ELECTROENCEPHALOGRAPHIC RECORDING DURING NATURALISTIC CONVERSATION: PRELIMINARY INSIGHTS FROM AN EVENT-RELATED POTENTIAL LANGUAGE STUDY

By Caitriona Douglas

Submitted in partial fulfilment of the requirements
for the degree of Masters of Science

at

Dalhousie University
Halifax, Nova Scotia
August, 2021

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

DEDICATION

This thesis is dedicated to my grandfather, David Douglas. My grandfather and I both began our Master's theses and related studies at the same time. There were a lot of jokes about who would become the highest educated family member first. We were planning to have a joint thesis celebration upon completion. Unfortunately, Covid-19 not only disrupted my originally planned thesis experiment, but also took the life of my grandfather before he had the chance to complete his Master's studies. I took a brief pause during the early stage of the Covid-19 in attempts to wait out the pandemic. I seriously debated whether I would return to complete my studies, as I was working on pandemic relief teams where my presence felt more immediately needed. He always inquired about my university progress, would joke about the competition, and encouraged me to complete my degree. For some context, I grew up in a city that was at least a 6-hour flight away from all family, other than my parents. If anyone asks me about where I plan to travel, I always am planning to trips to see family. My grandfather would always tell me to remember to go on an adventure or to plan a fun trip instead. Similarly, while he was happy to hear I was aiding Covid relief efforts, he encouraged me to return to school after the immediate crisis had passed. He emphasized both the importance of doing one's part, while also remembering to do something for oneself. While we won't be able to celebrate our theses competition together, I can keep his memory alive by going on many adventures, whether that be exploring new places, or investigating scientific concepts.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

EEG hyperscanning, which refers to recording simultaneous EEG data from multiple people, are becoming more popular in language cognition research. Studies have mainly focused on neuro-oscillatory dynamics as it is difficult to examine unpredictable occurrences in a highly time-locked manner as necessitated by event-related potential (ERP) studies.

We developed EEG hyperscanning methods to detect ERPs in response to words people hear during a conversation. Specifically, we examined if the N400 response differed between hearing low and high lexical frequency words. Pairs of participants (n=35) had a scripted conversation together while EEG hyperscanning occurred. A control group of pairs of participants (n=41) watched a recording of the same conversations while undergoing EEG hyperscanning. We then transcribed the recorded dialogues to mark the onset times of each word of interest to which ERPs were obtained.

There was a significantly greater N400 response to low frequency words in comparison to high frequency words in both groups, replicating previous findings. These results demonstrate that the N400 ERP experiments can be obtained using stimuli generated by experimental participants engaged in conversation. This opens up new opportunities for more naturalistic ERP studies of language.

ACKNOWLEDGEMENTS

I would like to acknowledge the contributions of many people, without whom, this thesis and degree would not have been possible. First of all, my supervisor, Dr. Aaron Newman whose advice and support has been invaluable. Secondly, I'd like to thank my friends. My friends have served as a spring-board for ideas, sources of words, and have tested out the preliminary version of scripts and experimental tasks. I would also like to thank my committee (Dr. Tim Bardouille, Dr. Steven Aiken, and Dr. Elissa Asp) for their time and insightful feedback. I'd like to thank Dr. Antoine Tremblay for his supervision during the pilot phase of these experiments (and the original idea of detecting ERPs in a completely free and naturalistic conversation). I'd like to thank Cindy Hamon-Hill, Christina Torrealba, Andrew Son, Sean Wang, Erin Morely, and Chinenye Ndulue for their help with running the EEG studies and transcription. I'd like to thank Dr. Colin Conrad for various insightful discussions and for serving as a speaker in the stimulus dialogues. I would also like to thank my parents, and my partner, Vicki , for their encouragement and moral support.

CHAPTER 1 INTRODUCTION

1.1 EVENT-RELATED POTENTIALS IN PSYCHOLINGUISTIC RESEARCH

1.1.1 Introduction to Psycholinguistic Research on Event-Related Potentials

Psycholinguistic research refers to the study of the psychological, cognitive, and neurological underpinnings of language production, language processing, and language development and learning across the lifespan. Obviously, there are many highly relevant reasons for studying psycholinguistics, like the better detection and treatment of speech and language disorders. But psycholinguistics also provides a scientific window beyond only the perspective of medical function versus dysfunction. It also concerns how humans interact with others and their environment from both a social and cultural standpoint. Afterall, language is used by humans to interact with each other, convey meaning or key information to others, and understand the information and meaning conveyed by others, be it through means like posting a sign with instructions, having a conversation with another, or reading a book (O'Connell and Kowal, 2011).

One facet of psycholinguistic research is the study of event-related potentials (ERPs) using electroencephalography (EEG). EEG is a neuroimaging technique that records the electrical potentials generated by neurons in the brain, using electrodes placed on the scalp (Luck, 2014; Newman, 2019). EEG has high temporal resolution, on the order of milliseconds, which allows researchers examine what happens in the brain prior to a stimulus occurring, exactly when the stimulus occurs, and immediately after the stimulus occurs. Event-related potentials (ERPs) are obtained from EEG waveforms that are time-locked to events of interest (like the onset of a word), and are typically averaged over a number of similar events to identify reliable patterns associated with the stimulus. ERPs are thought to be related to, or indicators of, specific cognitive processes (Luck, 2014).

EEG research began in the same time period as other basic biological signal monitoring research, like electrocardiography (EKG), when the research focus and philosophy revolved around detecting specific biological signals that were anticipated to provide precise insight into biological functions and processes (Collura, 1993). As a result, research methods focused on isolating these biological responses and removing any environmental interference. ERP research began to flourish from the late 1970s onwards, as cognitive scientists determined that ERPs included both exogenous (automatic) responses to stimuli, like a response to hearing a sound, and endogenous responses that reflect cognitive processing, such as a response to hearing a grammatical error in a sentence (Woods, 1990). ERP research often uses very simple tasks where participants are shown a series of stimuli that are identical or highly similar. Often, only a single stimuli feature is varied while all other features are matched in order to compare conditions that differ in a specific feature of experimental interest. Researchers then examine the responses of participants to the different stimuli type in comparison to the predicted or routine stimuli type (Woods, 1990; Luck, 2014).

An ERP component refers to one of the component waves that makes up an ERP waveform (Woodman, 2010). To characterize these components, ERP components are given names that reflect their timing, the direction of voltage change, their scalp localization, or task-related properties. For example, the N400 is a negative ERP component that appears 400 ms after stimulus presentation (Kutas & Hillyard, 1980). Similarly, the P600 refers to a positive component occurring 600 ms post-stimulus and the N100 is a negative component that occurs 100 ms post-stimulus (Luck, 2014). The N400 ERP was detected because a distinct negative going potential occurred in response to semantic anomalies like "I take my coffee with milk and *dog*." (where *dog* is the semantic anomaly) 400 ms after participants saw the word "dog". The N400 ERP is thought to reflect meaning processing, lexical access, semantic memory, and mark the process of learning novel words (Kutas & Federmeier, 2011).

The N400 ERP is detected in many scenarios other than semantic anomalies. For example, the N400 ERP is modulated by lexical frequency. Lexical frequency refers to how often a word is used in language. A word like "chair" has a higher lexical frequency as it occurs 62838 times per million words (log 4.8) according to the Contemporary Corpus of American English (COCA- a corpus that contains words and tabulates their linguistic properties, including lexical frequency, drawing from a collection of over 1 billion words collected from text excerpts of contemporary books, congress proceedings, tv shows, and news articles) (Davies, 2008). Whereas the word "galoshes" has a lower lexical frequency as it occurs 214 times per million words (log 2.3) according to COCA (Davies, 2008). In response both to words occurring within sentences and individual words, words with a lower lexical frequency consistently elicit a larger amplitude of

N400 response in comparison to words of higher lexical frequencies (Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber, Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017).

A technique that is commonly used in ERP studies is called a word-by-word presentation technique (also known as rapid serial visual presentation or RSVP), where a sentence is presented to a participant by displaying only one word on the screen at a time. However, in daily life people typically read phrases and complete sentences at their own pace and can re-read parts as needed. So far, EEG studies that compare the EEG results of RSVP tasks and free reading methods, where eye-tracking is used to determine the timing of when participants read words presented in blocks of text (e.g. paragraph) have generated similar ERP results (Metzner et al., 2015). However, Metzner and colleagues (2015) found that text comprehension was greater in free reading tasks compared to RSVP tasks. Further, during the free reading task, when participants encountered syntactic or semantic anomalies in the text, they looked back (i.e. regression) at the preceding text. These regressions corresponded with the occurrence of the P600 ERP component (associated with detecting syntactic anomalies or encountering passages that are difficult to parse; Leckey & Federmeier, 2020). Perhaps this indicates that the component may be part of a cognitive cue or process to re-evaluate information when an irregularity is detected to ensure that the irregularity was not self-generated due to miscomprehension. The difference in comprehension between the two tasks also provides some evidence that an artificial presentation method can alter language cognition.

1.1.2 The influence of context on ERPs

There is also a growing recognition of the importance of the context surrounding stimulus presentation at a smaller-scale level in ERP research. As previously mentioned, the N400 ERP is a negative ERP component that appears 400 ms after stimulus presentation (Kutas & Hillyard, 1980) that is thought to reflect meaning processing, lexical access, semantic memory, and mark the process of learning novel words (Kutas & Federmeier, 2011). The N400 ERP is known to be modulated by context. Larger N400s occur in response to semantic anomalies — words whose meaning does not fit with prior context but that are not grammatically incorrect — than to semantically congruent words (Kutas & Hillyard, 1980). For example, in the sentence "The peanut was in *love*", "love" is a semantic anomaly. However, if there is context given prior to exposure to this stimuli sentence, like a story about an anthropomorphized peanut that falls in love with an almond, an enhanced N400 is not observed (Nieuwland & Van Berkum, 2006). Conversely, if there is a sentence that would typically not be thought as containing a semantic anomaly like "The peanut was *salted*" presented at the end of the story about the romance between two anthropomorphic nuts, participants process "salted" as a semantic anomaly with a characteristic N400 response of greater magnitude (Nieuwland & Van Berkum, 2006).

