaze (when present) where always present. As the other reported effects (PMN, N400) were in line with the interpretations proposed for the effects in the first two experiments, this raises the question why and how the effects found in the P600 time-window — and similarly in the P3b time-window on the gaze cue as they are interpreted to index the same underlying mechanism respectively— differed from the results of the previous experiments.

To possibly explain this difference, it is important to point out another difference in the role of the respective regions in the two types of experiments. The main change was the different roles gaze and linguistic context took in the experiments. While in Experiment 1 and 2 in Chapter 4 on page 37 gaze could be utilized to form expectations about the upcoming referent in the unfolding utterance, in the last experiment, a gaze cue could instead be anticipated, based on linguistic context expectations for a referent. In order to change these roles, the sentences that were presented with the visual scenes were altered. While the sentences in the first two experiments were of the form *"Verglichen mit dem Haus, ist das Auto verhältnismäßig klein, denke ich"* (*'Compared to the house, the car is relatively small, I think'*), the sentences in the third experiment were of the form *"Das Haus ist größer als das abgebildete Auto, denke ich"* (*'The house is bigger than the displayed car, I think'*). This could be more generally expressed as a sequence of **NP1–Gaze <Predict>–NP2–Comparative** for the former experiments and **NP1–Comparative <Predict>–Gaze–NP2** for the latter experiment.[1] When inspecting the order of the utterances' elements, it becomes clear that, besides the shift in role of the linguistic context and gaze, also the point of sentence evaluation shifts. In the first and second experiment, whether the sentence was true or false was only detectable once the comparative was uttered. Hence, neither the gaze region nor the noun region where sufficient for this evaluation process. In the third experiment, however, the new word order also lead to an overlap of the expectation confrimation and the sentence evaluation. On the example of the sentence *'The house is bigger than the displayed car, I think'*, in order for the participants to decide whether the sentence was true or false

[1] The aditional tag <Predict> is added to the Element that provides grounds for the forming of expectations.

given the visual context, they needed to hear which of the objects is actually compared to the *'house'*. Gaze, when present, was used to confirmation the expectations, however, it could additionally be used to evaluate the sentence at that point. The very same holds for the noun region in this experiment when gaze was absent. This overlap of expectation confirmation with sentence evaluation could possibly explain the different patterns in the P600 distribution.

While there were no relevant effects on the comparative between conditions[2] in any of the presented experiments, in order to investigate the difference in the P600 time-window between experiments, it is worth aligning the regions according to their appearance in the experiments. Figure 6.1 aligns the ERPs of the corresponding regions for Experiment 2 (which showed similar effects as Experiment 1) and Experiment 3 beginning with the region used to form expectations (gaze in Experiment 2 and the comparative in Experiment 3). For Experiment 2, this is followed by the region used to confirm the expectations (noun) and finally the element used to evaluate the sentence (comparative). For Experiment 3, the second ERP shows the gaze cue while the last one shows the noun. Both gaze and noun in Experiment 3 are used to confirm and evaluate the sentence depending on whether gaze was present or absent.

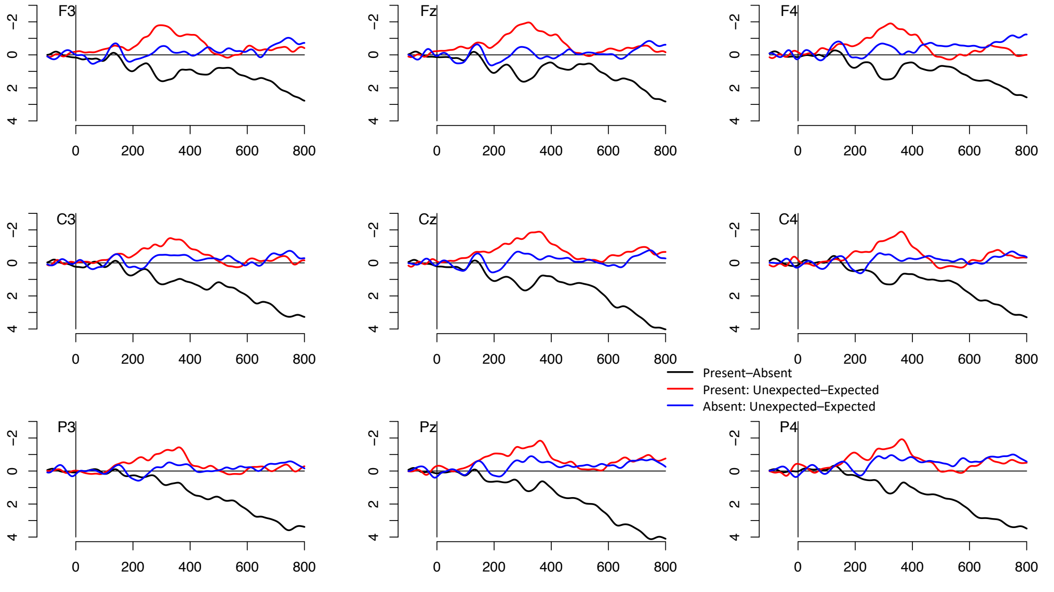


Figure 6.1: Regions of interest aligned by their appearance across the experiments. *Both only when gaze present. **Both only when gaze absent.

[2]The comparative in the first two experiments differed not only by the reported experimental conditions but further rendered the sentence true or false. Hence, this region suffered from data sparsity and was, thus, not reported. In Experiment 3, no effects were found on the comparitor.

What can be seen is that, when evaluation of the sentence is possible — on the comparative in Experiment 2 and gaze or noun in Experiment 3 — the corresponding region shows an overall more positive deflection. This is illustrated in Figure 6.2, which aligns the ERPs according to their role. Importantly, the same region does show no effects when used to form expectations and does not show the same strong positive deflection (gaze cue in Experiment 2 and comparative in Experiment 3). This indicates that the positivity stems from a process related to the evaluation mechanism rather than from a property specific to the linguistic content of the corresponding words.

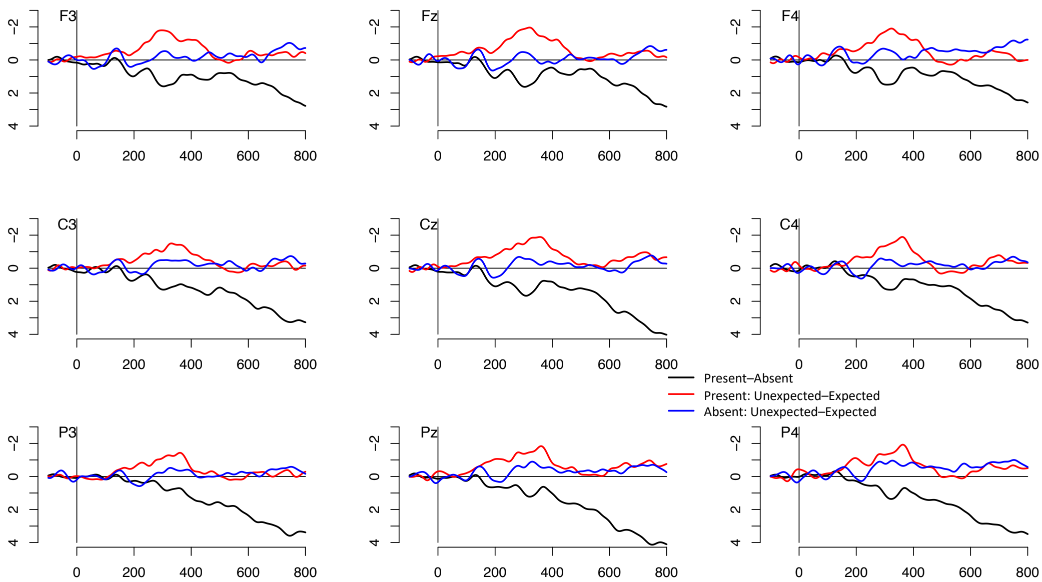


Figure 6.2: Regions of interest aligned by their role across the experiments. *Confirmation only for experiment 3.

This could be interpreted in a way such that overall the positive deflection on the comparative in Experiment 2 and the gaze/noun in Experiment 3 respectively expresses the sentence evaluation mechanism necessary to fulfill the task of the trial. Recall that each trial in all experiments required the participant to judge whether the sentence was true or false. This is in line with interpretations of the P300 (see Picton (1992) for a review) and P600 (Hahne and Friederici, 2002; Haupt et al., 2008; Schacht et al., 2014) as reflecting effects induced by the task. Such task-induced positivities are, however, often reported only for anomalous or unexpected condition. The positivities found in Experiment 3 — both on the gaze and noun in their respective conditions — where, however, present for both expected and unexpected referents. This might be considered to be more in line with a related interpretation of the P3 as reflecting decision making through the involvement of the Locus Coeruleus-Norepinephrine (LC-NE) System (Pineda et al., 1989; Pineda, 1995; Nieuwenhuis et al., 2005). Decision making, in this context, is understood as the identification and processing of a task-relevant stimulus and its mapping to the appropriate response (Ratcliff, 1978; Gold and Shadlen, 2001; Nieuwenhuis et al., 2005). The argument that the P600 represents such a so called LC/NE-P3 was, for example made by Sassenhagen et al. (2014) and, similarly, by Coulson et al. (1998).

While the interpretation of the observed positivities as reflecting decision making seems reasonable for the results in Experiment 3, it could not explain the findings in the P600 time-window in Experiment 1 and 2. It could be argued that the positivities in the respective experiments index different mechanisms (situational integration in Experiments 1 and 2, and decision making in Experiment 3). However, as there was no measurable difference between the unexpected and expected conditions in Experiment 3, which could be expected if the two processes of reevaluation of the mental representation and decision making overlap, the two effects could also be more closely related. It could be argued that both, situational integration as well as decision making, require the involvement of a representation of the situation. While the P600 observed in Experiment 1 and 2 expresses the necessity to revise the current representation of the situation, the positivities in Experiment 3 express the evaluation of the situation representation in order to make a decision (whether the sentence was true or false in the presented experiments). If that was the case, then it is reasonable to assume that the evaluation of the situational representation subsumes its revision.

This interpretation would be in line with an explanation for the P3b as proposed by Verleger et al. (2005). They propose that the P3b might reflect a system of monitoring whether the first decision to classify a stimulus and (prepare to) act accordingly led to the correct processing. This interpretation could be linked to the proposed situational integration account so that if the monitoring system encounters input that does not fit the current representation of the situation (noun in Experiments 1 and 2), an increased positivity is elicited, indexing the need to adjust the representation. Eventually, when the situational representation reaches the state to evaluate into a decision (gaze and noun in Experiment 3), this is also expressed by the monitoring system. Additionally, this is in line with findings that suggest task-relevance to be an eliciting factor for such a type of positive deflection (Duncan-Johnson and Donchin, 1977) and findings of error related P3 occurrences (Falkenstein et al., 2000). To disentangle these interpretations of the P3b and P600, as indexing both an revision as well as a monitoring mechanism, future work could separate the regions on which sentence evaluation and expectation confirmation are proposed to happen to investigate whether such a positivity would be found for both mechanisms. One possibility to achieve this might be to introduce a Congruency factor, as used in the first two presented experiments. As the gaze cue in Experiment 3 was always Congruent, participants could completely rely on the gaze cue, providing certainty about the continuation. If gaze was less reliable, participants might require the naming of the referent to form a decision on the correct answer, rather than being able to evaluate the sentence on the gaze cue. This could additionally serve to further investigate the extent to which the reported effects are influenced by a cues reliability. Introducing a Congruency factor could further provide support for the proposed incremental updating and revision of the mental representation of the situation utilizing gaze cues equal to linguistic content: If gaze is used identically to linguistic content in terms of how it informs the mental representation of the situation, even if the referent is necessary to be certain about the sentential content, a full representation of the situation would be reached after the gaze to an object. This is because at that point, both objects and the comparative would be part of the mental representation of the situation. If the following referent was not the same as the gazed at object, this full representation would be required to be revised. Following the findings presented in this thesis, two possible outcomes are predicted: a) If gaze is utilized identical to linguisitc content to inform the mental representation, an increased P3b response would be expected for both Expected and Unexpected gaze cues, indexing the evaluation process controlled by the monitoring mechanism. This should then be followed by another increased P600 modulation on the referent, indexing the required re-evaluation of the mental representation. b) If the utilization of gaze to inform the mental representation was dependent on the reliability of the gaze cue, a reduced reliability could lead to an increased P3b response only when the gazed-at object was not expected given the preceding context, while the following referent should elicit an increased P600 response in any condition as the final evaluation should always be in this region.

In sum, while there remain numerous open questions to be explored for future research, the presented studies demonstrated robust evidence that gaze is processed as an integral component of the communicative signal. In the following, a theoretical model will be outlined that incorporates the findings and interpretations.

At any given time in a situated verbal interaction, the context interlocutors were presented with — as for example the beginning of an unfolding utterance, the present objects, speaker gaze, etc. — was used to inform a mental representation of the situation. If the current state of the mental representation provides grounds to anticipate an upcoming referent, listeners utilize this to form expectations. Hence, there are two possibilities of processing for any new element of the communicative signal listeners encounter: a) when the current state of the mental representation enables anticipation of upcoming referents, and b) when it does not. If expectations were formed, they are utilized to form an extended mental representation of the situation that already in part assumes the expected referent to be part of the situation. The expectations are then further matched with the actual incoming signal.