Findings that suggest the influence of real world and personal experience on cognitive processes can be found be found in earlier ERP research as well. Müller and Kutas (1996) found that there was a significantly greater amplitude in N400 response when participants heard their own names versus when they heard other proper names (proper names are first names like, *Mary*, *John*, or *Sam*). There was no difference in N400 response when comparing proper names (not including the participant's own name) and common nouns (i.e., regular nouns like *bird*, *chair*, or *table*). Furthermore, other factors such as intonation and loudness, and whether words began with different consonant types failed to modulate the N400 response. This indicates that the difference in N400 response was due to participants' hearing their own names rather than their names having linguistic properties that differed substantially from the stimulus set of proper names that would generate a difference in N400 response. In line with these findings, Fischler and colleagues (1987) instructed participants to take on an assumed name and respond to that name as if it were their own throughout the study. The N400 responses to stimuli containing their actual name and false names were more similar to each other in comparison to the N400 responses to participants' assumed name. This suggests that people process names that they consider as referring to them differently than names of others. This is a consciously modulated response that can quickly adapt to new information, such as being asked to be referred to by another name and ignore their own name, as the participants elicited a greater N400 response to their assumed name and had similar lower amplitude N400 responses to their own name (which they were asked to ignore) and false names. There cumulatively, both context and real-world experience appears to modulate the N400 ERP, pointing to the importance of conducting experiments that can examine these responses in naturalistic scenarios. Examining only how language is processed when isolated from context and the real-world experiences of language may lead to deficits in the ecological validity of the scientific understanding of language cognition.

1.1.3 Proposals and Calls for an Understanding of Naturalistic Language Processing

In the early 2010s, there were a few articles published by neuroscientists calling for more research in naturalistic settings (e.g. watching a movie, reading books or signs, a cooperative social interaction (playing a game, playing a musical duet, etc.) with a specific focus on the importance of looking at cognition in social scenarios rather than only conducting research examining an isolated individual. The general arguments were that as humans are a highly social species, examining only individuals in isolated conditions may not give a comprehensive picture of human cognition and that the isolation itself may be processed by participants as an abnormal experience (Konvalinka & Roepstorff, 2012; Schilback et al., 2013). Calls for experimentation that allowed social interaction, social participation, or that involved more than one participant generally became known as calls for two-brain neuroscience or a two-brain approach (Liu et al., 2018).

Subsequently, similar calls were reflected in papers published by psycholinguistic researchers. Their calls emphasized the importance of developing methods that allowed for the presentation of stimuli in ways that resembled how people interacted and encountered language in the real world rather than only using artificial and isolated experimental tasks. These papers also discussed the importance of developing two-brain neuroimaging approaches for psycholinguistics, as human interaction is one of the main settings where people encounter language — be it in conversation, a parent reading a story to a child, or a teacher lecturing a class (Brennan, 2016; Khulen et al., 2015; Hari, Malinen, & Parkkonen, 2015; Schoot, Hagoort, & Saegert, 2016).

The limitations of some techniques were apparent. Functional MRI and magnetoencephalography (MEG) experiments both involve large, expensive scanners that need to be electromagnetically shielded and kept far from other equipment (including other scanners), creating logistical and financial barriers to running scans for multiple participants at once. Therefore, an experimental technique that was more appropriate to studying social interaction was EEG, as the equipment is far less expensive. This means multiple EEG systems can be purchased, and multiple systems and people can be present in the same room during an experiment. EEG also has high temporal resolution, which allows for event-level analyses of linguistic social interactions. At the time the papers calling for the development of naturalistic methods in psycholinguistics came out (Brennan, 2016; Khulen et al., 2015; Hari, Henriksson, Malinen, & Parkkonen, 2015; Schoot, Hagoort, & Saegert, 2016), there had been some research that examined the EEG activity of people listening to or narrating short stories using oscillatory analyses (which average over longer period of time than ERPs; Khulen, Allefeld, Haynes, 2012), but none that examined ERPs or that had multiple subjects completing an experimental task together simultaneously. Khulen and colleagues (2012) recorded EEG from participants assigned as either listeners or speakers separately (i.e. completely unique and independent EEG recording session for speakers and listeners). Listeners would observe a pre-recorded narration from two superimposed video recordings of speaking participants and were asked to only attend to a single speaker. Listeners asked to attend to the same speaker had higher correlations of EEG activity in the However, a method called EEG hyperscanning, that allows simultaneous EEG recording of multiple brains at once (Babilloni et al., 2006), showed promise in terms of allowing both multi-participant data collection and event-level analysis.

1.2 EEG HYPERSCANNING

1.2.1 Brief Introduction and History of EEG Hyperscanning

EEG hyperscanning refers to simultaneously recording EEG data from multiple people at once (Babilloni et al., 2006). Accounts of EEG hyperscanning date back to the early history of EEG. William Gray Walker, who played a major role in advancing the use of EEG as a clinical tool, theorized about similar brain rhythms between individuals as markers of interpersonal cooperation and compatibility (Hayward, 2001). Other than Walker's accounts, the earliest EEG hyperscanning study dates back to 1965. It is a twin study that claimed to demonstrate the ability to induce alpha rhythms in pairs of identical twins placed in separate rooms by inducing alpha rhythms in one of the twins that were then mirrored by the other despite their separation (Duane & Behrendt, 1965). As Barrazza, Dumas, Liu and colleagues (2019) point out, no EEG hyperscanning studies were pursed again until 2006. This is likely due to the unsettling relationship between EEG hyperscanning and attempts to demonstrate the existence of extrasensory perception.

1.2.2 EEG Hyperscanning for Naturalistic Social Interaction Research

Fabio Babiloni and colleagues implemented the first modern EEG hyperscanning in 2006. Babiloni and colleagues detected synchronized neural oscillatory activity in the pre-frontal areas between participants while they played a cooperative card game. Similarly, interbrain synchronization has been reported in studies of musical ensembles and duets (Linderberger et al., 2009). Interbrain synchronization was also detected by Dumas and colleagues (2010) in a task in which participants were asked to mirror another's hand gestures. EEG hyperscanning has now been implemented in many diverse social task scenarios, like finger-tapping tasks, board games, a prisoner's dilemma task, viewing movies and recordings, and in a classroom learning scenario (for recent reviews, see Liu et al., 2018; Czeszumski et al., 2020).

So far, EEG hyperscanning experiments have mainly focused on identifying inter-brain synchrony by examining neural oscillatory activity rather than on the exact time-locked responses of participants to events that occur. Neural oscillatory activity refers to the rhythmic fluctuations in the excitability of neurons or populations of neurons that are detectable as patterns of cycling electrical activity with distinct characteristics of frequency (number of cycles per second) and amplitude (Cohen, 2014). One of the reasons most research focuses on neural oscillatory activity is because EEG hyperscanning research requires a separate EEG system and amplifier for each participant, which introduces challenges in synchronizing each participants data to each other and the experimental stimuli (Barrazza et al., 2019; for recent reviews, see Liu et al., 2018; Czeszumski et al., 2020). As previously mentioned, ERP studies require precise time-locking of the EEG data to events of interest. Variance of even tens of milliseconds in the synchronization of EEG and stimuli results in uninterpretable data. In hyperscanning studies, experimental designs typically attempt to examine naturalistic behavior and rely on unpredictable, participant-generated stimuli. For this reason, the majority of EEG hyperscanning studies so far have not focused on ERPs but rather on neural oscillatory activity. Neural Oscillatory activity is measured as power in specific frequency bands, averaged over periods of hundreds of milliseconds or even seconds. Averaging over longer periods of time and mapping a one-dimensional signal from the time domain into a two-dimensional function of time and frequency creates a time-smearing effect. As a result, neural oscillatory research requires a lower degree of temporal accuracy (Cohen, 2014).

The synchronization issue and temporal resolution of EEG hyperscanning can be resolved by presenting exactly time-locked stimuli to both participants simultaneously while also generating trigger codes or stimulus time locking independently for each participant. ERPs have been successfully detected using EEG hyperscanning. Loehr and colleagues (2013) demonstrated that participants exhibited two ERPs, the FRN and P300 (associated with error detection), in response to both a participant's own errors and their partner's errors during a piano duet task. In this case, the stimulus time-locking could be time locked to the recorded button presses on the electric piano keyboards (Loehr et al., 2013). Tardif and colleagues (2018) also have demonstrated that an ERP called the LPP can be detected in partnered pairs completing a social viewing task where trigger codes were sent to both EEG systems at the onset of each stimulus. Both studies' findings are congruent with ERP findings in similar, single-participant EEG studies. However, a limitation of both studies is that the examination of naturalistic behaviour is confined to interaction that can be measured by a machine (e.g. piano key button presses) or requires that participants passively observe stimuli with pre-determined event timing. If multiple participants' EEG data could be directly synchronized as single recording simultaneously alongside a recording of naturalistic participant generated behaviour, then post-hoc time locking could be generated by marking the events in the behavioural recording. Such a method would allow for participants to interact at their own pace and in a less constrained manner without jeopardizing the temporal resolution required for ERP analyses.

1.2.3 EEG Hyperscanning for Naturalistic Psycholinguistic Research

When the program of research related to this thesis began in 2016, no hyperscanning ERP studies had been published in the field of psycholinguistics. Indeed, there were very few psycholinguistic studies that employed EEG hyperscanning methods at all. Even as of the present date, the majority of EEG hyperscanning research in psycholinguistics has been focused on neural entrainment to speech rhythms and on inter-brain synchronization (Liu et al., 2018; Czeszumski, 2020). So far, experiments that allow a free conversation or an interactive dialogue have been focused on whether inter-brain synchrony decreases or increases based on relationship type and have been focused on if there is synchronization between mothers and infants (Nguyen et al., 2021) and synchronization differences between couples and strangers (Kinreich et al., 2017). However, a major timing challenge is still presented for examining exact responses to naturalistic speech. First, the general ERP research guidelines recommend eliminating any potential noise as muscle activation created by jaw muscle tension and facial movements creates considerable noise in EEG data (muscle activation also generates electrical activity that is detectable by EEG) (Luck, 2014; Newman, 2019). Second, an audio recording of participant speech would need to be directly time-locked and synchronized with the participants' EEG recording. This recording would need to be analyzed post-hoc by researchers to generate event-timings for precise stimuli occurrences like the onset or offset of words.