There again are two different paths of processing of that signal depending on the nature of the encountered element that differ in the initial processing step: In case of an auditory signal, a phonological matching mechanism matches the incoming initial

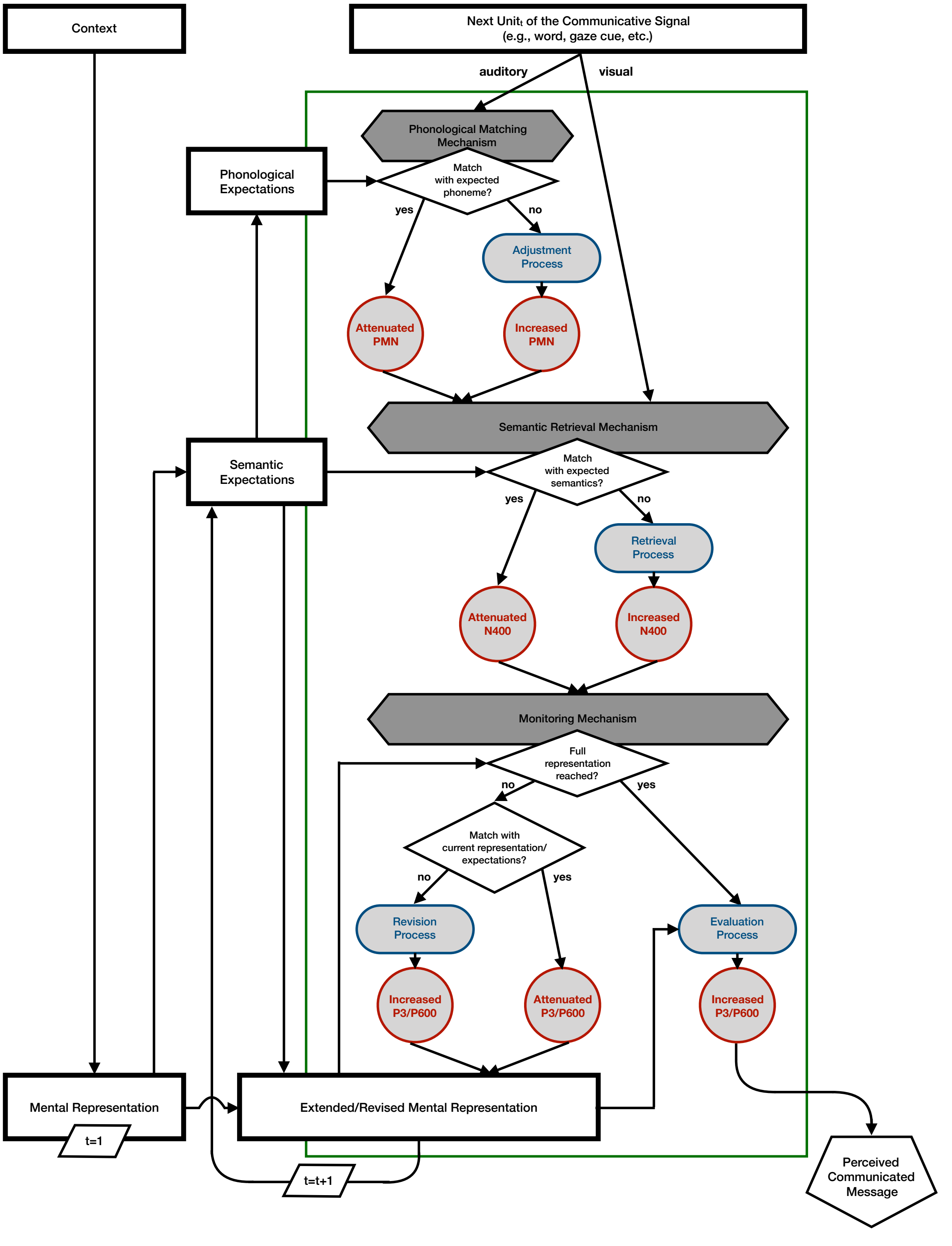


Figure 6.3: Outline of the proposed model.

phoneme with the expected phoneme in a constrained context. It is proposed that, if no specific word can be predicted (e.g., because possible synonyms could be used, etc.), the anticipated semantics — and, possibly, the word candidates activated by these anticipations — are matched with any candidate activated by the perceived phoneme. If a mismatch is detected, an increased PMN indexes the primary adjustments of the expectations based on the physical form of the stimulus influenced by the amount of information contained in the phoneme. This means that unexpected phonemes carry more information as they entail revision to the actual referent. The output of this mechanism then further informs a semantic retrieval mechanism which is indexed by an N400, where an increased response signifies the necessity to retrieve the perceived referents semantics as they do not match the anticipated semantics based on the expectations.

The output of the preceding phonological matching mechanism supports this retrieval so that — given auditory input — the adjustment to the actual referent based on the physical form of the input eases the retrieval of the semantics as indexed by a shorter N400 effect as compared to effects found in response to, for example, written words. This semantic retrieval mechanism is also involved for non-auditory elements of the communicative signal, such as gaze to objects and, potentially, other stimuli that are part of the communicative signal (e.g.: gestures, etc.). If, however, no phonological matching was possible the latency of the N400 effect is more similar to the classical N400 as in response to written words.

The retrieved semantics are then integrated with the current representation of the situation governed by a monitoring mechanism. This mechanism constantly monitors the current state of the mental representation of the situation and, once a full representation of the situation is reached, evaluates the content of the that representation to arrive at an interpretation of the communicated signal. This evaluation process is indexed by a positive deflection that can be expressed as an P3b or P600 effect depending on the effects of the preceding mechanisms. If, however, the current element of the communicative signal does not lead to a full representation of the situation, the element is integrated with the current mental representation. If the element does not match the expectations and, hence, the current extended representation of the situation, an revision process updates this representation. This revision process is indexed by an increased P600 effect. The revised mental representation of the situation is then again utilized to form expectations about the upcoming referent until a full representation of the situation is reached.

If, however, the current state of the mental representation does not support the forming of expectations, the encountered element of the communicative signal is processed without a matching with expectations. Again, the initial step of processing differs depending on whether the encountered element is auditory or not. If it is auditory, in an initial step the first phoneme provides information about the word, which results in an increased PMN response indexing the phonological matching mechanism. The output of the mechanism again informs the semantic retrieval mechanism, so that the N400 indexing the retrieval of the semantics of the word is shorter than if no phonological information was provided as is the case for non-auditory input such as gaze cues. While the monitoring mechanism is initiating the evaluation process — expressed by a positivity (P3b/P600) — if the element leads to a full representation of the situation, no revision process is initiated as the lack of grounds for expectations did lead to no extended mental representation of the situation that requires revision. For a comparison see Figure 6.3 on page 121.

CHAPTER 7

Conclusion

In sum, the experimental evidences presented in this thesis have demonstrated the robust and replicable influence of speech-related gaze cues on several underlying processes, including auditory processing, lexical retrieval, and integration with sentence meaning (and possibly decision making). While most models of language comprehension postulate that predictions are determined by linguistic context, the present results show that information from a variety of sources — in this case speaker gaze — is utilized to enrich communication and to anticipate the course of a sentence, ultimately facilitating language comprehension.

I propose the following time frame of involved mechanisms: Firstly, my findings strongly indicate that, in situated communication, every relevant aspect of the communicative signal available to interlocutors is utilized incrementally to inform a mental representation of the situation. This includes (but is mostly likely not limited to) linguistic content as well as visual content, such as gaze. When the contextual information is sufficient, interlocutors use the current state of the situational model to anticipate upcoming content to facilitate comprehension. These expectations are then rapidly matched with the actual input and, if necessary, revised. This involves an early matching mechanism when presented with speech that can already utilize the first phoneme of an incoming word (PMN). Also, a mechanism of word retrieval is involved that is inhibited when another referent than the expected one is mentioned (N400). Additionally, a mechanism of monitoring of and integration into a mental representation of the situation mediates decision making (P3/P600). The monitoring mechanism is sensitive to the current state of the situational model so that mismatches between the model and input are detected and repaired through the integration of the perceived input. When the model contains enough information to evaluate the content of the sentence, the mechanism enables decision making. This interpretation, however, again underlines the necessity of stimuli relevance. In other words, only if a reaction of the listener is required, this mechanism should be involved. Unlike the notion of experimental task dependency, however, I would argue that the mechanism is involved in any type of required reaction also including reactions such as answering or commenting. While this could also be considered a task, it is still to be investigated whether all or, if not, which required reactions elicit the involvement of the mechanism.

While some of the terms may differ, taken together, my findings are consistent with the retrieval-integration account (Brouwer et al., 2017): Retrieval difficulty (N400) is observed for the Incongruent, Averted and Unexpected conditions, while integration difficulty (P3/P600) is found only when revision of the situation model is necessary in the Incongruent condition. Extending the predictions of that model (N400 and P600), however, my findings demonstrate that, in situated communication, phonological information as well is influenced by contextual expectability, as expressed by the reported PMN modulation. Further, this effect possibly influences the retrieval difficulty indexed by the subsequent N400 component. Also, my findings suggest additional processes of monitoring and evaluation to be indexed by the P600 beyond the integration of the meaning of a referent.

In brief, when combining my findings with previous research in the field, I propose that in situated interactions:

- Any relevant information — including the linguistic signal, visual context, and speaker behaviour — that potentially facilitate comprehension is utilized incrementally

- Expectations about the unfolding utterance are formed as soon as evidence is accessible (linguistic context, gaze to an possible referent, etc.)

- A mental representation of the situation is informed by those expectations

- Expectations are confirmed by subsequently incoming information of any type (linguistic, visual, etc.) and, if necessary, adjusted. This involves different mechanisms:

    1. A matching mechanism of the perceived input (PMN) that is inhibited if the stimulus does not fit the expected form (phonological mismatch) and that potentially alleviates word retrieval (shorter N400 after PMN)
    2. A word retrieval mechanism (N400) that is inhibited if expectations are not met (e.g., a referent different from the expected one is mentioned or gazed at)
    3. A monitoring mechanism enables a rapid reevaluation of the mental representation and integration of the actual referent (P3/P600)

- The same monitoring system is sensitive to the current state of the mental representation of the situation and mediates decision making (depending on the requirement of the situation, such as task, response-readiness, etc.) when enough information is provided

Taken together, these conclusions strongly suggest that speech-related gaze cues are an integral dimension of the communicative signal, immediately contributing to the construction, prediction and confirmation as well as revision of the mental representation of the situation.

Appendices

APPENDIX A

Additional Plots - Experiment 1

The following tables serve as a reminder of the contrasts as used in the linear mixed effects models.

Table A.1: Contrast matrix as embedded in the factor Gaze Direction.

Direction	d1	d2	d3
UP	-1/2	-1/2	0
DOWN	-1/2	1/2	0
LEFT	1/2	0	-1/2
RIGHT	1/2	0	1/2

Table A.2: Contrast matrix as embedded in the factor Longitude.

Longitude	l1	l2
Frontal	-2/3	0
Central	1/3	-1/2
Parietal	1/3	1/2

Table A.3: Contrast matrix as embedded in the factor Condition.

Congruency	c1	c2
Congruent	-1/3	-1/3
Incongruent	2/3	-1/3
Averted	-1/3	2/3

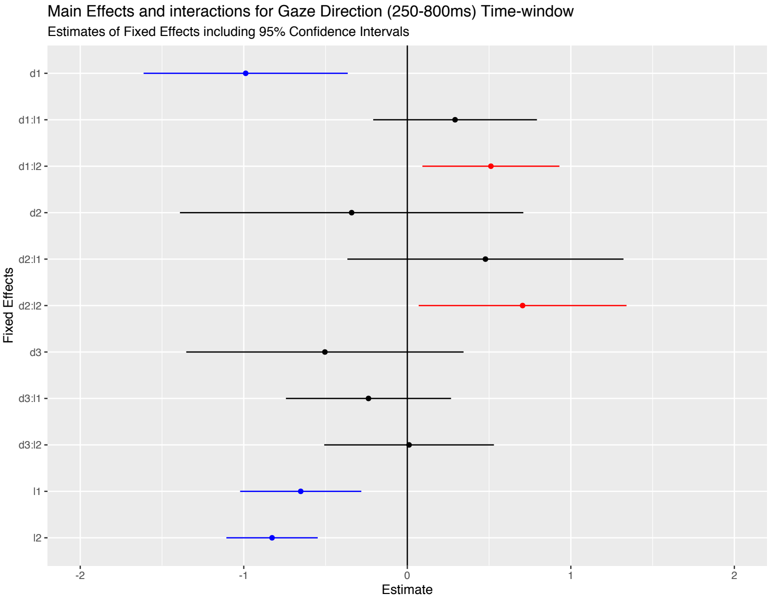


Figure A.1: Fixed Effects for the Positivity in the 250-800ms time-window in Experiment 1. Blue and red bars indicating significant negative and positive effects respectively.

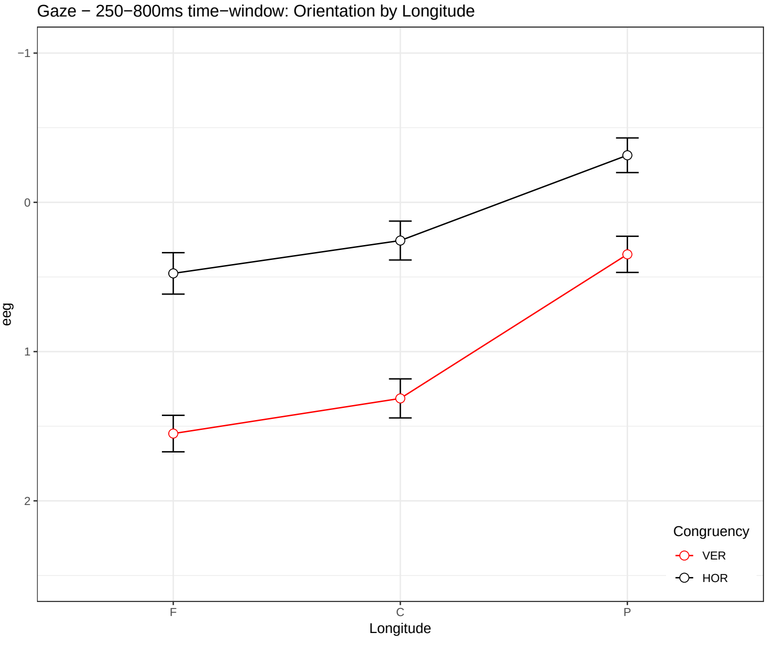


Figure A.2: Plot of means in the 250 – 800ms time-window for Horizontal (black) and Vertical (red) gaze direction in the Gaze region.

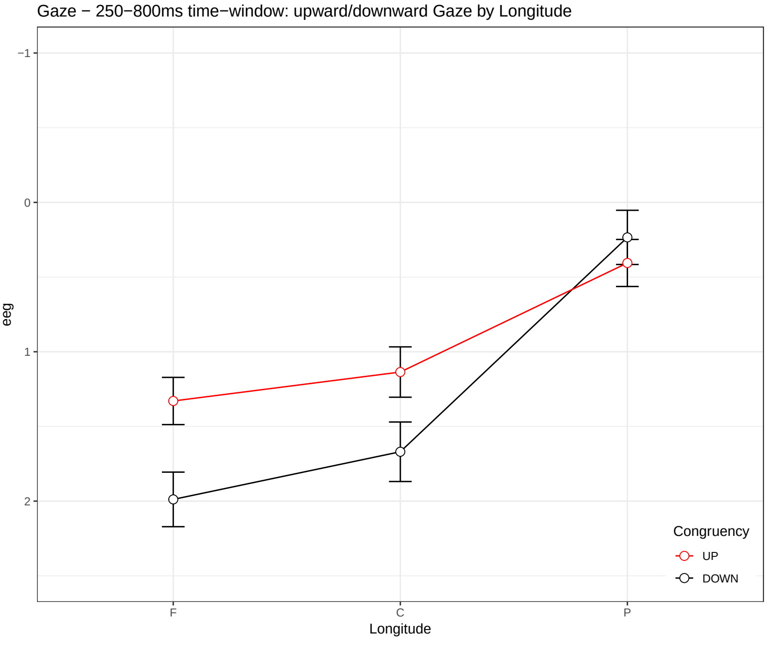


Figure A.3: Plot of means in the 250 – 800ms time-window for Downward (black) and Upward (red) gaze direction in the Gaze region.

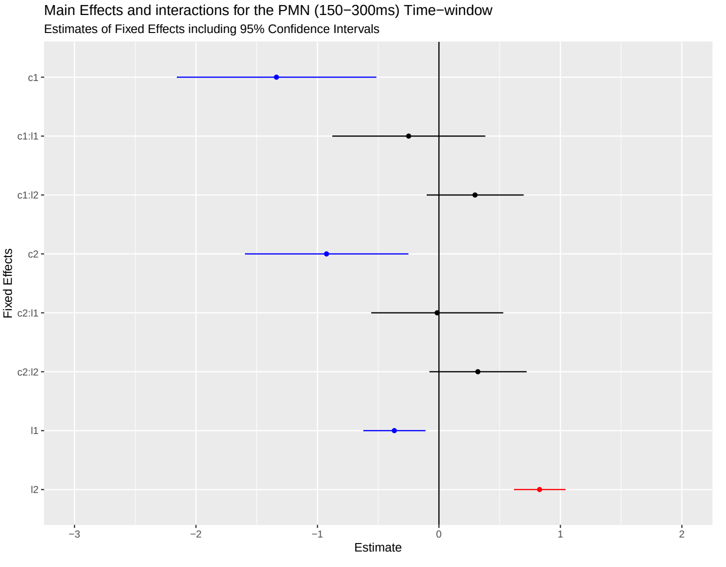


Figure A.4: Fixed Effects for the PMN time-window (150 – 300ms) in the noun region in experiment 1.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 4.8 on page 55 and its description.

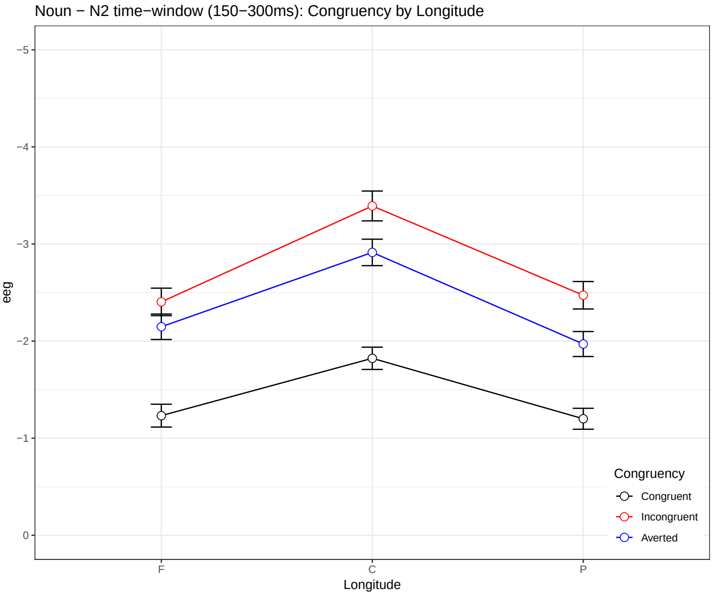
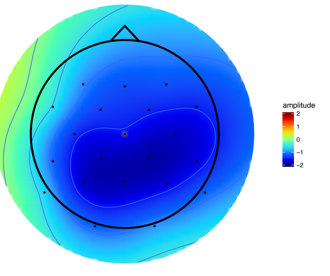
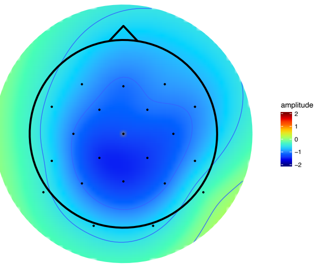


Figure A.5: Plot of means in the PMN time-window (150 – 300ms) for Congruent (black), Incongruent (red) and Averted (blue) conditions in the Noun region. The corresponding analysis can be found in Table 4.8 on page 55 and its description.



(a) Incongruent - Congruent Condition



(b) Averted - Congruent Condition

Figure A.6: Topographical scalp map for the (a) Incongruent - Congruent Condition and (b) Averted - Congruent Condition for the PMN time-window (150 – 300ms) in the noun region. The corresponding analysis can be found in Table 4.8 on page 55 and its description.

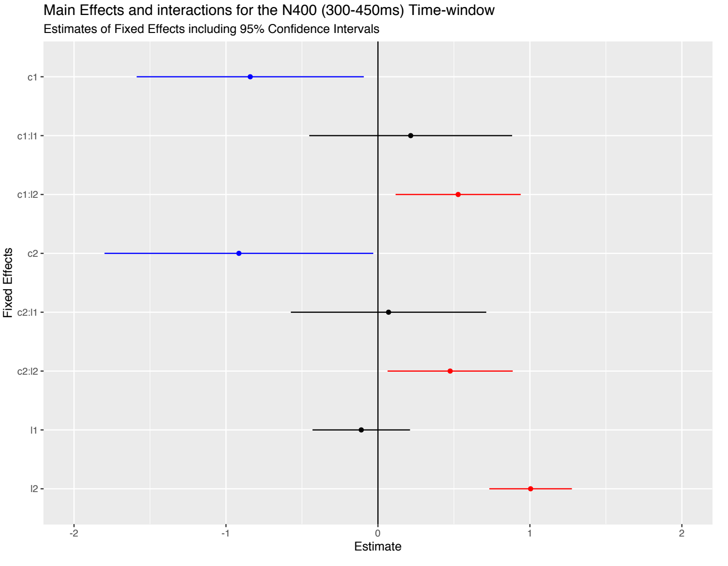


Figure A.7: Fixed Effects for the N400 time-window (300 – 450ms) in the noun region in experiment 1.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 4.9 on page 56 and its description.

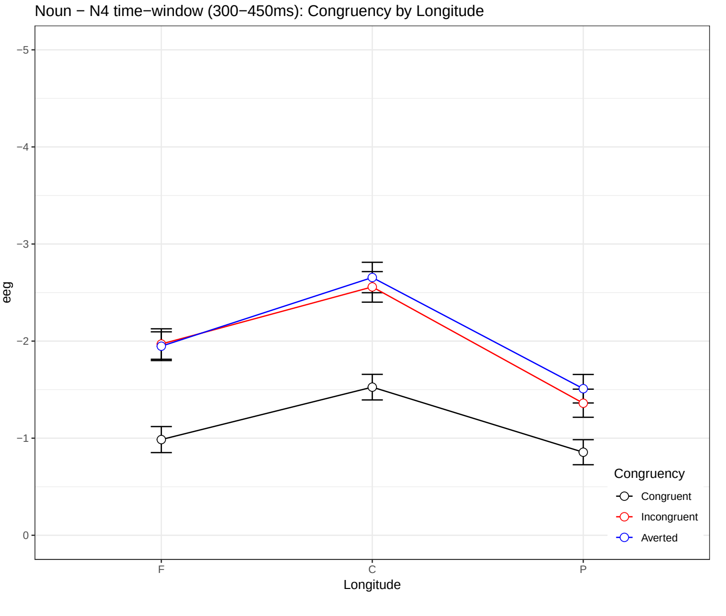
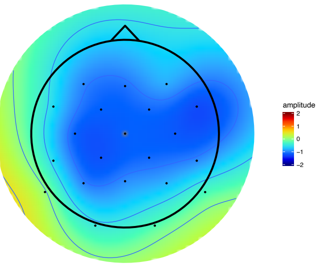
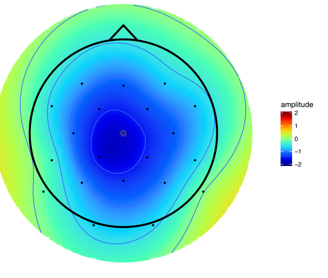


Figure A.8: Plot of means in the N400 time-window (300 – 450ms) for Congruent (black), Incongruent (red) and Averted (blue) conditions in the Noun region. The corresponding analysis can be found in Table 4.9 on page 56 and its description.



(a) Incongruent - Congruent Condition



(b) Averted - Congruent Condition

Figure A.9: Topographical scalp map for the (a) Incongruent - Congruent condition and (b) Averted - Congruent condition (bottom) for the N400 time-window (300 – 450ms) in the noun region. The corresponding analysis can be found in Table 4.9 on page 56 and its description.

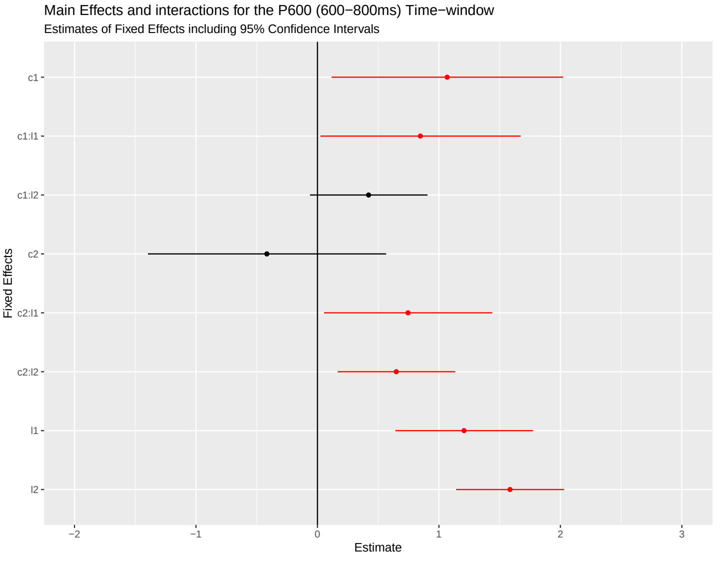


Figure A.10: Fixed Effects for the P600 time-window (600 – 800ms) in the noun region in experiment 1.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 4.11 on page 58 and its description.

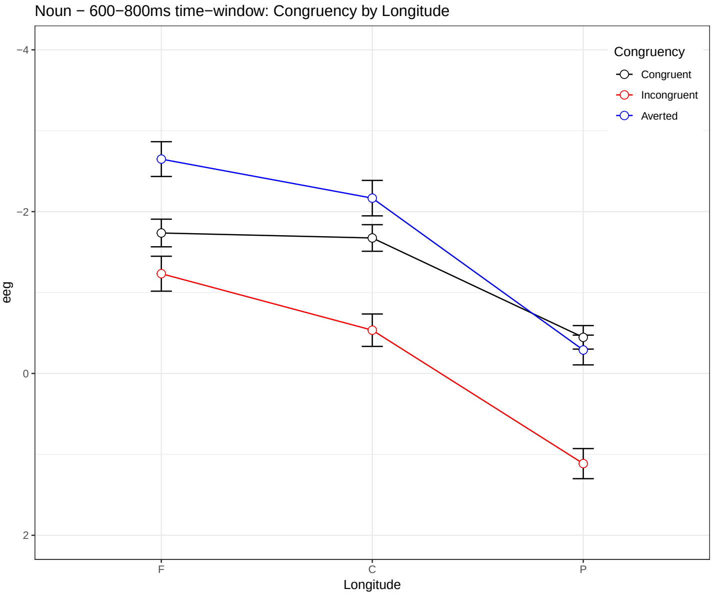
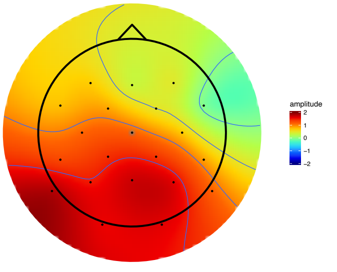
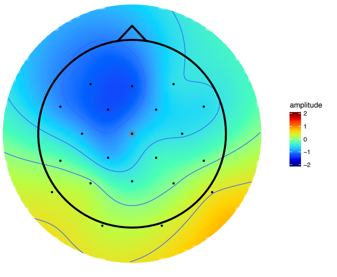


Figure A.11: Plot of means in the P600 time-window (600 – 800ms) for Congruent (black), Incongruent (red) and Averted (blue) conditions in the Noun region. The corresponding analysis can be found in Table 4.11 on page 58 and its description.



(a) Incongruent - Congruent Condition



(b) Averted - Congruent Condition

Figure A.12: Topographical scalp map for the (a) Incongruent - Congruent condition and (b) Averted - Congruent condition (bottom) for the P600 time-window (600 – 800ms) in the noun region. The corresponding analysis can be found in Table 4.11 on page 58 and its description.

APPENDIX B

Additional Plots - Experiment 2

The following tables serve as a reminder of the contrasts as used in the linear mixed effects models.

Table B.1: Contrast matrix as embedded in the factor Longitude.

Longitude	l1	l2
Frontal	-2/3	0
Central	1/3	-1/2
Parietal	1/3	1/2

Table B.2: Contrast matrix as embedded in the factor Condition.

Congruency	c1	c2
Congruent	-1/3	-1/3
Incongruent	2/3	-1/3
Mutual	-1/3	2/3

Table B.3: Contrast matrix as embedded in the factor Condition for the comparison between the Congruent baseline and the Averted filler type.

Congruency	co1
Congruent	-1/2
Averted	1/2

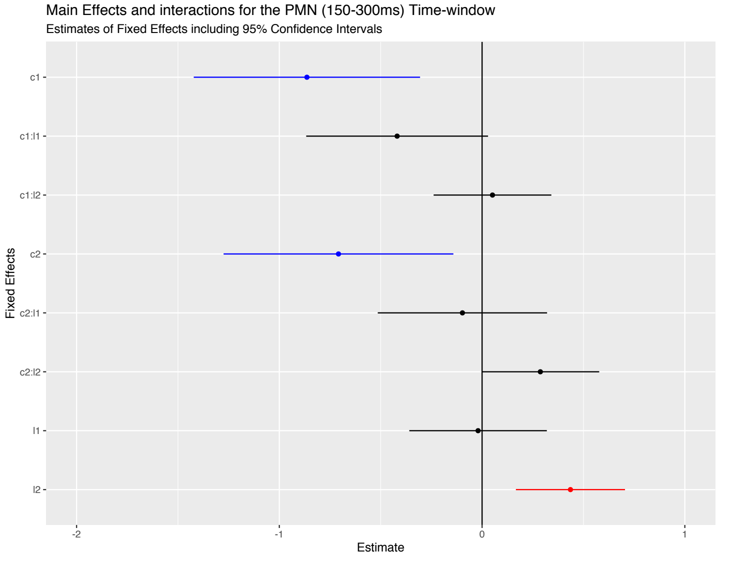


Figure B.1: Fixed Effects for the PMN time-window (150 – 300ms) in the noun region in experiment 2.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 4.20 on page 72 and its description.