An EEG study by Fjaellingsdal and colleagues (2020) in which a participant (fitted with an EEG cap) and a non-EEG equipped confederate completed a sentence co-creation task in which each person took turns producing words to create a sentence (e.g. P says "I", confederate says "walked", P says "home"). For 30% of the trials, the confederate would utter an incorrect gender marker referring to a character involved in the sentence (e.g., "Susan walked to *his* house"). In response to these gender incongruent markers, there was a greater N400 ERP amplitude than for trials where the gender was congruent which is in line with single participant studies (Guajardo & Wicha, 2014; White et al., 2009). This and previously-mentioned ERP hyperscanning studies (Loehr et al., 2013; Tardiff et al., 2018) indicate that ERP responses and effects are in line with single participant studies conducted under more controlled circumstances.

There is still a question of whether participants will process more naturalistic experimental scenarios differently than isolated experimental settings. It has been established that ERP responses are influenced by modality as there are differences in ERP latency between studies using the visual and auditory stimulus display methods (Kutas & Federmeier, 2011) ERPs are also highly sensitive to modulations of experimental paradigms. For example, the N400 ERP component occurs earlier when participants listen to sentences read at a natural speaking rate, in comparison to when the participants listen to sentences where there are brief (750 ms) pauses between each word (Holcomb & Neville, 1991). Similarly, a small change like changing the font of text used for the linguistic stimuli influences the amplitude of the N/P150 ERP components (Chauncey, Holcomb, & Grainger, 2008). Given, Debruille, Brodeur, and Porras (2012) found that people exhibited greater N300 ERP responses to photos of dummies in comparison to faces. If ERPs are modulated by modality changes and social changes (human faces vs dummies), then it is foreseeable that an in-person live social interaction setting may produce different ERP effects than scenarios where participants passively observe stimuli.

1.3 THE CURRENT INVESTIGATION

1.3.1 Research Questions

The main research question for the program of research related to the thesis was whether or not it would be possible to detect event-related potentials (ERPs) in response to words people heard while they were participating in a conversation; more specifically whether we could replicate the modulation of N400 amplitude by lexical frequency (with more negative amplitudes for lower frequency words) in a conversational setting. Secondly would ERPs detected during a conversational scenario be similar or different to ERPs when participants only observed, rather than acted as a participant in, the conversation?

1.3.2 Our Approach

Initially, in planning this study we aimed to analyze participants' responses to content words (like nouns) that they heard over the course of a free (i.e., unconstrained by the experimenters) conversation in hopes of replicating the N400 lexical frequency effect (Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber, Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017). However, in pilot testing, participants rarely used content words and instead mainly relied on proper names and referential words (ex: "it", "thing", "stuff"). In addition to this, content words were often highly irregular (ex: "thrombocytopenia", "emoji", etc.) so they did not appear in linguistic corpora used to categorize lexical frequency. In addition to these factors, the words used in each conversation could differ substantially (e.g., two philosophy majors would discuss very different subjects and use very different words than two mathematics majors). Therefore, as the aim of the program of research was to validate a method to detect ERPs in conversation, we decided to develop a scripted dialogue to ensure that all the participants engaged in or viewed the same conversation and linguistic stimuli.

We developed six scripted dialogues that contained 120 low and 120 high frequency words, that would serve as target stimuli for which ERPs would be measured. Pairs of friends were recruited to take part in a conversation study where they would have a scripted dialogue together while we recorded EEG activity from both participants using an EEG hyperscanning set-up. Participants were given scripts that contained the sentences they were to read aloud, and indicated breaks when their partner was to speak. Audio of the conversations were recorded simultaneously alongside the EEG recordings. The audio of the conversations were transcribed and the target nouns' onset time were marked and used as time markers of when the stimuli occurred during the EEG data analysis stage. As we were also interested in if the act of participating in a conversation impact N400 amplitude, we also conducted a control study in which participant pairs were asked to watch a recording of two people engaged in the same scripted dialogues that were used in the coversation study.

1.3.3 Hypotheses

Three separate hypotheses were made to address our research questions. First, there would be a significantly greater N400 ERP amplitude response to low frequency words in comparison to high frequency words during the conversational task, replicating past results (Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber, Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017). Second, it was hypothesized that there would be a significantly greater N400 ERP amplitude response to low frequency words in comparison to high frequency words during the control task, replicating past results Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber, Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017). Third, it was hypothesized that that there would be a significant difference between the frequency effect contrast size of the control group and the conversation group. In-person interaction and passive viewing are different experiences, in the same vein of how reading written text differs from hearing speech. It is predicted that there will be task-related differences that result in a significantly different N400 ERP amplitude as viewing a video is different perceptual experience than viewing and participating in a live in-person conversation. Prior research has established that changing task modalities (from visual to auditory) and changes in task stimulus aesthetic (like changes in font) modulate the N400 response (Holcomb & Neville, 1991; Chauncey, Holcomb, & Grainger, 2008). However, as there has been little EEG research under these conditions, we did not predict the direction of this predicted effect (i.e., which group's ERP amplitude would be greater).

SUMMARY OF HYPOTHESES

H1: There will be a significant difference between low and high frequency nouns that occur during the conversational task. Specifically, the prediction is that low frequency nouns will correspond with a greater N400 ERP amplitude than high frequency nouns over the midline central electrodes including Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, and CP2.

H2: The control group will exhibit a significant difference in N400 ERP amplitude between the responses to low and high frequency nouns. Specifically, the prediction is that low frequency nouns will correspond with a greater N400 ERP amplitude than high frequency nouns over the midline central electrodes including Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, and CP2.

H3: The control group will have a significantly different N400 ERP amplitude response in comparison to the conversational group over midline central electrodes including Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, and CP2. However, no specific prediction regarding the nature of the difference between groups is made.

CHAPTER 2 METHODOLOGY

2.1 PARTICIPANTS

We recruited two groups of participants, who participated in this study in pairs. We asked that all participants recruit a friend to participate in the study with them that they would feel comfortable talking to for half an hour. We did not disallow any relationships that met this familiarity criterion, so pairings included casual acquaintances, coworkers, friends, romantic partners, and family members. For the conversation group, we recruited 25 pairs of participants (50 people in total). The data of 35 participants was used in the analysis and the data of 15 participants was rejected due to participants disclosing during or after the study that they failed to meet the eligibility criteria (described in the next paragraph) or due to equipment failure. For the control group, we recruited 28 pairs of participants (54 people in total). The data of 41 participants was used in the analysis and the data of 13 participants was rejected due to participants disclosing during or after the study that they failed to meet the eligibility criteria (described in the next paragraph) or due to experiment equipment failure.

All participants were current university students ranging from ages 18-30 in Halifax, Nova Scotia. Participants were screened prior to the study to verify their eligibility. All participants were native English speakers and had (by self-report) normal or corrected to normal hearing and normal or corrected to normal vision. Potential participants who had medical conditions or who were taking prescription medications known to affect EEG activity, had hairstyles or unremovable headgear that would prevent close contact between the EEG electrodes and the scalp, or who were claustrophobic (because the testing room was small) were excluded from participating in the study.

All participants received either $20 or course credit points in return for participation. Participants were recruited using posters on Dalhousie University campus, ads on Facebook, by word of mouth, and by the Dalhousie Psychology and Neuroscience department's SONA course credit participant pool. This study was approved by Dalhousie University Social Sciences and Humanities Research Ethics Board, and informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

2.2 STIMULI

The experimental stimuli ("target words", to which ERPs were measured) were 240 nouns contained in 6 scripted conversations on different topics between two people. These stimuli were divided into 120 high frequency and 120 low frequency words. High frequency nouns were classified as nouns with a frequency ≥ 1800 (log 3.25) occurrences per million words, according to the Contemporary Corpus of American English (COCA; Davies, 2008). Low frequency nouns are classified as nouns with a COCA frequency ≤ 1500 (log 3.17) occurrences per million words. All nouns were between 1-3 syllables in length. The high and low frequency word lists were matched at a list level for mean orthographic & phonological neighbour counts, phonological neighbour frequency, number of phonemes, and number of morphemes, based on the English Lexicon Project repository (Balota et al., 2007) and the COCA. The low and high frequency word lists were compared using t-tests on each of those variables and none were significant (all $p > 0.05$). The characters, subject of conversation, and storyline differed in every script. Each script contained 40 low-frequency and 40 high-frequency nouns, divided such that each participant spoke 20 high and 20 low frequency target words, and heard their interlocutor speak 20 high and 20 low frequency target words.

The videos shown to the control group featured two people seated across from each other at a table having the six scripted conversations. The roles were performed by a male and female speaker (members of the research lab). The two people were reading the lines from the scripts using laptops set-up as teleprompters off-screen behind each other's heads. Each script was separate video. Each video was about 3-4 minutes in length.

2.3 PROCEDURE

2.3.1 Conversation Group

Figure 1 displays a diagram of the experimental set-up for the scripted conversational dialogue study. Participants were randomly assigned either to be participant 1 or participant 2 and were given paper copies of the six, 2-page dialogues specific to their assigned role in the conversation (each script had only the lines for that participant to read shown). Once participants were equipped with EEG caps (see below for details), they were asked to sit across from each other at a table in a small EEG testing room. A member of the research team provided verbal instructions, including that each participant would read their lines and then need to listen carefully to their partner's response to determine when it was their turn to speak their next line. Each participant was given a script that contained only their own lines and blank lines that indicated when their partner would be talking, so they did not see their partner's lines (see Appendix A for an example of a single participant's script). Participants had to rely on social speaker-listener and dialogue content cues to determine when their partner was finished speaking and they should speak their lines.

Participants were told that we were interested in their comprehension and language processing of the stimuli produced by their partner (i.e., their partner's lines). Participants were instructed to remain as still as possible and refrain from blinking or reading ahead when they were listening to their partner. After answering any questions, the researcher started EEG recording and closed the door, and participants began their first scripted conversation. The participants were monitored on a live video and audio feed by the researcher, who would re-enter between scripts to re-adjust equipment or lower electrode impedances when needed, or to give advice to their participants (like reminding them not to move). The participants could communicate with the research assistant at any time and request breaks when needed, report problems, or ask questions. The participants read through all 6 scripts in sequence, which took approximately 20 minutes in total. After the task was complete, the research assistant re-entered the testing room and removed the EEG equipment from the participants and thanked them for their participation.