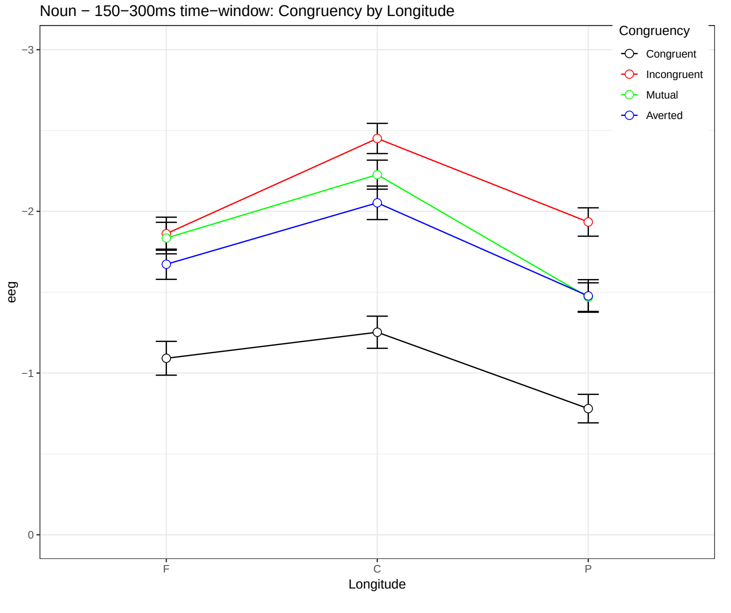
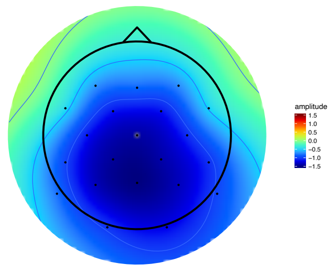
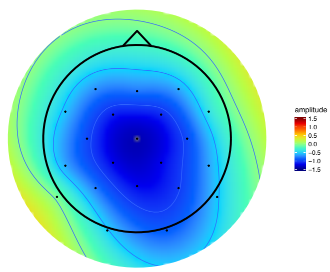


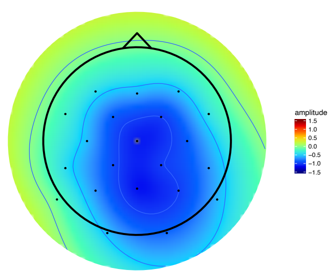
Figure B.2: Plot of means in the PMN time-window (150 – 300ms) for Congruent (black), Incongruent (red) and Mutual (blue) conditions in the Noun region. The corresponding analyses can be found in Table 4.20 on page 72 and Table 4.21 on page 73 and their descriptions.



(a) Incongruent - Congruent Condition



(b) Mutual - Congruent Condition



(c) Averted - Congruent Condition

Figure B.3: Topographical scalp map for the the (a) Incongruent - Congruent condition, (b) Mutual - Congruent condition, and (c) Averted filler type - Congruent Condition for the PMN time-window (150 – 300ms) in the noun region. The corresponding analyses can be found in Table 4.20 on page 72 and Table 4.21 on page 73 and their descriptions.

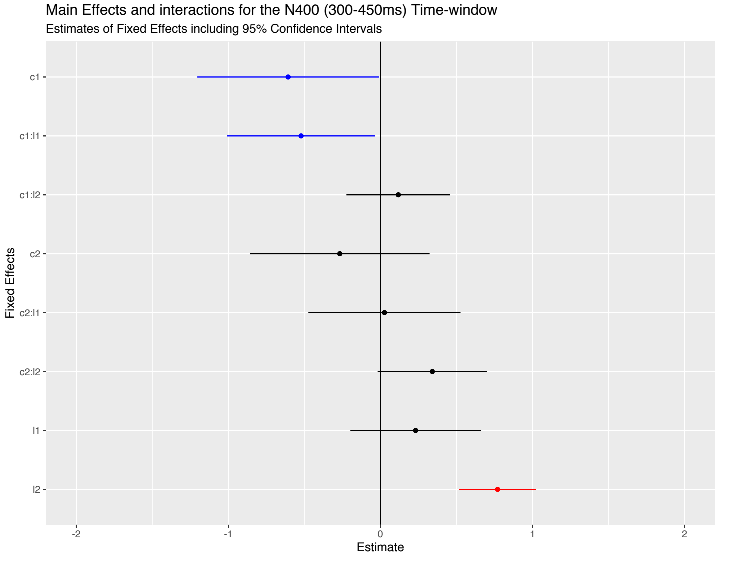


Figure B.4: Fixed Effects for the N400 time-window (300 – 450ms) in the noun region in experiment 1.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 4.22 on page 74 and its description.

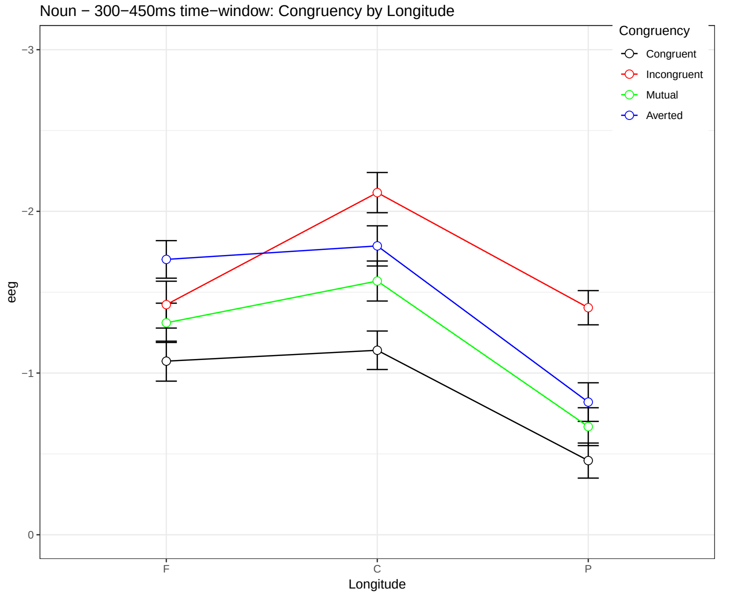
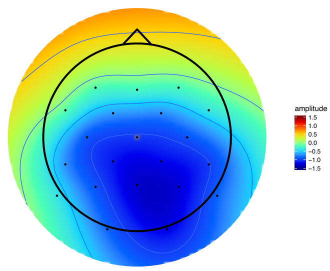
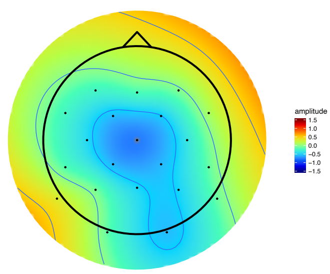


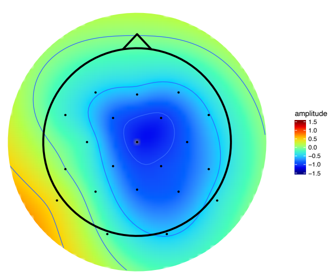
Figure B.5: Plot of means in the N400 time-window (300 – 450ms) for Congruent (black), Incongruent (red) and Averted (blue) conditions in the Noun region. The corresponding analyses can be found in Table 4.22 on page 74 and Table 4.23 on page 75 and their descriptions.



(a) Incongruent - Congruent Condition



(b) Mutual - Congruent Condition



(c) Averted - Congruent Condition

Figure B.6: Topographical scalp map for the the (a) Incongruent - Congruent condition, (b) Mutual - Congruent condition, and (c) Averted filler type - Congruent Condition for the N400 time-window (300 – 450ms) in the noun region. The corresponding analyses can be found in Table 4.22 on page 74 and Table 4.23 on page 75 and their descriptions.

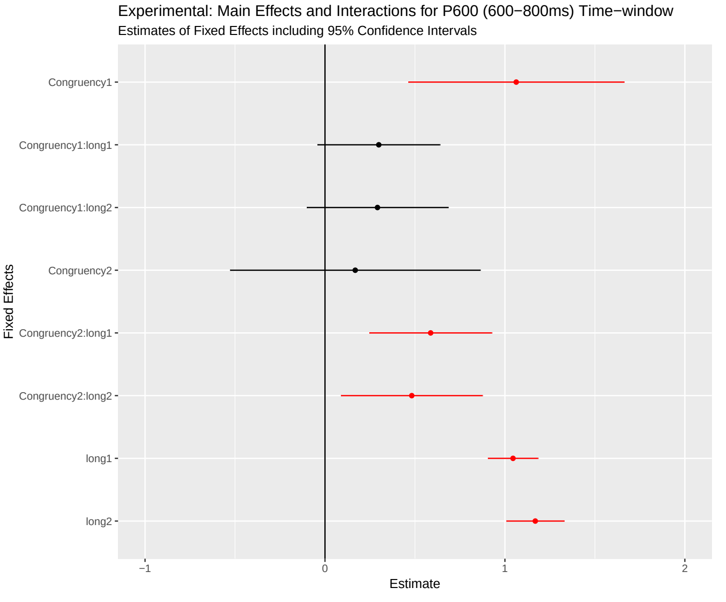


Figure B.7: Fixed Effects for the P600 time-window (600 – 800ms) in the noun region in experiment 2.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 4.24 on page 76 and its description.

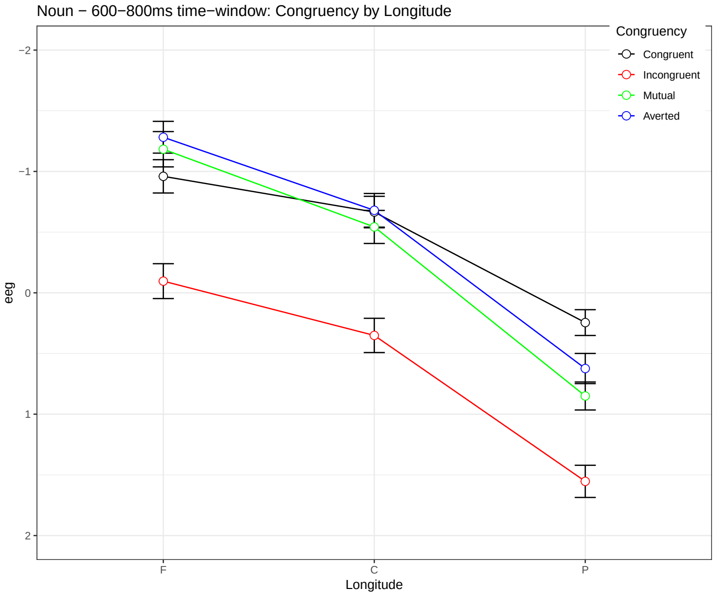
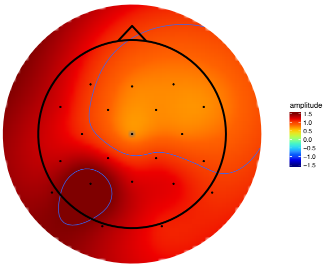
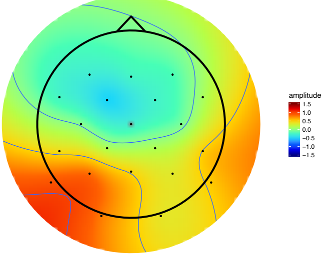


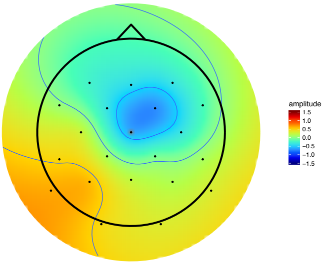
Figure B.8: Plot of means in the P600 time-window (600 – 800ms) for Congruent (black), Incongruent (red) and Averted (blue) conditions in the Noun region. The corresponding analyses can be found in Table 4.24 on page 76 and Table 4.25 on page 77 and their descriptions.



(a) Incongruent - Congruent Condition



(b) Mutual - Congruent Condition



(c) Averted - Congruent Condition

Figure B.9: Topographical scalp map for the the (a) Incongruent - Congruent condition, (b) Mutual - Congruent condition, and (c) Averted filler type - Congruent Condition for the P600 time-window (600 – 800ms) in the noun region. The corresponding analyses can be found in Table 4.24 on page 76 and Table 4.25 on page 77 and their descriptions.

APPENDIX C

Additional Plots - Experiment 3

The following tables serve as a reminder of the contrasts as used in the linear mixed effects models.

Table C.1: Contrast matrix as embedded in the factor Condition.
P - Present gaze , A - Absent gaze , E - Expected continuation , U - Unexpected continuation

Condition	c1	c2	c3
PE	-1/2	-1/2	0
PU	-1/2	1/2	0
AE	1/2	0	-1/2
AU	1/2	0	1/2

Table C.2: Contrast matrix as embedded in the factor Longitude.

Longitude	l1	l2
Frontal	-2/3	0
Central	1/3	-1/2
Parietal	1/3	1/2

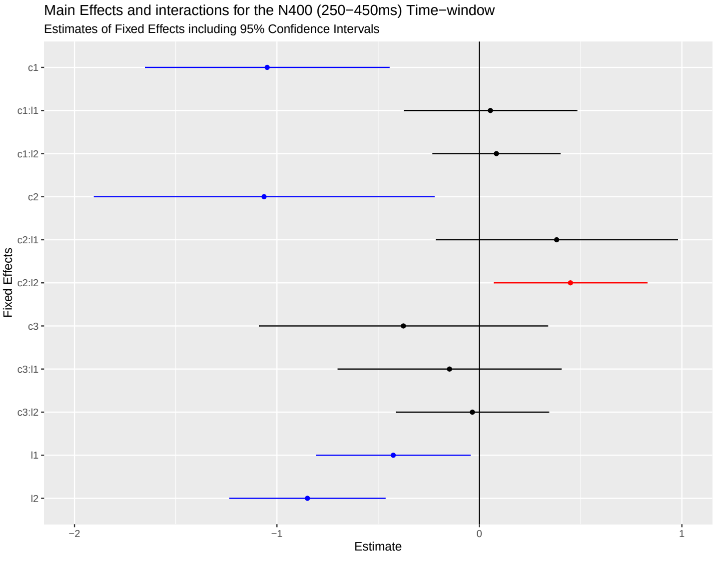


Figure C.1: Fixed Effects for the N400 time-window (250 – 450ms) in the Gaze region in experiment 3.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 5.4 on page 99 and its description.

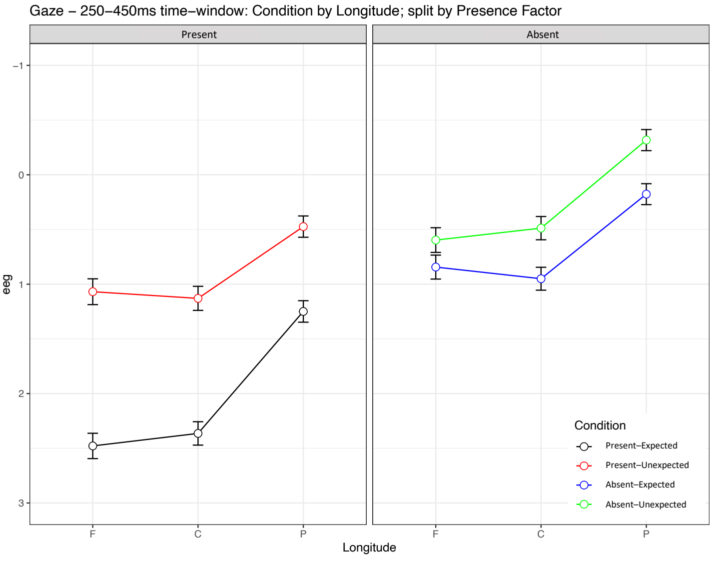
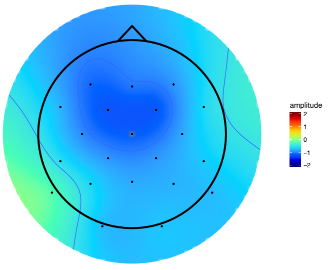
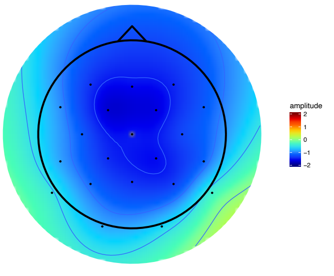


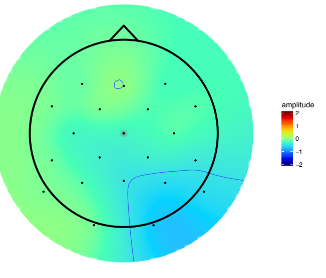
Figure C.2: Plot of means in the N400 time-window (250 – 450ms) for Present-Expected (black), Present-Unexpected (red), Absent-Expected (blue), Absent-Unexpected (green) conditions in the Gaze region. The corresponding analysis can be found in Table 5.4 on page 99 and its description.



(a) Present - Absent condition



(b) Present:
Unexpected - Expected condition



(c) Absent:
Unexpected - Expected condition

Figure C.3: Topographical scalp map for the (a) Present - Absent condition, (b) Present: Unexpected - Expected condition, and (c) Absent: Unexpected - Expected condition for the N400 time-window (250 – 450ms) in the Gaze region. The corresponding analysis can be found in Table 5.4 on page 99 and its description.

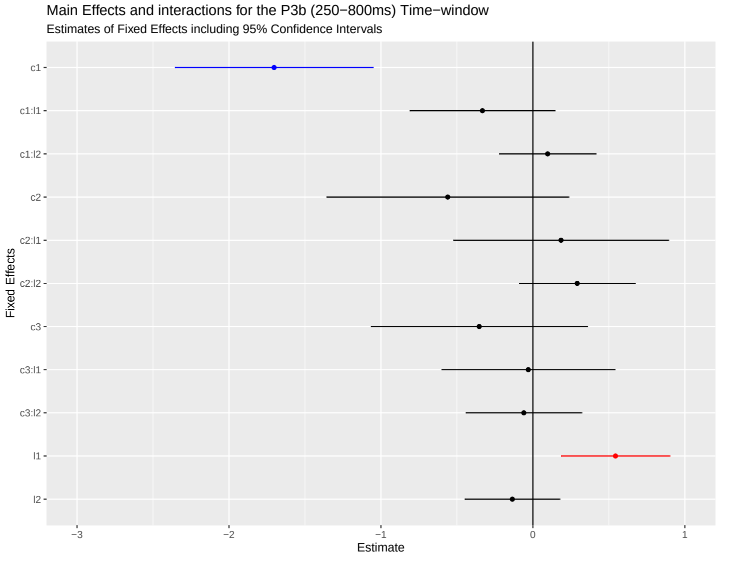


Figure C.4: Fixed Effects for the P300 time-window (250 – 800ms) in the Gaze region in experiment 3.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 5.3 on page 97 and its description.

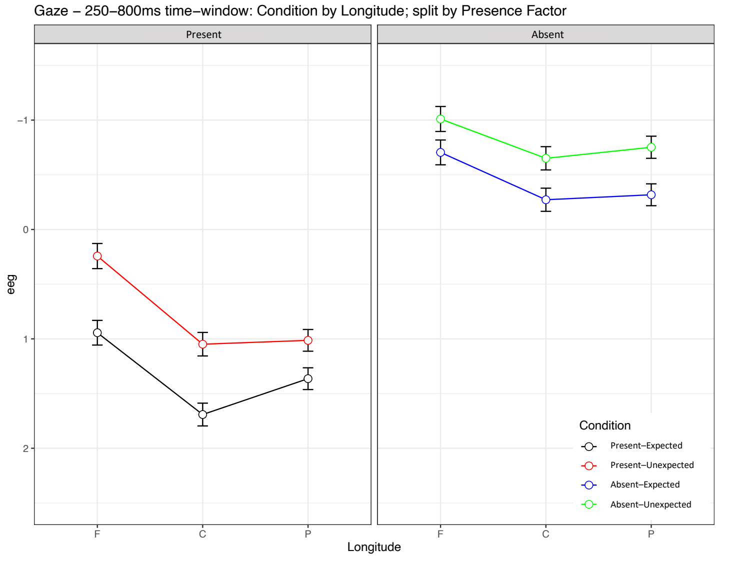
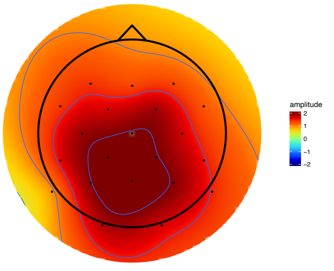
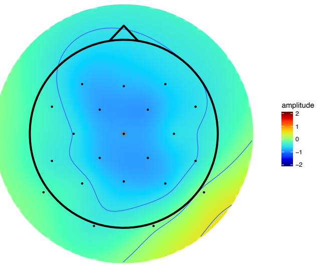


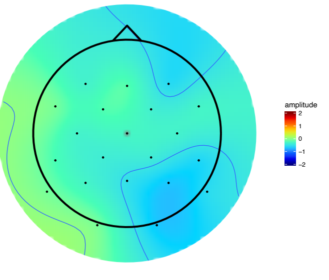
Figure C.5: Plot of means in the P300 time-window (250 – 800ms) for Present-Expected (black), Present-Unexpected (red), Absent-Expected (blue), Absent-Unexpected (green) conditions in the Gaze region. The corresponding analysis can be found in Table 5.3 on page 97 and its description.



(a) Present - Absent condition



(b) Present:
Unexpected - Expected condition



(c) Absent:
Unexpected - Expected condition

Figure C.6: Topographical scalp map for the (a) Present - Absent condition, (b) Present: Unexpected - Expected condition, and (c) Absent: Unexpected - Expected condition for the P300 time-window (250 – 800ms) in the Gaze region. The corresponding analysis can be found in Table 5.3 on page 97 and its description.

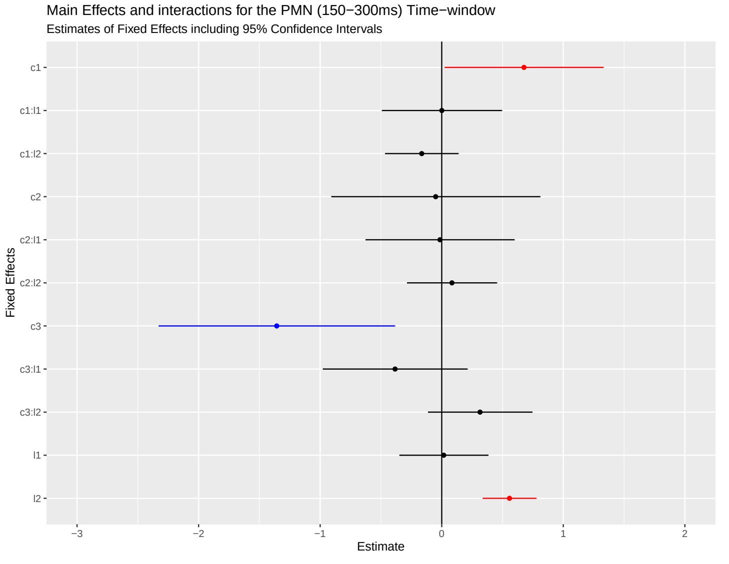


Figure C.7: Fixed Effects for the PMN time-window (150 – 300ms) in the Noun region in experiment 3.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 5.5 on page 103 and its description.

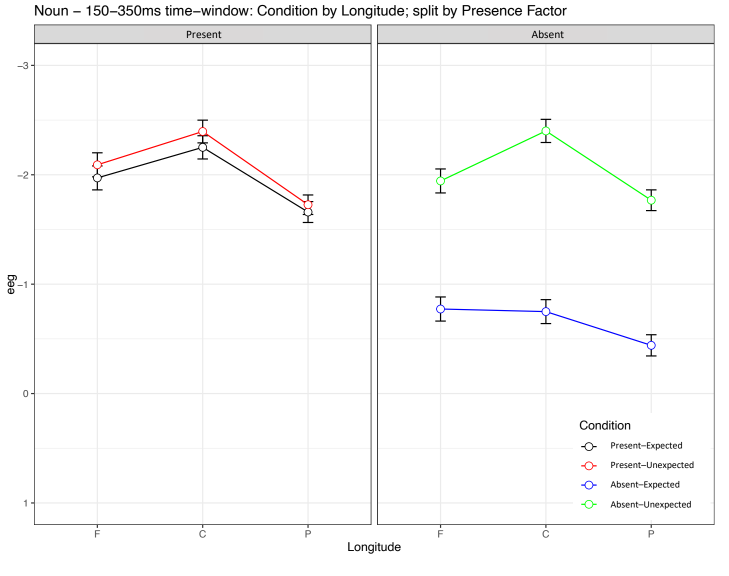
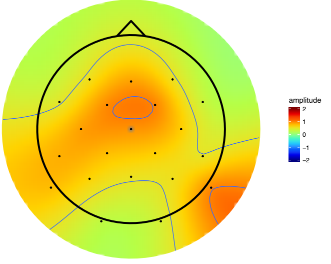
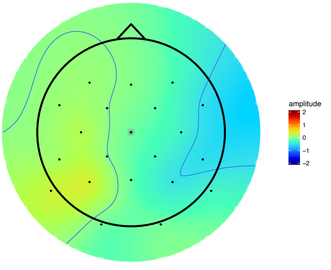


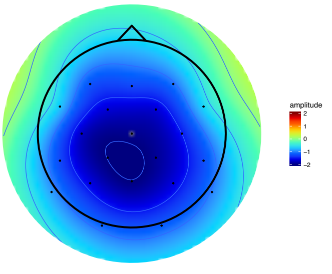
Figure C.8: Plot of means in the PMN time-window (150 – 300ms) for Present-Expected (black), Present-Unexpected (red), Absent-Expected (blue), Absent-Unexpected (green) conditions in the Noun region. The corresponding analysis can be found in Table 5.5 on page 103 and its description.



(a) Present - Absent condition



(b) Present:
Unexpected - Expected condition



(c) Absent:
Unexpected - Expected condition

Figure C.9: Topographical scalp map for the (a) Present - Absent condition, (b) Present: Unexpected - Expected condition, and (c) Absent: Unexpected - Expected condition for the PMN time-window (150 – 300ms) in the Noun region. The corresponding analysis can be found in Table 5.5 on page 103 and its description.

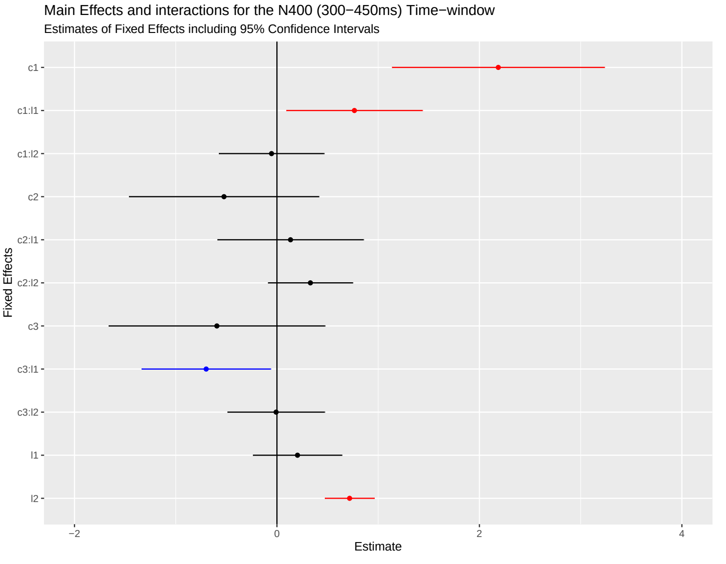


Figure C.10: Fixed Effects for the N400 time-window (300-450ms) in the Noun region in experiment 3.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 5.6 on page 104 and its description.

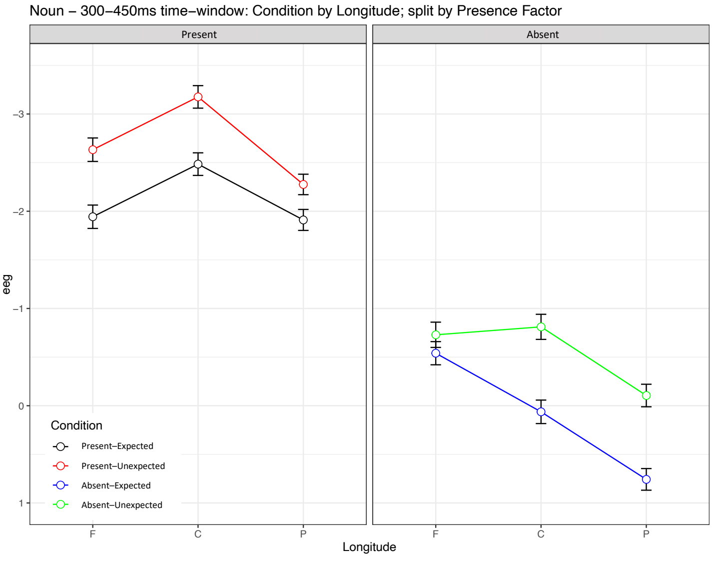
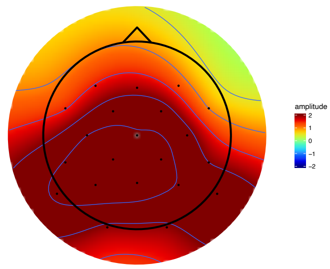
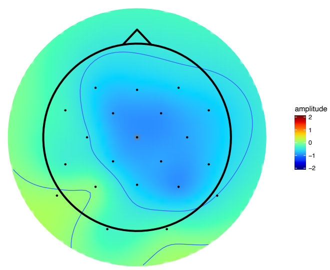


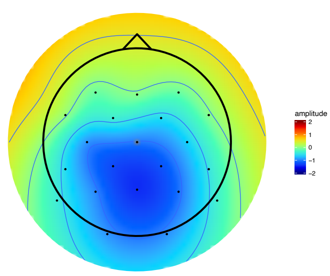
Figure C.11: Plot of means in the N400 time-window (300 – 450ms) for Present-Expected (black), Present-Unexpected (red), Absent-Expected (blue), Absent-Unexpected (green) conditions in the Noun region. The corresponding analysis can be found in Table 5.6 on page 104 and its description.