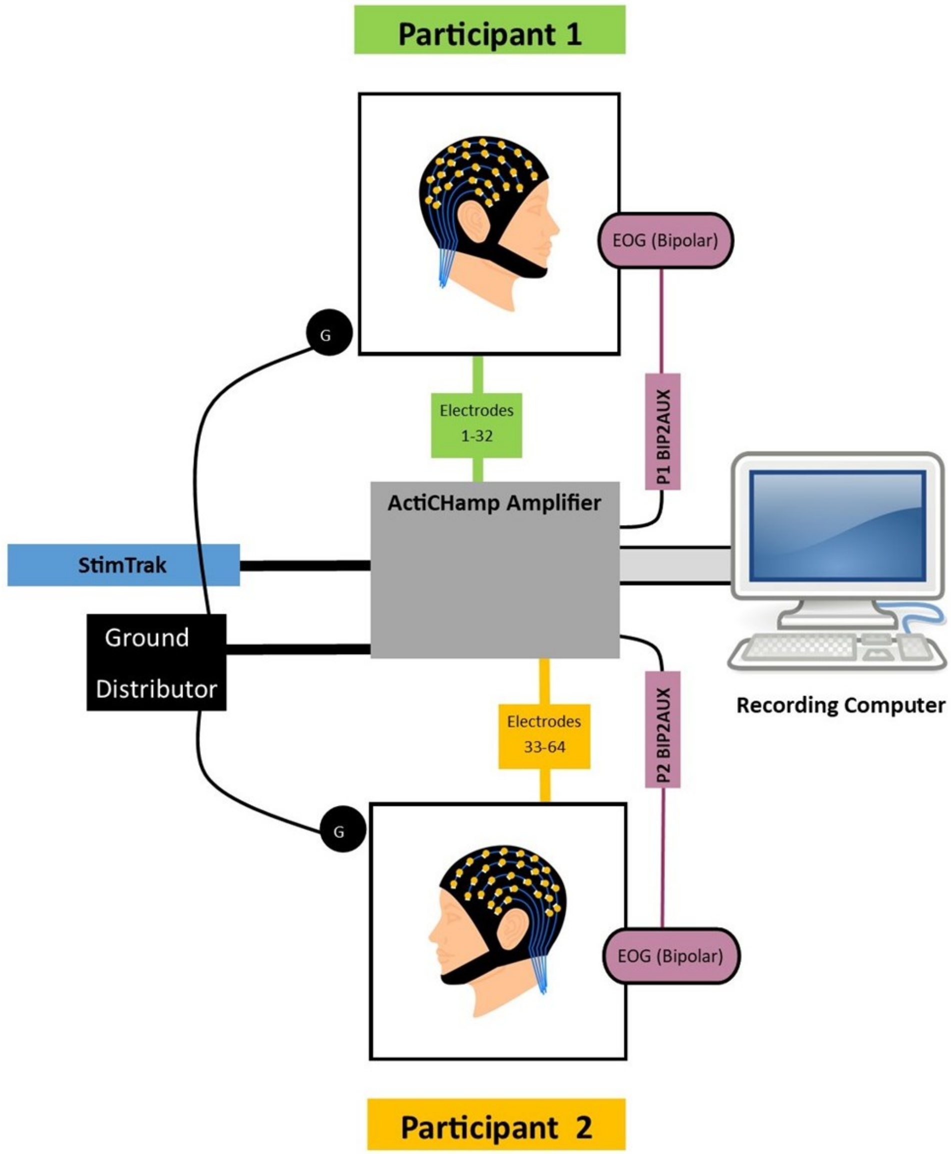

Figure 1: A diagram of the EEG hyperscanning set-up using for both the conversation and control group.

2.3.2 Control Group

The procedures for the control group were similar to that of the conversation group, except as follows. Although participants were still tested in pairs, rather than sitting facing each other and reading scripts, control participants sat side by side facing a 23” computer monitor (ASUS VH232H; Asus, Taipei City, Taiwan) and a single audio speaker (Mackie MR5 mk3, LOUD Technologies, Woodinville, WA, USA) , and watched videos of the six scripted conversations. Participants were instructed to attend to the videos and to try to minimize motion and blinking during the videos. They were also informed that we were interested in their comprehension and language processing of the video stimuli and that we would administer a short questionnaire based on the content of the videos, at the end of each of the six videos. The questionnaires’ purpose was to assess if participants were paying attention to the videos. If participants answered fewer than 3 of the questions correctly, their data would be rejected due to inattention. After each video of a scripted conversation, the participants would complete a questionnaire. As the videos were 3-4 minutes each and the questionnaires took about 2-5 minutes to complete, the task was approximately 20 minutes in total.

2.4 EEG DATA ACQUISITION

We used a Brain Products ActiChamp (Brain Products, Gilching, Germany) 64 channel EEG amplifier with a ground splitter (Brain Products) to record EEG data and audio data from two participants simultaneously, using 32 Ag-AgCl active electrodes (ActiCap; Brain Products) and one ground electrode per participant. Participants were fitted with elastic EEG caps that each included a ground electrode located over the anterior midline of the scalp (FCz), that was connected the ground splitter device. Each participant had 28 electrodes placed on their scalp following the International 10-10 system, 4 EMG electrodes placed over their left and right their masseter and larynx areas, and a bipolar electrode pair (connected to the amplifier via a Brain Products Bip2Aux adapter) to measure EOG, placed with one electrode above one eye, and the other lateral to the outer canthus of the same eye (see Figure 2 for depiction. We performed an impedance check to verify that all electrodes had a low impedance level (< 25 kOhm) which indicated that there was a good connection between the scalp and the electrode. If any electrodes had a high impedance, we rubbed the electrode gently and added more electrolyte gel if needed, until the connection improved and the impedance lowered.

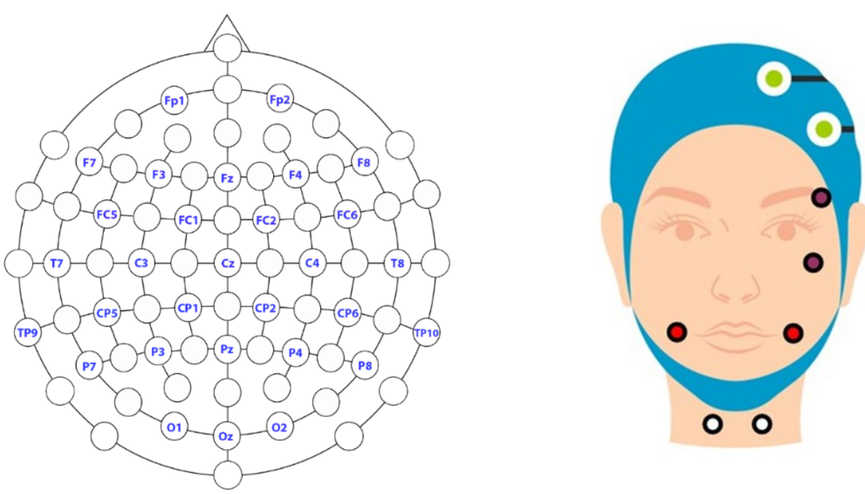

Figure 2: The electrode montage for each participant. The left image displays the 28-electrode EEG montage with positions based on the International 10-10 System. The right image illustrates the placement of the EOG and EMG electrodes.

We recorded the audio of participants' conversations using a StimTrak device (BrainVision), which has a built-in microphone (with Gain set to "Mic" and Trigger sensitivity set at 1 V). The mic was placed in the middle of the table where the participants where seated across from each other for the conversation study. For the control study, the mic was placed in the middle of a table between the two participants who were seated side by side. The StimTrak device allowed us to collect audio data to a channel in each EEG raw data file, meaning that the audio recording was perfectly synchronized with the EEG data. In the control study, the microphone was used similarly, to record the audio produced by the speaker that participants listened to. We used a sampling rate of 10,000 Hz (to ensure audio quality sufficient for transcription), a low pass filter at 2,500 Hz and a 60 Hz notch filter when recording the EEG data. Average reference was used for data acquisition.

2.5 DATA PREPROCESSING AND DATA ANALYSIS

The first step of data analysis involved extracting the audio data from the EEG data using the MNE-Python (v. 0.22; Gramfort et al., 2013) package, and exporting the audio to a WAV file using the SciPy Python package (v. 1.3.1; Virtanen et al., 2020). We then reviewed the audio file using Praat (Boersam & Weenink, 2021). We identified the time points of the onset of each target noun and manually recorded these in a spreadsheet, where we also encoded for each word the listener/speaker roles of the two participants, and the low/high frequency status of each item. For both groups, we noted for exclusion any trials where there were disruptions such as a participant speaking over the scripted dialogue, noise, or movement, breaks in the dialogue speech, mispronunciation, and or sentence production errors by marking these as trials to reject as part of the audio transcription process.

Then we created event markers for EEG data analysis by importing this spreadsheet into Python and using the timing and stimulus information to define the markers for ERP segmentation. All subsequent preprocessing steps were performed using Python and the MNE-Python package.

After the audio was extracted to WAV files, the audio channel was removed from the EEG data and the EEG data were decimated to a sampling rate of 500 Hz. The continuous raw data were filtered for later analysis using a 0.1–20 Hz zero-phase FIR bandpass filter with transition bandwidths of 0.1 and 5.0, respectively. A copy of the data were also made and bandpass filtered from 1.0–20 Hz; these data were used as the input to independent components analysis (see below) to define and remove artifacts. The data from the four EMG channels (which were recorded average-referenced) were reduced to two bipolar channels per participant, one for the pair of electrodes on the left side and one for those on the right.

Data were segmented into epochs consisting of 200 ms prior to, and 1000 ms after, the event codes marking the onset of each critical word. Data were then separated into individual datasets for each participant, for further processing. Segmented data were first visually inspected for large, paroxysmal artifacts or excessively noisy channels, and such trials/channels were excluded from further preprocessing (in both the copy of the data used for ICA, and the copy of the data filtered from 0.1–20 Hz for final analysis)). Independent component analysis (ICA; Jung et al., 2000) using the FastICA algorithm (Hyvärinen, 1999) was used to identify and remove artifacts, including blinks and noise from eye movements and facial muscle motion. This was done manually by a trained member of the research team, and reviewed by another member of the team. We used the information obtained from the EMG electrodes and bipolar electrodes to help identify bad trials due to noise created by facial motion and speech through visual inspection of the data that were excluded from further data analysis both prior to ICA and post ICA. After ICA, each trial was again visually inspected, and any trials with remaining artifacts were excluded from both copies of the data. Following this, the ICA weight matrix was applied to the 0.1–20 Hz filtered data to exclude artifacts, and this dataset was then re-referenced to the average of all electrodes.