(a) Present - Absent condition



(b) Present:
Unexpected - Expected condition



(c) Absent:
Unexpected - Expected condition

Figure C.12: Topographical scalp map for the (a) Present - Absent condition, (b) Present: Unexpected - Expected condition, and (c) Absent: Unexpected - Expected condition for the N400 time-window (300 – 450ms) in the Noun region. The corresponding analysis can be found in Table 5.6 on page 104 and its description.

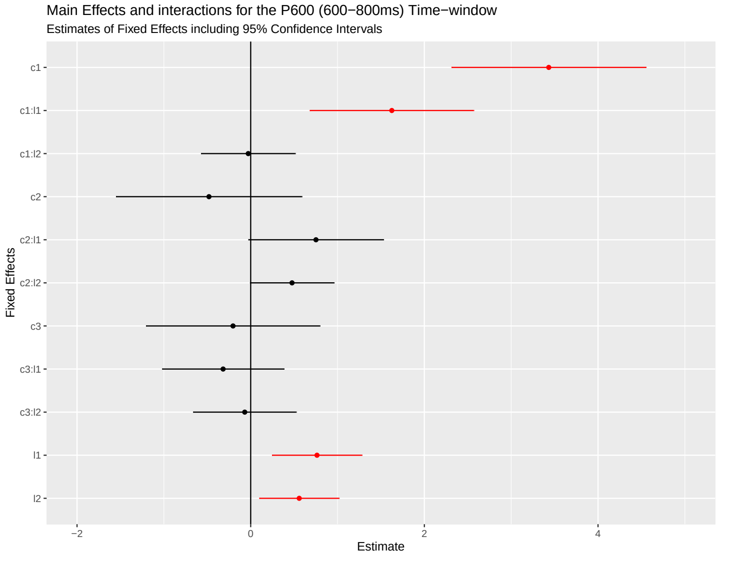


Figure C.13: Fixed Effects for the P600 time-window (600 – 800ms) in the Noun region in experiment 3.
Blue and red bars indicating significant negative and positive effects respectively. The corresponding analysis can be found in Table 5.7 on page 105 and its description.

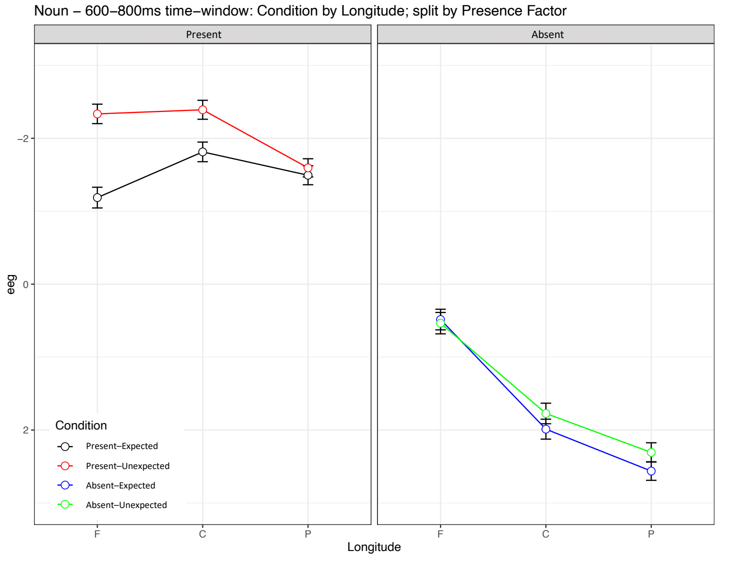
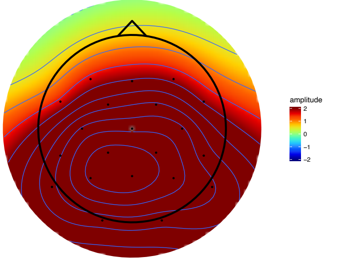
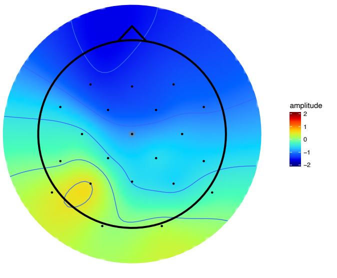


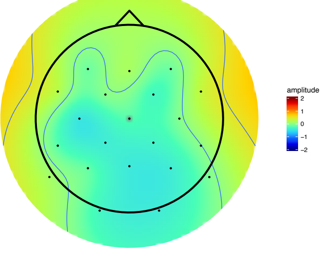
Figure C.14: Plot of means in the P600 time-window (600 – 800ms) for Present-Expected (black), Present-Unexpected (red), Absent-Expected (blue), Absent-Unexpected (green) conditions in the Noun region. The corresponding analysis can be found in Table 5.7 on page 105 and its description.



(a) Present - Absent condition



(b) Present:
Unexpected - Expected condition



(c) Absent:
Unexpected - Expected condition

Figure C.15: Topographical scalp map for the (a) Present - Absent condition, (b) Present: Unexpected - Expected condition, and (c) Absent: Unexpected - Expected condition for the P600 time-window (600 – 800ms) in the Noun region. The corresponding analysis can be found in Table 5.7 on page 105 and its description.

APPENDIX D

Example of Model Reduction Process (PMN in Experiment 1)

In a first step, the maximal model is computed:

```
full.N2 = lmer(eeg ~ 1 + Congruency*long+(1+Congruency*long|subject),
               REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
               control = lmerControl(calc.derivs=FALSE),
               na.action = na.omit)
```

```
## Warning: Model failed to converge with 2 negative eigenvalues: -2.4e-02 -1.6e+01
```

As this model fails to converge, the random structure is reduced.

To do that, after the contrasts were embedded in the respective factors, the following code is used to extract those contrasts in separate variables:

```
mm <- model.matrix( ~ Congruency*long, EXP1.NP2.N2.avg.top.cond)

EXP1.NP2.N2.avg.top.cond$c1 <- mm[,2]
EXP1.NP2.N2.avg.top.cond$c2 <- mm[,3]
EXP1.NP2.N2.avg.top.cond$l1 <- mm[,4]
EXP1.NP2.N2.avg.top.cond$l2 <- mm[,5]
```

These variables are used as the contrasts in the following models, starting with the zero-correlation model:

```
zcp.N2 = lmer(eeg ~ 1 + (c1 +c2)*(l1 + l2)+(1+(c1 +c2)*(l1 + l2)||subject),
              REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
summary(rePCA(zcp.N2))
```

```
## $subject
## Importance of components:
##                          [,1]   [,2]   [,3]    [,4]   [,5]    [,6]    [,7]
## Standard deviation     1.0685 0.8659 0.7072 0.56207 0.5331 0.28440 0.22898
## Proportion of Variance 0.3576 0.2348 0.1566 0.09894 0.0890 0.02533 0.01642
## Cumulative Proportion  0.3576 0.5924 0.7490 0.84794 0.9369 0.96227 0.97869
##                           [,8]    [,9]
## Standard deviation     0.18792 0.18092
## Proportion of Variance 0.01106 0.01025
## Cumulative Proportion  0.98975 1.00000
```

```
VarCorr(zcp.N2)
```

```
##  Groups    Name        Std.Dev.
##  subject   (Intercept) 1.10622
##  subject.1 c1          2.21727
##  subject.2 c2          1.79683
##  subject.3 l1          0.59016
##  subject.4 l2          0.37542
##  subject.5 c1:l1       1.46746
##  subject.6 c1:l2       0.38995
##  subject.7 c2:l1       1.16634
##  subject.8 c2:l2       0.47515
##  Residual              2.07508
```

As the model zcp.N2 supports the full random structure, correlations can be reintroduced (which leads back to the full model):

```
ext.N2 = lmer(eeg ~ 1 + (c1 +c2)*(l1 + l2)+(1+(c1 +c2)*(l1 + l2)|subject),
              REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
summary(rePCA(ext.N2))
```

```
## $subject
## Importance of components:
##                           [,1]   [,2]   [,3]    [,4]   [,5]   [,6]      [,7]
## Standard deviation     1.4297 0.9733 0.7077 0.55234 0.5063 0.2616 0.0007397
## Proportion of Variance 0.4959 0.2298 0.1215 0.07401 0.0622 0.0166 0.0000000
## Cumulative Proportion  0.4959 0.7257 0.8472 0.92121 0.9834 1.0000 1.0000000
##                              [,8]        [,9]
## Standard deviation     0.0001742 0.000003292
## Proportion of Variance 0.0000000 0.000000000
## Cumulative Proportion  1.0000000 1.000000000
```

```
VarCorr(ext.N2)
```

```
##  Groups   Name        Std.Dev. Corr
##  subject  (Intercept) 1.10639
##           c1          2.24483   0.239
##           c2          1.82063   0.334  0.500
##           l1          0.59609   0.300 -0.007  0.127
##           l2          0.43994  -0.570 -0.172  0.095  0.484
##           c1:l1       1.65171  -0.005 -0.049  0.020 -0.256 -0.337
##           c1:l2       1.14168  -0.178 -0.670 -0.291 -0.186 -0.141  0.768
##           c2:l1       1.38137  -0.061 -0.390 -0.218 -0.361 -0.498  0.524  0.695
##           c2:l2       1.15729  -0.276 -0.576 -0.577 -0.331 -0.344  0.341  0.669
##  Residual             2.05488
##
##
##
##
##
##
##
##
##
##   0.909
##
```

The correlations lead to a degenerate model (components with little to no contribution). This is possibly caused by the interactions of the factors in the random structure. However, as the zero-correlation model identified all components, in a first step only the correlations of the interactions are removed:

```r
red.N2 = lmer(eeg ~ 1 + (c1 +c2)*(l1 + l2) +
                (1 + c1 +c2 + l1 + l2 | subject) +
                (0 + c1:l1 | subject) +
                (0 + c1:l2 | subject) +
                (0 + c2:l1 | subject) +
                (0 + c2:l2 | subject),
              REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
summary(rePCA(red.N2))
```

```
## $subject
## Importance of components:
##                          [,1]   [,2]   [,3]    [,4]    [,5]    [,6]   [,7]
## Standard deviation     1.2322 0.7079 0.6873 0.56336 0.51216 0.30501 0.2310
## Proportion of Variance 0.4666 0.1540 0.1452 0.09754 0.08061 0.02859 0.0164
## Cumulative Proportion  0.4666 0.6206 0.7658 0.86335 0.94396 0.97255 0.9889
##                           [,8] [,9]
## Standard deviation     0.18963    0
## Proportion of Variance 0.01105    0
## Cumulative Proportion  1.00000    1
```

```r
VarCorr(red.N2)
```

```
##  Groups    Name        Std.Dev. Corr
##  subject   (Intercept) 1.10616
##            c1          2.24188   0.241
##            c2          1.81791   0.335  0.500
##            l1          0.59510   0.302 -0.005  0.127
##            l2          0.41759  -0.601 -0.183  0.099  0.516
##  subject.1 c1:l1       1.46761
##  subject.2 c1:l2       0.39312
##  subject.3 c2:l1       1.16790
##  subject.4 c2:l2       0.47885
##  Residual              2.07310
```

The PCA still shows one unidentified component. So the interactions with the least variance explained are removed:

```
red2.N2 = lmer(eeg ~ 1 + (c1 +c2)*(l1 + l2) +
                (1 + c1 +c2 + l1 + l2 | subject) +
                (0 + c1:l1 | subject) +
                (0 + c2:l1 | subject),
              REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
summary(rePCA(red2.N2))
```

```
## $subject
## Importance of components:
##                          [,1]   [,2]   [,3]    [,4]    [,5]    [,6]      [,7]
## Standard deviation     1.2297 0.7066 0.6864 0.56110 0.51108 0.30421 0.0001269
## Proportion of Variance 0.4798 0.1585 0.1495 0.09991 0.08289 0.02937 0.0000000
## Cumulative Proportion  0.4798 0.6383 0.7878 0.88774 0.97063 1.00000 1.0000000
```

```
VarCorr(red2.N2)
```

```
##  Groups    Name        Std.Dev. Corr
##  subject   (Intercept) 1.10499
##            c1          2.24121   0.238
##            c2          1.81832   0.334  0.500
##            l1          0.59445   0.301 -0.006  0.129
##            l2          0.41772  -0.600 -0.181  0.100  0.516
##  subject.1 c1:l1       1.46741
##  subject.2 c2:l1       1.16520
##  Residual              2.07665
```

This did not solve the issue. The l2 component also shows low variance but high correlations with both l1 and the intercept, so those correlations are removed:

```
red3.N2 = lmer(eeg ~ 1 + (c1 +c2)*(l1 + l2) +
                (1 + c1 +c2 + l1 | subject) +
                (0 + l2 | subject) +
                (0 + c1:l1 | subject) +
                (0 + c2:l1 | subject),
              REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
              control = lmerControl(calc.derivs=FALSE),
```

```
               na.action = na.omit)
summary(rePCA(red3.N2))
```

```
## $subject
## Importance of components:
##                          [,1]   [,2]   [,3]   [,4]    [,5]    [,6]    [,7]
## Standard deviation     1.2299 0.7054 0.6851 0.5610 0.49448 0.26331 0.18007
## Proportion of Variance 0.4817 0.1584 0.1494 0.1002 0.07786 0.02208 0.01032
## Cumulative Proportion  0.4817 0.6401 0.7895 0.8897 0.96760 0.98968 1.00000
```

```
VarCorr(red3.N2)
```

```
##  Groups    Name        Std.Dev. Corr
##  subject   (Intercept) 1.10579
##            c1          2.24505   0.239
##            c2          1.81854   0.334  0.500
##            l1          0.58964   0.306 -0.006  0.128
##  subject.1 l2          0.37425
##  subject.2 c1:l1       1.46604
##  subject.3 c2:l1       1.16593
##  Residual              2.07839
```

While this model is identified, it is reasonable to remove the non-correlated random effects for a more parsimonious approach.

```
red4.N2 = lmer(eeg ~ 1 + (c1 +c2)*(l1 + l2)+(1 + c1 +c2 + l1 |subject),
               REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
               control = lmerControl(calc.derivs=FALSE),
               na.action = na.omit)
summary(rePCA(red4.N2))
```

```
## $subject
## Importance of components:
##                          [,1]   [,2]   [,3]    [,4]
## Standard deviation     1.2039 0.6704 0.4841 0.25540
## Proportion of Variance 0.6593 0.2044 0.1066 0.02967
## Cumulative Proportion  0.6593 0.8637 0.9703 1.00000
```

```
VarCorr(red4.N2)
```

```
## Groups   Name        Std.Dev. Corr
## subject  (Intercept) 1.10558
##          c1          2.24225   0.240
##          c2          1.81704   0.335  0.500
##          l1          0.58437   0.308 -0.006  0.129
## Residual             2.12140
```

```
anova(red4.N2,red3.N2)
```

```
## Data: EXP1.NP2.N2.avg.top.cond
## Models:
## red4.N2: eeg ~ 1 + (c1 + c2) * (l1 + l2) + (1 + c1 + c2 + l1 | subject)
## red3.N2: eeg ~ 1 + (c1 + c2) * (l1 + l2) + (1 + c1 + c2 + l1 | subject) +
## red3.N2:     (0 + l2 | subject) + (0 + c1:l1 | subject) + (0 + c2:l1 |
## red3.N2:     subject)
##         Df   AIC   BIC  logLik deviance  Chisq Chi Df         Pr(>Chisq)
## red4.N2 20 16768 16893 -8364.0    16728
## red3.N2 23 16710 16854 -8332.3    16664 63.565      3 0.0000000000001017 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This however would significantly reduce model fit. A last change could be to reintroduce the previously discarded interactions:

```
red5.N2 = lmer(eeg ~ 1 + (c1 +c2)*(l1 + l2) +
                (1 + c1 +c2 + l1 | subject) +
                (0 + l2 | subject) +
                (0 + c1:l1 | subject) +
                (0 + c1:l2 | subject) +
                (0 + c2:l1 | subject) +
                (0 + c2:l2 | subject),
              REML=FALSE, data=EXP1.NP2.N2.avg.top.cond,
              control = lmerControl(calc.derivs=FALSE),
              na.action = na.omit)
summary(rePCA(red5.N2))
```

```
## $subject
```

```
## Importance of components:
##                           [,1]   [,2]   [,3]    [,4]    [,5]    [,6]    [,7]
## Standard deviation     1.2320 0.7074 0.6863 0.56199 0.49539 0.26412 0.22885
## Proportion of Variance 0.4684 0.1544 0.1454 0.09746 0.07573 0.02153 0.01616
## Cumulative Proportion  0.4684 0.6228 0.7681 0.86559 0.94132 0.96285 0.97901
##                           [,8]    [,9]
## Standard deviation     0.1879 0.18086
## Proportion of Variance 0.0109 0.01009
## Cumulative Proportion  0.9899 1.00000
```

```
VarCorr(red5.N2)
```

```
##  Groups    Name        Std.Dev. Corr
##  subject   (Intercept) 1.10618
##            c1          2.24421   0.239
##            c2          1.81947   0.334  0.500
##            l1          0.59033   0.305 -0.006  0.128
##  subject.1 l2          0.37527
##  subject.2 c1:l1       1.46771
##  subject.3 c1:l2       0.38990
##  subject.4 c2:l1       1.16607
##  subject.5 c2:l2       0.47483
##  Residual              2.07488
```

```
anova(red3.N2,red5.N2)
```

```
## Data: EXP1.NP2.N2.avg.top.cond
## Models:
## red3.N2: eeg ~ 1 + (c1 + c2) * (l1 + l2) + (1 + c1 + c2 + l1 | subject) +
## red3.N2:     (0 + l2 | subject) + (0 + c1:l1 | subject) + (0 + c2:l1 |
## red3.N2:     subject)
## red5.N2: eeg ~ 1 + (c1 + c2) * (l1 + l2) + (1 + c1 + c2 + l1 | subject) +
## red5.N2:     (0 + l2 | subject) + (0 + c1:l1 | subject) + (0 + c1:l2 |
## red5.N2:     subject) + (0 + c2:l1 | subject) + (0 + c2:l2 | subject)
##         Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## red3.N2 23 16710 16854 -8332.3    16664
## red5.N2 25 16713 16869 -8331.7    16663 1.1333      2     0.5674
```

While this model is also identified, it does not significantly improve model fit. Hence, due to the parsimonious approach, `red3.N2` is picked as the final model, as it contains a simpler random structure while explaining the data equally well.

```
ext2.N2 <- red3.N2

summary(ext2.N2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: eeg ~ 1 + (c1 + c2) * (l1 + l2) + (1 + c1 + c2 + l1 | subject) +
##     (0 + l2 | subject) + (0 + c1:l1 | subject) + (0 + c2:l1 |      subject)
##    Data: EXP1.NP2.N2.avg.top.cond
## Control: lmerControl(calc.derivs = FALSE)
##
##      AIC      BIC   logLik deviance df.resid
##  16710.5  16854.0  -8332.3  16664.5     3757
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3581 -0.6328  0.0220  0.6499  3.4557
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  subject   (Intercept) 1.2228   1.1058
##            c1          5.0402   2.2450    0.24
##            c2          3.3071   1.8185    0.33  0.50
##            l1          0.3477   0.5896    0.31 -0.01  0.13
##  subject.1 l2          0.1401   0.3743
##  subject.2 c1:l1       2.1493   1.4660
##  subject.3 c2:l1       1.3594   1.1659
##  Residual              4.3197   2.0784
## Number of obs: 3780, groups:  subject, 30
##
## Fixed effects:
##               Estimate Std. Error        df t value        Pr(>|t|)
## (Intercept)   -2.17261    0.20470  30.02270 -10.614 0.0000000000111 ***
## c1            -1.33759    0.41817  29.99047  -3.199         0.00325 **
## c2            -0.92529    0.34219  30.03928  -2.704         0.01117 *
```