Participants in the conversation group heard half the stimuli and spoke the other half; ERPs were only examined for the stimuli that an individual heard, not those that they spoke aloud. However, in the control group, neither participant spoke; rather both participants watched the conversations spoken by other people. Thus all participants in the control group heard all the words. For the purposes of comparing results between these two experiments, for the control group we only included in analysis half of the words each participant heard. This was done by randomly assigning participants in each dyad as 1 and 2, and including data in analysis only for the words they would have heard had they received that participant assignment in the conversation experiment.

2.6 STATISTICAL ANALYSES

For statistical analysis of the ERP data, the dependent variable was the mean amplitude of the potential at each electrode between 400-600 ms after noun onset, computed for each subject, trial, and channel. We compared and contrasted the low and high frequency noun conditions at 400-600 ms in order to capture the period of the potential N400 effect as this is within the established time window of the N400 ERP (Kutas & Federmeier, 2011), and consistent with past N400 studies conducted in our lab (e.g., Newman et al., 2012; Muise-Hennessey et al., 2016). Prior to statistical analysis, the data was examined for outliers. Individual data points (at the trial/electrode level) were rejected due to meeting criteria as an outlier using a threshold of $z = \pm 2.5$.

Statistical analysis was performed using linear mixed effects modelling, implemented in the *mgcv* package (Wood, 2011) in the R statistical software (v. 4.0.4; R Core Team, 2021). Separate modelling was performed for each group/experiment separately, an then for the comparison between groups. In each case, we fitted a family of models varying in their fixed and random effect structures as well as the depth of interactions between fixed effects, and compared the models using the Akaike information criterion (AIC — Akaike, 1973; Wagenmaker & Farrell, 2004). The model with the lowest AIC was selected as the best one, and if multiple models had AIC values within 4 points of each other, the one with the simplest structure (fewest terms and least complex interaction structure) was selected. In the analysis of each group separately, the fixed effects tested were frequency (low/high) and channel (all 28 scalp electrodes), and their interaction; random effects included random intercepts for subjects, random slopes for channels within subjects, random slopes for words within scripts, and random slopes for participants within dyads. For the between-group analysis, group (conversation/control) was included as an additional fixed effect (testing up to the three-way interaction of frequency × channel × group). In each analysis, the final model was checked using the *gam.check* function from the *mcgv* library in R to confirm that the residuals were normally distributed (which in all cases they were).

In examining the results of the best-fitting models, the cut-off for significance was $p \leq .05$. Significant main effects and interactions were further explored using pairwise *t*-tests, which were corrected for multiple comparisons using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995).

CHAPTER 3 RESULTS

3.1 BEHAVIOURAL DATA

A component of the thesis research was demonstrating that the novel experimental methods outlined were suitable for assessing cognitive responses to the stimuli. As a result, behavioural data and inspected to verify firstly that participants were attentive during the experimental tasks and secondly to ensure that participants were successfully able to complete the experimental tasks.

3.1.1 Conversation Group

As mentioned in the methods chapter, trials were rejected from the conversation experiment due to participant error, such as the listening participant interrupting or interjecting during the speaker's turn or the speaking participant making an error (e.g. stutter, mispronunciation, missed word, etc.). The rejection rate due to these exclusions ranged from 0 (no errors) to 25.21%. The average number of trials rejected per experimental session due to participant error was 4.52%.

3.1.2 Control Group

In order to ensure that participants were paying attention the video dialogues they watched, participants completed a questionnaire after each dialogue they watched. It consisted of five questions about the content contained in the dialogue. The threshold for inattentiveness was defined answering fewer that three of the five questions correctly. All participants answered at least three out of five questions correctly, so no participants were excluded from the study due to inattentiveness.

3.2 ERP DATA

3.2.1 Conversation Group

Visual inspection of the ERP scalp topographies as depicted in Figure 1 depicted a negative going potential in response to both low and high frequency nouns, largest in the central parietal electrode region during the 400 to 600 ms time window characteristic of the N400. The average waveform plot of electrodes in the central parietal region (Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, CP2) is shown in Figure 3. The choice of electrodes was selected a priori as a selection of electrodes in the central parietal region as this is the typical scalp distribution of the N400 ERP (Kutas & Federmeier, 2011). This demonstrated a negative potential that occurred in the 200-600 ms time window, again characteristic of the N400 both in timing and in showing a more negative amplitude to low frequency in comparison to high frequency nouns, as depicted in Figure 4.

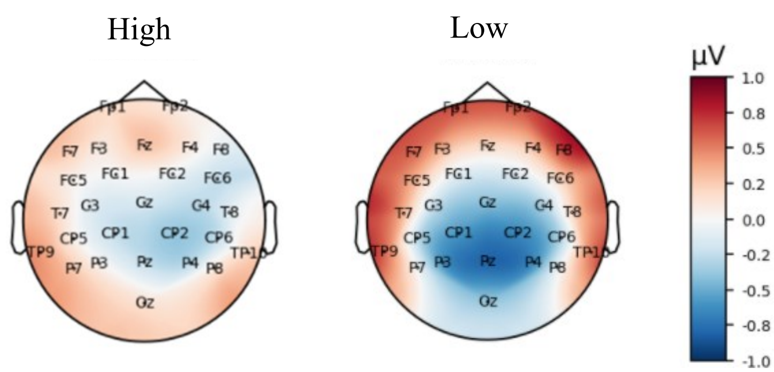

Figure 3: Scalp topographies averaged between the 400-600ms time window characteristic of the N400 when, participants in the conversation group hear either low (right plot) or high (left plot) frequency words. A negative potential occurs in the central parietal regions of both conditions and is greater in amplitude in the low frequency condition as compared to the high frequency condition.

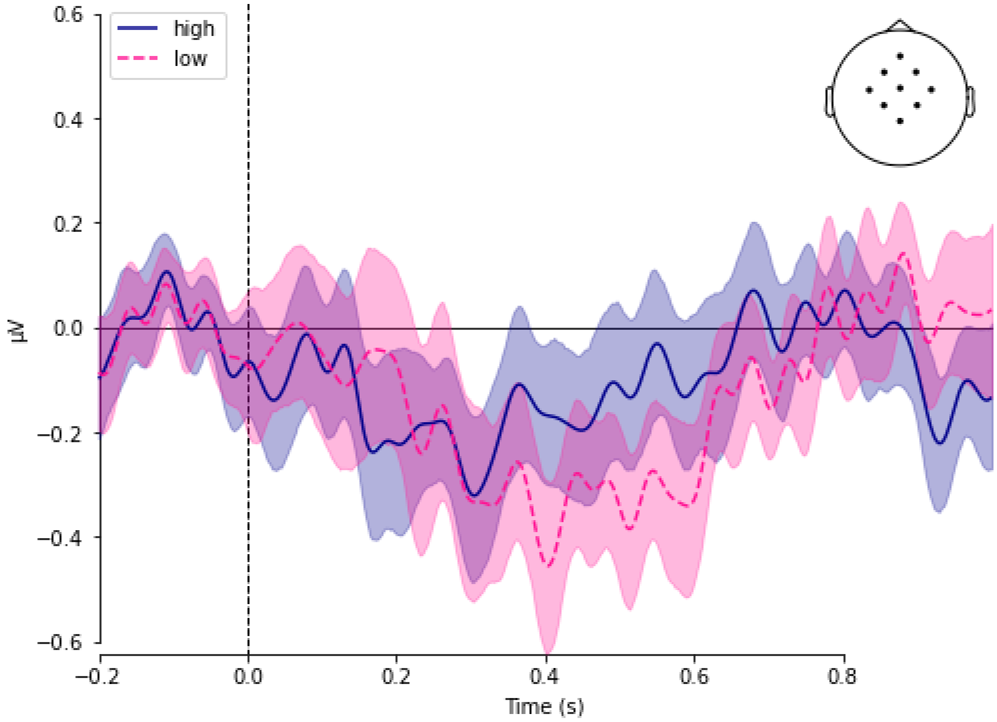

Figure 4:The ERP waveforms averaged over the central parietal region (electrodes Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, CP2) of the conversation group hearing either high frequency (in blue) or low frequency (in pink) nouns. The shaded areas represent the 95% confidence interval.

Prior to statistical analysis, the data was examined for outliers by applying a *z* transform to the data from all trials and channels, separately for each participant. Individual data points were removed if they exceeded of threshold of $z = \pm 2.50$. In total, 1.87% of the data was rejected as outliers.

Linear mixed effects (LME) modeling was performed on the mean amplitude of all electrodes during the 400-600 ms time window that was centered around the typical peak of the N400 ERP to analyze the amplitude and scalp distribution, as described above in Methods. The optimal model included the main effects of frequency and channel, the two-way interaction of frequency by channel and the subject, and channel by subject random intercepts and slope, as shown in Table 1. The factors of word frequency, channel, and the interaction of word frequency and channel were all significant.

Coefficient	**df**	***F***	***p*-value**
Frequency	1	11.54	<0.01
Channel	27	17.55	<0.01
Frequency X Channel	27	5.84	<0.01
Subject	34	0.00	1.00
Subject X Channel	952	0.39	<0.01

Table 1: ANOVA table of the optimal model for the N400 at the 400-600ms time window for the conversation group.

The effects of channel and the interaction of word frequency and channel further were further investigated by conducting post-hoc analyses at each electrode that examined the contrast of low and high frequency as shown in Table 2. When controlling for false discovery rate using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995), nine electrodes (Fp1, Fp2, P3, Pz, F8, FC6, T8, TP10, and Oz) had significant *p*-values as displayed in Table 2 and depicted in Figure 5.

Electrode	Estimated Difference (low-high)	*t*	*p* (FDR-BH)	Lower CL	Upper CL
Fp1	0.42	3.4	<0.01	0.05	0.18
Fp2	0.5	4.03	<0.01	0.07	0.2
F7	0.26	2.14	0.08	0.01	0.14
F3	0.18	1.49	0.19	-0.02	0.11
Fz	0	0.03	1	-0.06	0.07
F4	0.18	1.49	0.19	-0.02	0.11
F8	0.55	4.42	<0.01	0.08	0.21
FC5	0.24	1.95	0.08	0	0.13
FC1	-0.12	-1	0.42	-0.1	0.03
FC2	-0.07	-0.58	0.69	-0.08	0.05
FC6	0.42	3.4	0	0.05	0.17
T7	0.2	1.65	0.15	-0.01	0.12
C3	-0.1	-0.81	0.53	-0.09	0.04
Cz	-0.28	-2.3	0.06	-0.14	-0.01
C4	-0.04	-0.33	0.8	-0.07	0.05
T8	0.37	3	0.01	0.03	0.16
TP9	0	0	1	-0.07	0.07
CP5	-0.24	-1.95	0.08	-0.13	0
CP1	-0.27	-2.19	0.07	-0.14	-0.01
CP2	-0.24	-1.98	0.08	-0.13	0
CP6	-0.06	-0.52	0.7	-0.08	0.05
TP10	0.3	2.45	0.04	0.02	0.15
P7	-0.24	-1.97	0.08	-0.13	0
P3	-0.57	-4.62	<0.01	-0.22	-0.09
Pz	-0.43	-3.5	<0.01	-0.18	-0.05
P4	-0.24	-1.97	0.08	-0.13	0
P8	-0.06	-0.48	0.71	-0.08	0.05
Oz	-0.32	-2.58	0.03	-0.15	-0.02

Table 2*:* The estimated difference between low and high frequency noun conditions (low-high) at each electrode for the conversation group.

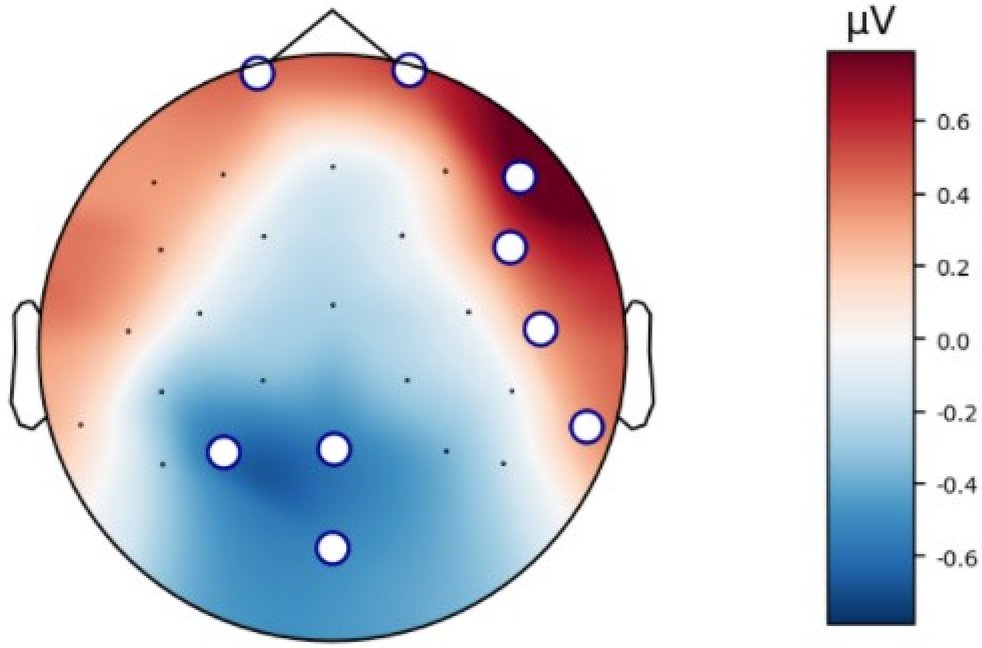

Figure 5: The contrast of the difference between the low and high frequency word conditions (low-high) during the 400-600 ms time window for the conversation group. Electrodes with significant *p*-values using the Benjamini-Hochberg false discovery rate correction procedure (Benjamini & Hochberg, 1995) are marked by white circles. The Fp1, Fp2, P3, Pz, F8, FC6, T8, TP10, and Oz electrodes had significant *p*-values; low frequency words elicited more negative potentials than high frequency at Pz, P3, and Oz.

<u>3.2.2</u> <u>Control Group</u>

Visual inspection of the ERP scalp topographies as shown in Figure 6 detected a negative-going potential in response to low and high frequency nouns, in the central parietal electrode region during the 400 to 600 ms time window characteristic of the N400 in the control group. An average waveform plot of electrodes in the central parietal region (Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, CP2) is shown in Figure 7. The choice of electrodes was selected a priori as a selection of electrodes in the central parietal region as this is the typical scalp distribution of the N400 ERP (Kutas & Federmeier, 2011).

This shows a negative potential from 200-600 ms characteristic of the N400, with a larger amplitude negative deflection occurring in response to low frequency nouns in comparison to high frequency nouns, as predicted.

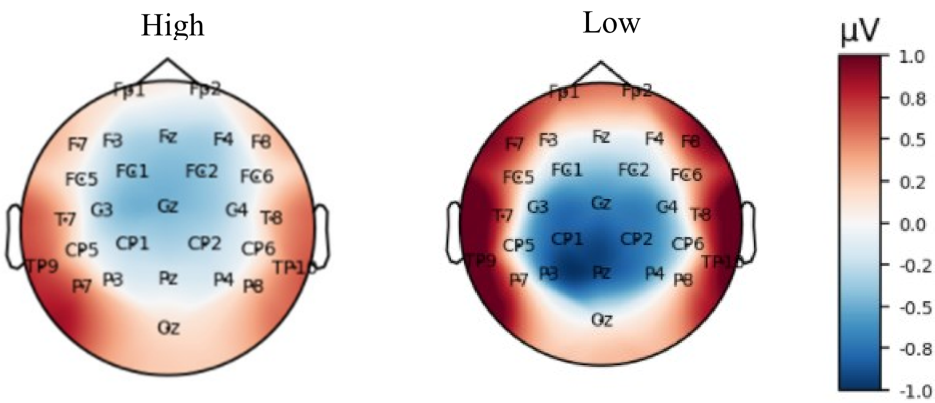

Figure 6: Scalp topographies averaged between the 400-600ms time window characteristic of the N400 when participants in the control group heard either low (right plot) or high (left plot) frequency words. A negative potential occurs in the central parietal regions of both conditions and is greater in amplitude in the low frequency condition as compared to the high frequency condition.

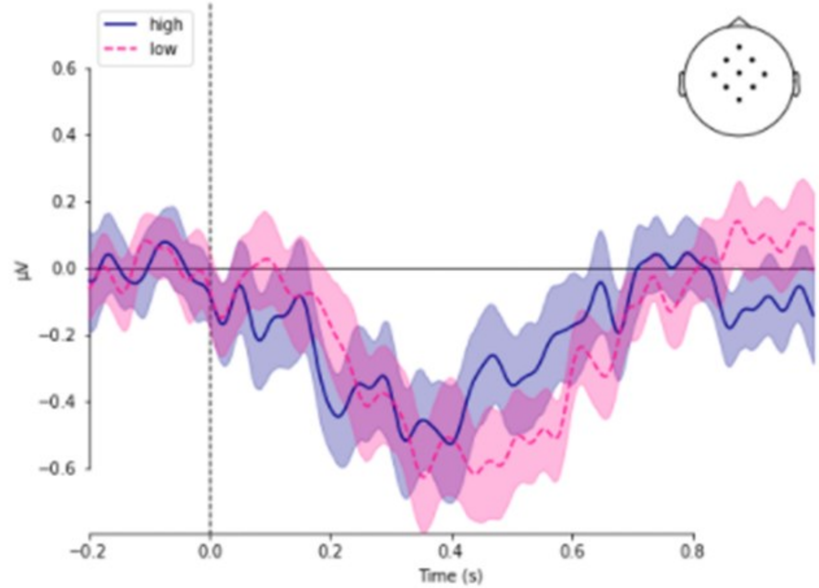

Figure 7: The ERP waveforms averaged over the central parietal region (electrodes Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, CP2) of the control group hearing either high frequency (in blue) or low frequency (in pink) nouns. The negative potential amplitude is greater across the 400-600ms time window in response to low frequency nouns as compared to high frequency nouns. The shaded areas represent the 95% confidence interval.

The data was examined for outliers. In total, 2.00% of the data was rejected due to meeting criteria as an outlier as it exceeded of threshold of $z = \pm 2.50$.

LME modeling was again performed on the mean amplitude of all electrodes during the 400-600 ms time window, as for the conversation group's data. The optimal model included the two-way interaction of frequency by channel and random intercepts for subject, and channel by subject random slopes, as shown in Table 3. The fixed effects of word frequency, channel, and the interaction of word frequency and channel were all significant.

Coefficient	**df**	**F**	***p*-value**
Frequency	1	23.20	<0.01
Channel	27	36.50	<0.01
Frequency X Channel	27	11.90	<0.01
Subject	39	0.00	1.00
Subject X Channel	1092	0.60	<0.01

Table 3: ANOVA table of the optimal model for the N400 at the 400-600ms time window for the control group.

To investigate the effects of channel and the interaction of word frequency and channel further, we conducted post-hoc analyses at each electrode that examined the contrast of low and high frequency (low-high contrast) as shown in Table 4. When controlling for false discovery rate using the Benjamini-Hochberg procedure, nineteen electrodes (Fp1, Fp2, F3, F7, FC5, T7, CP1, CP5, TP9, P3, Pz, F4, F8, FC6, T8, CP2, TP10, P4, and Oz) had significant p-values as displayed in Table 4 and depicted in Figure 8.