```
## l1            -0.36725    0.12935   30.01179  -2.839         0.00804 **
## l2             0.82867    0.10736   30.00541   7.719 0.0000000130280 ***
## c1:l1         -0.24927    0.32015   31.57719  -0.779         0.44202
## c1:l2          0.29702    0.20283 3563.03538   1.464         0.14318
## c2:l1         -0.01486    0.27599   31.52231  -0.054         0.95740
## c2:l2          0.32050    0.20283 3563.03538   1.580         0.11416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr) c1     c2     l1     l2     c1:l1  c1:l2  c2:l1
## c1     0.231
## c2     0.320  0.500
## l1     0.251 -0.005  0.103
## l2     0.000  0.000  0.000  0.000
## c1:l1  0.000  0.000  0.000  0.000  0.000
## c1:l2  0.000  0.000  0.000  0.000  0.000  0.000
## c2:l1  0.000  0.000  0.000  0.000  0.000  0.175  0.000
## c2:l2  0.000  0.000  0.000  0.000  0.000  0.000  0.500  0.000
```

Bibliography

Ackerman, P. T., Dykman, R. A., and Oglesby, D. M. (1994). Visual event-related potentials of dyslexic children to rhyming and nonrhyming stimuli. *Journal of Clinical and Experimental Neuropsychology*, 16(1):138–154.

Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Umeda, S., and Kashima, H. (2006). Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. *Neuropsychologia*, 44(10):1804–1810.

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, 38(4):419–439.

Altmann, G. T. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

Argyle, M. and Cook, M. (1976). *Gaze and mutual gaze.* Cambridge U Press.

Baron-Cohen, S. (1997). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *The maladapted mind: Classic readings in evolutionary psychopathology*, pages 207–239.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Barrett, S. and Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. *Neuropsychologia*, 27(7):913–922.

Barrett, S. E. and Rugg, M. D. (1990a). Event-related potentials and the phonological matching of picture names. *Brain and Language*, 38(3):424–437.

Barrett, S. E. and Rugg, M. D. (1990b). Event-related potentials and the semantic matching of pictures. *Brain and cognition*, 14(2):201–212.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bayliss, A. P., Frischen, A., Fenske, M. J., and Tipper, S. P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, 104(3):644–653.

Bayliss, A. P., Paul, M. A., Cannon, P. R., and Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic bulletin & review*, 13(6):1061–1066.

Becchio, C., Bertone, C., and Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, 12(7):254–258.

beim Graben, P., Gerth, S., and Vasishth, S. (2008). Towards dynamical system models of language-related brain potentials. *Cognitive neurodynamics*, 2(3):229–255.

Bentin, S., Kutas, M., and Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, 30(2):161–169.

Bentin, S., McCarthy, G., and Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and clinical Neurophysiology*, 60(4):343–355.

Bobes, M. A., Valdessosa, M., and Olivares, E. (1994). An erp study of expectancy violation in face perception. *Brain and cognition*, 26(1):1–22.

Böckler, A., Knoblich, G., and Sebanz, N. (2011). Observing shared attention modulates gaze following. *Cognition*, 120:292–298.

Bornkessel, I. and Schlesewsky, M. (2006a). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, 113(4):787.

Bornkessel, I. and Schlesewsky, M. (2006b). Generalised semantic roles and syntactic templates: A new framework for language comprehension. *TRENDS IN LINGUISTICS STUDIES AND MONOGRAPHS*, 165:327.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., and Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and language*, 117(3):133–152.

Brink, D. v. d. and Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by erps. *Journal of Cognitive Neuroscience*, 16(6):1068–1084.

Bronzino, J. D. (1995). Principles of electroencephalography. *The biomedical engineering handbook*, 1.

Brouwer, H. (2014). *The electrophysiology of language comprehension: A neurocomputational model*. PhD thesis, University of Groningen.

Brouwer, H. and Crocker, M. W. (2017). On the proper treatment of the n400 and p600 in language comprehension. *Frontiers in psychology*, 8.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., and Hoeks, J. C. (2017). A neurocomputational model of the n400 and the p600 in language processing. *Cognitive science*, 41(S6):1318–1352.

Brouwer, H., Fitz, H., and Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the p600 in language comprehension. *Brain research*, 1446:127–143.

Brown, C. M. and Hagoort, P. (1999). The cognitive neuroscience of language: challenges and future directions. *The neurocognition of language*, pages 3–14.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2):159–168.

Burkhardt, P. (2007). The p600 reflects cost of new information in discourse memory. *Neuroreport*, 18(17):1851–1854.

Butterworth, G. and Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British journal of developmental psychology*, 9(1):55–72.

Caldara, R., Jermann, F., Arango, G. L., and Van der Linden, M. (2004). Is the n400 category-specific? a face and language processing study. *Neuroreport*, 15(17):2589–2593.

Chao, L., Nielsen-Bohlman, L., and Knight, R. (1995). Auditory event-related potentials dissociate early and late memory processes. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 96(2):157–168.

Chapman, R. M. and Bragdon, H. R. (1964). Evoked responses to numerical and non-numerical visual stimuli while problem solving. *Nature*, 203(4950):1155.

Clark, H. H. and Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of memory and language*, 50(1):62–81.

Comerchero, M. D. and Polich, J. (1999). P3a and p3b from typical auditory and visual stimuli. *Clinical neurophysiology*, 110(1):24–30.

Connolly, J. F., D'arcy, R. C., Kujala, A., Alho, K., et al. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *NeuroReport*, 12(2):237–243.

Connolly, J. F. and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3):256–266.

Connolly, J. F., Phillips, N. A., Stewart, S. H., and Brake, W. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and language*, 43(1):1–18.

Connolly, J. F., Stewart, S., and Phillips, N. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and language*, 39(2):302–318.

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

Da Silva, F. L. (2009). Eeg: origin and measurement. In *EEg-fMRI*, pages 19–38. Springer.

Dambacher, M., Kliegl, R., Hofmann, M., and Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain research*, 1084(1):89–103.

D'Arcy, R. C., Connolly, J. F., Service, E., Hawco, C. S., and Houlihan, M. E. (2004). Separating phonological and semantic processing in auditory sentence processing: A high-resolution event-related brain potential study. *Human brain mapping*, 22(1):40–51.

Davis, P. (1939). The electrical response of the human brain to auditory stimuli. *American Journal of Physiology*, 126:475–476.

Delogu, F., Brouwer, H., and Crocker, M. W. (2019a). Event-related potentials index lexical retrieval (n400) and integration (p600) during language comprehension. *Brain and cognition*, 135:103569.

Delogu, F., Drenhaus, H., and Crocker, M. W. (2018). On the predictability of event boundaries in discourse: An erp investigation. *Memory & cognition*, 46(2):315–325.

Delogu, F., Jachmann, T., Staudte, M., Vespignani, F., and Molinaro, N. (2019b). Discourse expectations are sensitive to the question under discussion: Evidence from erps. *Discourse Processes*, pages 1–19.

Delong, K. A., Urbach, T. P., Groppe, D. M., and Kutas, M. (2011). Overlapping dual erp responses to low cloze probability sentence continuations. *Psychophysiology*, 48(9):1203–1207.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Desroches, A. S., Newman, R. L., and Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, 21(10):1893–1906.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., and Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental psychology*, 38(3):438.

Donchin, E. (1981). Surprise! surprise? *Psychophysiology*, 18(5):493–513.

Drenhaus, H., Zimmermann, M., and Vasishth, S. (2011). Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics*, 24(3):320–337.

Driver IV, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., and Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual cognition*, 6(5):509–540.

Dugdale, D. (1993). *Essentials of electromagnetism*. Macmillan International Higher Education.

Duncan-Johnson, C. C. and Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, 14(5):456–467.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6):581–604.

Falkenstein, M., Hoormann, J., Christ, S., and Hohnsbein, J. (2000). Erp components on reaction errors and their functional significance: a tutorial. *Biological psychology*, 51(2-3):87–107.

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

Federmeier, K. D. and Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation*, 51:1–44.

Ferreira, F., Foucart, A., and Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, 69(3):165–182.

Flom, R. E., Lee, K. E., and Muir, D. E. (2007). *Gaze-following: Its development and significance.* Lawrence Erlbaum Associates Publishers.

Folk, C. L. and Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human perception and performance*, 24(3):847.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, 6(2):78–84.

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive brain research*, 1(3):183–192.

Friederici, A. D., Steinhauer, K., and Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition*, 27(3):438–453.

Friesen, C. K. and Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3):490–495.

Friesen, C. K., Ristic, J., and Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2):319.

Ganis, G. and Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, 16(2):123–144.

Ganis, G., Kutas, M., and Sereno, M. I. (1996). The search for common sense: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8(2):89–106.

Ganis, G. and Schendan, H. E. (2013). Concealed semantic and episodic autobiographical memory electrified. *Frontiers in human neuroscience*, 6:354.

George, M. S., Mannes, S., and Hoffinan, J. E. (1994). Global semantic expectancy and language comprehension. *Journal of cognitive neuroscience*, 6(1):70–83.

Geyer, A., Holcomb, P., Kuperberg, G., and Perlmutter, N. (2006). Plausibility and sentence comprehension. an erp study. *J. Cogn. Neurosci. Suppl.*

Glenberg, A. M., Schroeder, J. L., and Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory and cognition*, 26(4):651–658.

Gold, J. I. and Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in cognitive sciences*, 5(1):10–16.

Goodwin, C. (1981). Conversational organization: Interaction between speakers and hearers.

Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the p600. *Language and cognitive processes*, 25(2):149–188.

Griffin, Z. M. (2004). *Why look? Reasons for eye movements related to language production*, pages 213–247. Psychology Press.

Griffin, Z. M. and Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 4(11):274–279.

Grillon, C., Courchesne, E., Ameli, R., Elmasian, R., and Braff, D. (1990). Effects of rare non-target stimuli on brain electrophysiological activity and performance. *International Journal of Psychophysiology*, 9(3):257–267.

Hagoort, P. (2007). The memory, unification, and control (muc) model of language. *Automaticity and control in language processing*, 1:243.

Hagoort, P., Brown, C., and Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and cognitive processes*, 8(4):439–483.

Hagoort, P. and Brown, C. M. (2000). Erp effects of listening to speech: Semantic erp effects. *Neuropsychologia*, 38(11):1518–1530.

Hagoort, P., Brown, C. M., and Osterhout, L. (1999). The neurocognition of syntactic processing. *The neurocognition of language*, pages 273–316.

Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *science*, 304(5669):438–441.

Hahne, A. and Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by erps. *Cognitive Brain Research*, 13(3):339–356.

Haith, M. M., Bergman, T., and Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science*, 198(4319):853–855.

Hanna, J. E. and Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4):596–615.

Hare, B. and Tomasello, M. (1999). Domestic dogs (canis familiaris) use human and conspecific social cues to locate hidden food. *Journal of Comparative Psychology*, 113(2):173.

Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., and Bornkessel-Schlesewsky, I. (2008). The status of subject–object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, 59(1):54–96.

Heimberg, D. R. (2002). *Zusammenbruch der Gestalt: eine Evozierte-Potentiale-Analyse der Informationsverarbeitung schizophren Erkrankter*. Waxmann Verlag.

Heinze, H.-J., Weissenborn, K., Münte, T., and Künkel, H. (1986). Endogene evozierte potentiale. *Aktuelle Neurologie*, 13(02):61–65.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, 3:31.

Hietanen, J. K., Leppänen, J. M., Nummenmaa, L., and Astikainen, P. (2008). Visuospatial attention shifts by gaze and arrow cues: An ERP study. *Brain Research*, 1215:123–136.

Hoeks, J. C., Stowe, L. A., and Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive brain research*, 19(1):59–73.

Holcomb, P. J. and Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and cognitive processes*, 8(4):379–411.

Holcomb, P. J. and Mcpherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and cognition*, 24(2):259–276.

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

Hommel, B., Pratt, J., Colzato, L., and Godijn, R. (2001). Symbolic control of visual attention. *Psychological science*, 12(5):360–365.

Jachmann, T. K., Drenhaus, H., Staudte, M., and Crocker, M. W. (2019). Influence of speakers gaze on situated language comprehension: Evidence from event-related potentials. *Brain and cognition*, 135:103571.

Jackson, A. F. and Bolger, D. J. (2014). The neurophysiological bases of eeg and eeg measurement: A review for the rest of us. *Psychophysiology*, 51(11):1061–1071.

Johnson, R. (1986). A triarchic model of p300 amplitude. *Psychophysiology*.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Jończyk, R. (2016). Affective word processing in native and nonnative english speakers: A neuropragmatic perspective. In *Affect-Language Interactions in Native and Non-Native English Speakers*, pages 103–131. Springer.

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

Kamp, H. and Reyle, U. (1993). From discourse to logic. *Dordrecht, The Netherlands: D. Reidel*.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., of Biochemistry, D., Jessell, M. B. T., Siegelbaum, S., and Hudspeth, A. (2000). *Principles of neural science*, volume 4. McGraw-hill New York.

Kappenman, E. S. and Luck, S. J. (2016). Best practices for event-related potential research in clinical populations. *Biological psychiatry: cognitive neuroscience and neuroimaging*, 1(2):110–115.

Kellenbach, M. L., Wijers, A. A., and Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: Event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research*, 10(1-2):67–75.

Kelly, S. D., Kravitz, C., and Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and language*, 89(1):253–260.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63.

Kim, A. and Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of memory and language*, 52(2):205–225.

Kingstone, A., Friesen, C. K., and Gazzaniga, M. S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science*, 11(2):159–166.

Knoeferle, P., Crocker, M. W., Scheepers, C., and Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95(1):95–127.

Kok, A. (2001). On the utility of p3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3):557–577.

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

Koornneef, A. W. and Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54(4):445–465.

Kos, M., Vosse, T. G., Van Den Brink, D., and Hagoort, P. (2010). About edible restaurants: conflicts between syntax and semantics as revealed by erps. *Frontiers in psychology*, 1:222.

Kreysa, H. (2009). *Coordinating speech-related eye movements between comprehension and production*. The University of Edinburgh.

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., and Connolly, J. F. (2004). Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with meg. *Cognitive brain research*, 21(1):106–113.

Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., and Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience*, 22(12):2685–2701.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., and Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active english sentences: Evidence from event-related potentials. *Brain and language*, 100(3):223–237.

Kutas, M. and Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*, 4(12):463–470.

Kutas, M. and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (erp). *Annual review of psychology*, 62:621–647.

Kutas, M. and Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological psychology*, 11(2):99–116.

Kutas, M. and Hillyard, S. A. (1980b). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language*, 11(2):354–373.

Kutas, M. and Hillyard, S. A. (1980c). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205.

Kutas, M. and Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia*, 20(5):579–590.

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

Kutas, M. and Hillyard, S. A. (1984). Brain potentials reflect word expectancy and semantic association during reading. *Nature*, 307(5947):161–3.

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 Supplement*, 39:325–330.

Langton, S. R. and Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition*, 6(5):541–567.

Laszlo, S. and Federmeier, K. D. (2007). Better the dvl you know: Acronyms reveal the contribution of familiarity to single-word reading. *Psychological Science*, 18(2):122–126.

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.

Luck, S. J. (1998). Sources of dual-task interference: Evidence from human electrophysiology. *Psychological Science*, 9(3):223–227.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

MATLAB (2017). *version 9.2.0 (R2017a)*. The MathWorks Inc., Natick, Massachusetts.

McCallum, W., Farmer, S., and Pocock, P. (1984). The effects of physical and semantic incongruites on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 59(6):477–488.

McClelland, J. L. and Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, 18(1):1–86.

McPherson, W. B. and Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, 36(1):53–65.

Meltzoff, A. N., Brooks, R., Shon, A. P., and Rao, R. P. N. (2010). Social” robots are psychological agents for infants : A test of gaze following. *Neural Networks*, 23(8-9):966–972.

Meyer, A. S., Sleiderink, A. M., and Levelt, W. J. M. (1998). Viewing and naming objects : eye movements during noun phrase production. *Cognition*, 66:25–33.

Miklösi, Á., Polgárdi, R., Topál, J., and Csányi, V. (1998). Use of experimenter-given cues in dogs. *Animal cognition*, 1(2):113–121.

Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1):92.

Mundy, P. and Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant behavior and development*, 21(3):469–482.

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., and Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, 36(3):217–226.

Newman, R., Connolly, J., Service, E., and McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, 40(4):640–647.

Newman, R. L. and Connolly, J. F. (2004). Determining the role of phonology in silent reading using event-related brain potentials. *Cognitive Brain Research*, 21(1):94–105.

Newman, R. L. and Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, 80(1):114–121.

Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the p3, and the locus coeruleus–norepinephrine system. *Psychological bulletin*, 131(4):510.

Nieuwland, M. S. and Van Berkum, J. J. (2005). Testing the limits of the semantic illusion phenomenon: Erps reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, 24(3):691–701.

Nigam, A., Hoffman, J. E., and Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of cognitive neuroscience*, 4(1):15–22.

O'Halloran, J. P., Isenhart, R., Sandman, C. A., and Larkey, L. S. (1988). Brain responses to semantic anomally in natural, continuous speech. *International journal of psychophysiology*, 6(4):243–254.

Osterhout, L. and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, 31(6):785–806.

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.

Phillips, C. and Sakai, K. L. (2005). Language and the brain. *Yearbook of Science and Technology*, pages 166–169.

Picton, T. W. (1992). The p300 wave of the human event-related potential. *Journal of clinical neurophysiology*, 9(4):456–479.

Pineda, J. (1995). Are neurotransmitter systems of subcortical origin relevant to the electrogenesis of cortical erps? *Electroencephalography and clinical neurophysiology. Supplement*, 44:143.

Pineda, J., Foote, S., and Neville, H. (1989). Effects of locus coeruleus lesions on auditory, long-latency, event-related potentials in monkey. *Journal of Neuroscience*, 9(1):81–93.

Polich, J. (2003). Theoretical overview of p3a and p3b. In *Detection of change*, pages 83–98. Springer.

Polich, J. (2007). Updating p300: an integrative theory of p3a and p3b. *Clinical neurophysiology*, 118(10):2128–2148.

Polich, J. and Kok, A. (1995). Cognitive and biological determinants of p300: an integrative review. *Biological psychology*, 41(2):103–146.

Polich, J. and Margala, C. (1997). P300 and probability: comparison of oddball and single-stimulus paradigms. *International Journal of Psychophysiology*, 25(2):169–176.

Praamstra, P. and Stegeman, D. F. (1993). Phonological effects on the auditory n400 event-related brain potential. *Cognitive Brain Research*, 1(2):73–86.

Pratarelli, M. E. (1994). Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. *Brain and Cognition*, 24(1):137–157.

Pritchard, W. S. (1981). Psychophysiology of p300. *Psychological bulletin*, 89(3):506.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, 85(2):59.

Ristic, J., Friesen, C. K., and Kingstone, A. (2002). Are eyes special? it depends on how you look at it. *Psychonomic Bulletin & Review*, 9(3):507–513.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, 18(4):367–379.

Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The p600-as-p3 hypothesis revisited: Single-trial analyses reveal that the late eeg positivity following linguistically deviant material is reaction time aligned. *Brain and language*, 137:29–39.

Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P. C., and Martín-Loeches, M. (2014). Differential task effects on n400 and p600 elicited by semantic and syntactic violations. *PloS one*, 9(3):e91226.

Schumacher, P. B. (2012). Context in neurolinguistics. *What is a Context?: Linguistic Approaches and Challenges*, 196:33.

Sharbrough, F., Chatrian, G. E., Lesser, R. P., Lüders, H., Nuwer, M., and Picton, T. W. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8:200–202.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., and Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of cognitive neuroscience*, 20(11):2037–2057.

Soltani, M. and Knight, R. T. (2000). Neural origins of the p300. *Critical Reviews in Neurobiology*, 14(3-4).

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

Staudte, M. and Crocker, M. W. (2010). When robot gaze helps human listeners: Attentional versus intentional account. In *Proceedings of the 32nd annual meeting of the cognitive science society*, pages 1637–1642.

Staudte, M. and Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, 120(2):268–291.

Staudte, M., Crocker, M. W., Heloir, A., and Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, 133(1):317–328.

Steinhauer, K., Connolly, J. F., Stemmer, B., and Whitaker, H. A. (2008). Event-related potentials in the study of language. *Concise encyclopedia of brain and language*, pages 91–104.

Sutton, S., Braren, M., Zubin, J., and John, E. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700):1187–1188.

Swerts, M. and Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, 53(1):81–94.

Symons, L. A., Hains, S. M., and Muir, D. W. (1998). Look at me: Five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. *Infant Behavior and Development*, 21(3):531–536.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension.

Thomsen, C. E. (1974). Eye contact by non-human primates toward a human observer. *Animal Behaviour*, 22(1):144–149.

Tipples, J. (2008). Orienting to counterpredictive gaze and arrow cues. *Perception & Psychophysics*, 70(1):77–87.

Tomasello, M., Call, J., and Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55(4):1063–1069.

Uono, S., Sato, W., and Kochiyama, T. (2014). Commonalities and differences in the spatiotemporal neural dynamics associated with automatic attentional shifts induced by gaze and arrows. *Neuroscience Research*, 87(C):56–65.

Van Berkum, J. J. (2009). The neuropragmatics of'simple'utterance comprehension: An erp review. In *Semantics and pragmatics: From experiment to theory*, pages 276–316. Palgrave Macmillan.

Van Berkum, J. J., Hagoort, P., and Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the n400. *Journal of cognitive neuroscience*, 11(6):657–671.

Van Berkum, J. J., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, 1146:158–171.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience*, 20(4):580–591.

Van De Meerendonk, N., Kolk, H. H., Vissers, C. T. W., and Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different erp patterns. *Journal of cognitive neuroscience*, 22(1):67–82.

Van Den Brink, D., Brown, C. M., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus n400 effects. *Journal of Cognitive Neuroscience*, 13(7):967–985.

van Dijk, T. A. and Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Herten, M., Kolk, H. H., and Chwilla, D. J. (2005). An erp study of p600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2):241–255.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., and Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2):394.

Van Petten, C. and Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & cognition*, 18(4):380–393.

Van Petten, C. and Kutas, M. (1991). Electrophysiological evidence for the flexibility of lexical processing. In *Advances in psychology*, volume 77, pages 129–174. Elsevier.

Van Petten, C. and Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33(4):485–508.

Vecera, S. P. and Rizzo, M. (2006). Eye gaze does not produce reflexive shifts of attention: Evidence from frontal-lobe damage. *Neuropsychologia*, 44(1):150–159.

Verleger, R. (1997). On the utility of p3 latency as an index of mental chronometry. *Psychophysiology*, 34(2):131–156.

Verleger, R., Jaśkowski, P., and Wascher, E. (2005). Evidence for an integrative role of p3b in linking reaction to perception. *Journal of Psychophysiology*, 19(3):165–181.

Vissers, C. T. W., Kolk, H. H., Van de Meerendonk, N., and Chwilla, D. J. (2008). Monitoring in language perception: evidence from erps in a picture–sentence matching task. *Neuropsychologia*, 46(4):967–982.

Walter, W. G. (1964). Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. *Nature*, 230:380–384.

Wu, Y. C. and Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6):654–667.

Yamaguchi, S. and Knight, R. T. (1991). P300 generation by novel somatosensory stimuli. *Electroencephalography and clinical neurophysiology*, 78(1):50–55.

Zwaan, R. A., Langston, M. C., and Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological science*, 6(5):292–297.

Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2):162.