Electrode	Estimated Difference (Low-High)	*t*	*p* (FDR-BH)	Lower CL	Upper CL
Fp1	0.5	4.82	<0.01	0.09	0.2
Fp2	0.57	5.58	<0.01	0.11	0.22
F7	0.42	4.07	<0.01	0.06	0.18
F3	0.28	2.68	0.01	0.02	0.14
Fz	0.17	1.62	0.14	-0.01	0.11
F4	0.34	3.35	<0.01	0.04	0.16
F8	0.61	5.93	<0.01	0.12	0.24
FC5	0.31	3.01	0.01	0.03	0.15
FC1	0.03	0.32	0.78	-0.05	0.07
FC2	0.07	0.67	0.54	-0.04	0.08
FC6	0.32	3.14	<0.01	0.03	0.15
T7	0.25	2.39	0.03	0.01	0.13
C3	-0.18	-1.73	0.12	-0.11	0.01
Cz	-0.14	-1.32	0.22	-0.1	0.02
C4	-0.17	-1.7	0.12	-0.11	0.01
T8	0.27	2.65	0.02	0.02	0.14
TP9	0.27	2.55	0.02	0.02	0.14
CP5	-0.27	-2.59	0.02	-0.14	-0.02
CP1	-0.51	-4.97	<0.01	-0.21	-0.09
CP2	-0.41	-3.98	<0.01	-0.18	-0.06
CP6	-0.14	-1.41	0.19	-0.1	0.02
TP10	0.33	3.15	<0.01	0.04	0.16
P7	-0.09	-0.87	0.43	-0.09	0.03
P3	-0.69	-6.73	<0.01	-0.26	-0.14
Pz	-0.59	-5.7	<0.01	-0.23	-0.11
P4	-0.3	-2.93	0.01	-0.15	-0.03
P8	0	0.02	0.98	-0.06	0.06
Oz	-0.24	-2.31	0.03	-0.13	-0.01

Table 4: The estimated difference between low and high frequency noun conditions (low-high) at each electrode for the control group.

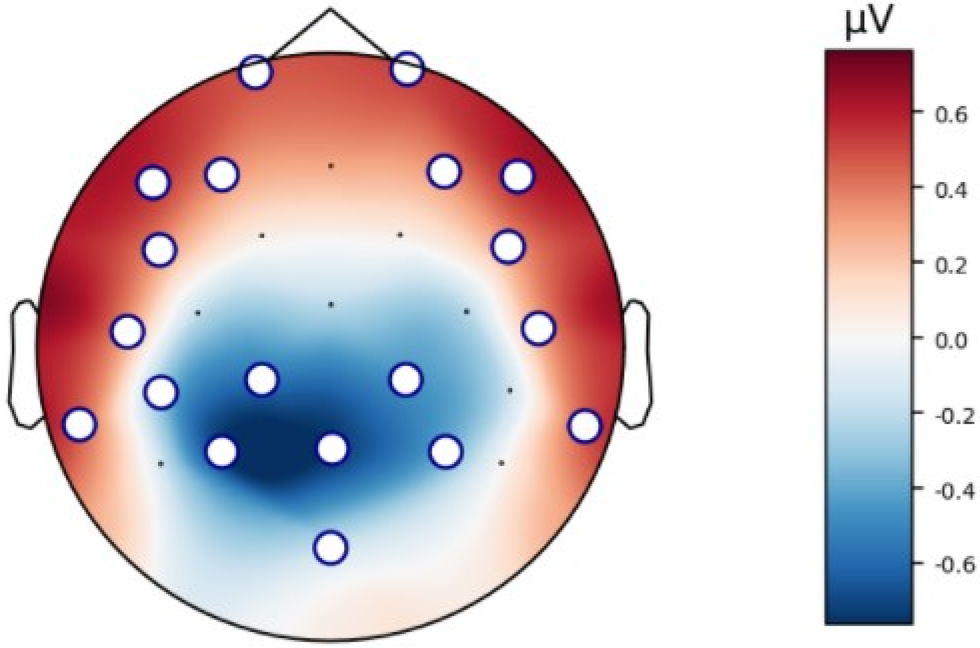

Figure 8: The contrast of the difference between the low and high frequency word conditions (low-high) during the 400-600ms time window for the control group. Electrodes with significant p-values using the Benjamini-Hochberg false discovery rate correction procedure (Benjamini & Hochberg, 1995) are indicated by white circles. The Fp1, Fp2, F3, F7, FC5, T7, CP1, CP5, TP9, P3, Pz, F4, F8, FC6, T8, CP2, TP10, P4, and Oz electrodes had significant p-values.

3.2.3 Conversation and Control Group Comparison

To examine the conversation and control groups results, LME modeling was performed on the mean amplitude of all electrodes during the 400-600 ms time window, as in the two previous analyses but now with both groups' data included and an additional fixed effect of group. The optimal model included the fixed effects of group, word frequency, and channel, the two-way interactions of word frequency X group, channel X group, and channel X frequency, the random intercept of subject, and the random slope of channel by subject, as shown in Table 5. The factors of word frequency, channel, and the interaction of channel by group and channel by frequency all had significant p-values. The factor of group and the interaction of word frequency by group were not significant. Figure 9 depicts the main effect of group, collapsed across frequency condition at each electrode. Figure 10 displays the main effect of frequency by group at each electrode.

Coefficient	**df**	***F***	***p*-value**
Group	1	0.31	0.58
Frequency	1	34.35	<0.01
Channel	27	39.87	<0.01
Group X Frequency	1	0.59	0.44
Group X Channel	27	4.02	<0.01
Frequency X Channel	27	16.72	<0.01
Subject	73	0.00	1.00
Subject X Channel	2044	0.49	<0.01
Group X Subject	73	0.00	1.00
Script X Word	230	0.00	1.00
Dyad X Participant Role	73	0.00	1.00

Table 5: ANOVA table of the optimal model for the N400 at the 400-600ms time window when comparing the conversation and control groups.

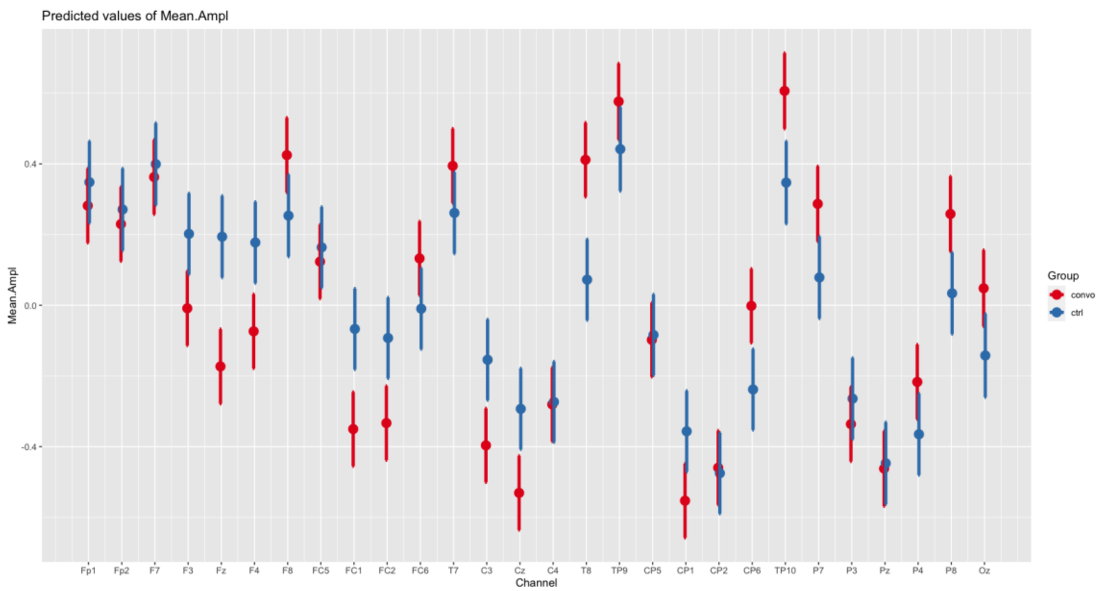

Figure 9: Model-estimated mean amplitudes for low and high frequency words across both groups, plotted for each electrode bars represent 95% confidence intervals.

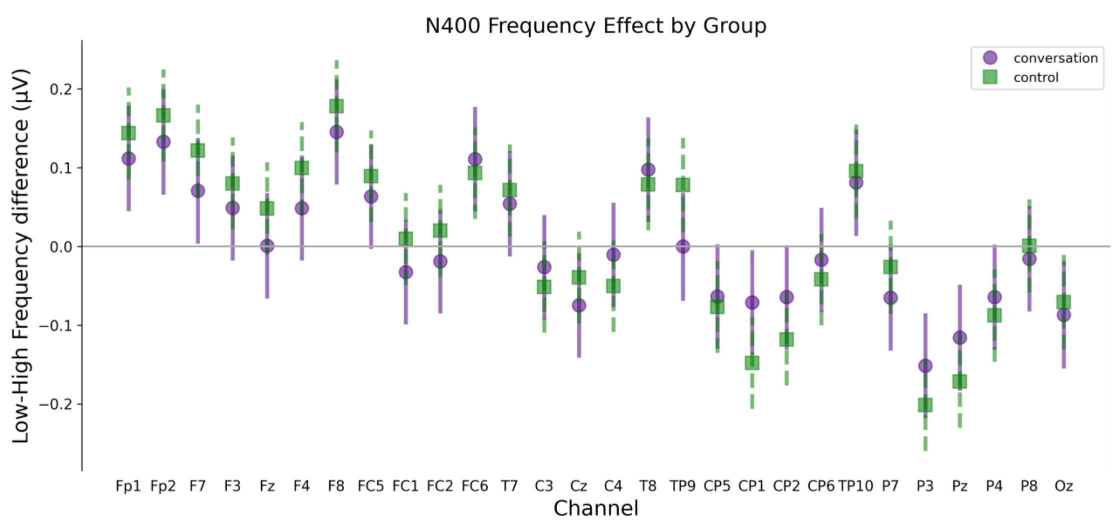

Figure 10: The model-estimated size of the difference between low and high frequency words (low-high) are plotted for each electrodes. Bars represent 95% confidence intervals. The conversation group is displayed by the solid purple markings and the control group is displayed by the dotted green markings.

The Group × Channel interaction indicates some difference in scalp distribution between groups, however critically there was no three-way interaction of Group, Channel, and Frequency, nor a Group × Frequency interaction. The absence of these interactions indicates that there was no difference in the size nor the scalp distribution of the N400 frequency effect between the two groups.

CHAPTER 4 DISCUSSION

4.1 GENERAL OVERVIEW

The aim of the thesis experiments was to determine if it was possible to replicate an established ERP effect using EEG hyperscanning methods during a conversational scenario. Specifically, the experiments sought to replicate a lexical frequency effect, where the N400 ERP amplitude is larger in response to low as compared to high lexical frequency words (Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber,Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017). To do this, a group of participant pairs completed a scripted conversation task, where participants had scripted conversations together that included a stimuli set of low and high frequency words inserted into the dialogue. A control group of participant pairs watched the same scripted dialogues together as passive observers. In addition to predicting that the lexical frequency effect would be replicated in both the conversation and control conditions, it was predicted that the ERP amplitudes would differ between the conversation and control conditions due to the difference in task (social interaction vs. passive viewing).

The effect of lexical frequency on the N400 ERP was replicated in both groups. This demonstrates that a) the EEG hyperscanning technique so far can be used to detect ERP effects that are in line with the prior literature and b) that semi-naturalistic testing paradigms (i.e. people having a conversation or watching a conversation) yield results that are in line with prior research. In addition, the N400 ERP lexical frequency effect was not modulated by task condition (i.e. having a conversation vs. watching a conversation). However, task condition modulated the N400 ERP when averaged across both frequency conditions (i.e. the average of low and high frequency conditions combined).

4.2 BEHAVIOURAL FINDINGS

A key component was developing an experimental task that resembled a conversation while maintaining control over the linguistic content participants were exposed to. As a result, participant pairs completed a scripted dialogue task where each received a script containing dialogue lines they were asked to read and cues of when they should listen to their partner as it was their partner's turn to speak. Each person acted as the other person's stimulus delivery system. As a result, if participants struggled with this task in terms of interrupting each other, failing to properly take turns reading their dialogue lines, or made speech production errors when reading the lines, the experiment and resulting analyses would be unfeasible. Participants were able to successfully complete the dialogue task with few errors. Overall, this indicates that participants had little trouble completing the task. Therefore, the dialogue task appears to be an appropriate method for simulating a conversational experience.

The control group was asked to observe a conversation by watching video recordings of two people having the same dialogues as used in the conversation group task. To verify that participants paid attention to the content of the dialogues, participants were asked to complete a short five question quiz after each dialogue. No participants were rejected due to poor performance on the comprehension questionnaire, indicating that they were paying attention to the task content. Therefore, the observation task appears to be an appropriate psycholinguistic stimulus presentation method.

4.3 ELECTROPHYSIOLOGICAL FINDINGS

4.3.1 Conversation Group

As depicted in Figures 3 and 4, there was a negative going potential that occurred in during the 200-600 ms time window across the central parietal electrode regions for both the low and high frequency noun conditions. These findings are in line with the typical presentation of the N400 ERP in response to hearing linguistic stimuli. Statistical analyses revealed a significant effect of frequency on N400 modulation, as well as a frequency × channel interaction which indicated that the frequency effect varied across scalp locations. Together, these are consistent with past reports of the N400 (Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber, Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017) insofar as amplitude in the N400 time window was significantly more negative for low than high frequency words over posterior midline electrode sites. Interestingly, however, there were some differences in the scalp distribution of the N400 relative to past studies, which have generally reported the largest negativity over the central parietal electrodes. Of the electrodes in the central parietal region predicted a priori to show the N400 (Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, CP2), only Pz exhibited a significantly more negative amplitude in response to low as compared to high frequency words. A similar effect was present at the other electrodes in the ROI, but the differences were not statistically significant. At the same time, the negativity was also significant at electrodes P3 and Oz, which are just outside our a priori ROI. As well, the frequency × channel interaction reflected several electrodes around anterior and very lateral locations where the difference was significant, but reversed: low frequency words elicited a more positive response than high frequency. These positive differences likely reflect the inverse part of the electrical source of the N400, since ERP effects are considered to be electrically dipolar, such that any negative effect should have a positive complement on the opposite side of the head.

In terms of why the N400 here had an unexpectedly posterior scalp distribution, all EEG components, including the N400 ERP are known to vary spatially because of individual differences in cortical anatomy and skull thickness. They can also vary spatially between studies due to differences in tasks and stimuli (Luck, 2014; Newman, 2019). Electrodes located at the posterior of the scalp are known to detect the N400 ERP over the temporal lobe and the N400 ERP can be detected at most electrode sites in some cases (Van Petten & Luca, 2006; Kutas & Federmeier, 2011). So, the significant difference detected between the low and high conditions at posterior electrodes Oz and P3 is in line with known N400 characteristics. Therefore, there is support for the hypothesis that the lexical frequency effect would be replicated in the conversation group.

4.3.2 Control Group

Both Figures 6 and 7 display a negative going potential that occurred in during the 200-600 ms time window across the central parietal electrode regions for both the low and high frequency noun conditions. The findings are in line with the typical presentation of the N400 ERP in response to hearing linguistic stimuli of varying lexical frequency (Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber, Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017) as the amplitude in the N400 time window was significantly more negative for low than high frequency words over posterior midline electrode sites. Specifically, there was both a significant effect of frequency on N400 modulation, and a frequency by channel interaction, indicating that the frequency effect varied across scalp locations. Similarly to the conversation group, there were some differences in the scalp distribution of the N400 relative to past studies, which have generally reported the largest negativity over the central parietal electrodes. Of the electrodes in the central parietal region predicted a priori to show the N400 (Fz, Cz, Pz, FC1, FC2, C3, C4, CP1, CP2), only electrodes CP1, CP2, and Pz all displayed a significant negative difference between the low and high frequency noun conditions. Again, a similar effect was present at the other electrodes in the a priori ROI, but the differences were not statistically significant. The posterior electrodes just outside of the ROI such as Oz, P3, and P4 also were significantly negatively different between the low and high frequency conditions, again corresponding with the conversation group results. The control group results also featured a frequency × channel interaction which was reflected in electrodes around anterior and very lateral locations where the difference was significant but reversed: low frequency words elicited a more positive response than high frequency. Both the positive differences in these regions and the unexpectedly posterior N400 scalp distribution are anticipated to be due to the same factors mentioned regarding the conversation results. Therefore, there is support for the hypothesis that the lexical frequency effect would be replicated in the control group. In addition, the control group results regarding the N400 lexical frequency effect, largely mirrored those of the conversation group.

4.3.3 Conversation and Control Group Comparison

We also hypothesized that there would be a significant difference between the ERP responses of the conversation group and the control group. The interaction of group and frequency was not significant, meaning that the act of having a conversation versus watching a conversation had no effect on a difference of response to words of different lexical frequencies. When the difference contrasts of the low and high frequency conditions were plotted for each channel and each group (see Figure 10), there were no statistically significant differences between the groups, indicating that there was no interaction of group, channel, and frequency. Overall, this means that the N400 lexical frequency effect and its scalp distribution did not differ between the conversation task and control group.

However, the interaction of group and channel was statistically significant, suggesting an overall difference in scalp distribution between groups and across both high and low frequency words. As shown in Figure 9, five of the electrodes in the central parietal ROI (Cz, Fz, C3, FC1, and FC2) had significant differences between the conversation and control group, where the conversation task elicited more negative amplitudes than the control task. A further limitation to drawing conclusions regarding the group and channel interaction is that the EEG data was referenced using an average reference method (referencing to the average of all electrodes). While the average reference method is known to preserve the timing and scalp topography of ERPs, it has been found to alter the significant difference level between task-related effects and at a channel level (Yang, Fan, Wang, Li, 2017), which could foreseeably be at play for the task-level (i.e. conversation task vs. control observation task) differences in the thesis results. Therefore, while the N400 lexical frequency effect did not differ between groups, there was a significant effect of task condition of scalp distribution of the N400 component when averaged across word frequency conditions. This provides some support for the last hypothesis, that the N400 response would be modulated by conversational interaction versus conversational observation. However, with the considerable limitations at play in this circumstance, the task-related differences should be viewed as an interesting finding that requires much more research to elucidate.

4.4 CONCLUSIONS AND FUTURE RECOMMENDATIONS

The main finding of the thesis is that it is possible to detect an N400 ERP effect in response to naturalistically presented linguistic stimuli that occur in conversational scenarios. Furthermore, the results of two methods of presenting linguistic stimuli naturalistically (i.e. by having participants observe a conversation or have a scripted dialogue) are in line with prior N400 ERP research. Specifically, we also found support for the prediction that there would be significantly a greater amplitude of N400 ERP response to low frequency nouns as compared to high frequency nouns.

Overall, the observation of an N400 effect under more naturalistic testing conditions is in line with other ERP studies that used semi-naturalistic testing paradigms and found similar results to more traditional and less-naturalistic settings. A study that asked participants to listen to naturalistic speech interactions of a mother and her infant detected ERPs in both adult and infant participants (Cass-Lam et al., 2021). Fjaellingsdal and colleagues (2020) also demonstrated that ERP responses (including the N400 ERP) occurred in participants equipped with a non-hyperscanning EEG setup who engaged in a spontaneous word-by-word sentence construction task with a non-EEG equipped confederate. Our study takes this a step further by providing evidence that people participating in a task where they alternate between speaker and listener conversational roles exhibit N400 ERPs in response to target nouns.

There was a minor difference between the thesis results and the N400 ERP literature- the N400 ERP responses in both groups were more posteriorly located than anticipated. This could be due a few different reasons. First, ERP component scalp distribution is known to vary due to individual differences (both anatomically and neurologically) and varies between studies and tasks (Luck, 2014; Newman, 2019). Second, differences in reference selection can influence scalp distribution of ERP components (Yang, Fan, Wang, Li, 2017, Luck, 2014; Newman, 2019). As the participants in one group had a conversation, the mastoids (another common reference point) were not an appropriate choice for re-referencing as they were prone to more muscle noise from the speech, whereas most N400 ERP studies either reference to a built-in reference electrode or the mastoid electrodes (Rugg, 1990; Van Petten & Kutas, 1990; Hauk & Pulvermüller, 2004; Barber, Vergara, & Carreiras, 2004; Payne, Lee, & Federmeier, 2015;Vergara-Martínez, Comesaña, & Perea, 2017).

Although these are the most likely reasons for the more posteriorly located N400 ERPs, it should be mentioned that there is some debate around characteristics of the N400 ERP component. Specifically, there is debate about whether the FN400 component (a more anteriorly located component) is a different and distinct ERP component from the N400 that characterizes different cognitive processes. A negative component occurs at about 400 ms that is known to be modulated by recency and familiarity as newly presented words elicit a greater amplitude of response in comparison to words that are repeated several times throughout an experiment that is strongest in central frontal electrodes (Curran, 1999). Predominant psycholinguistic researchers like Marta Kutas and Kara Federmeier believe that this negative potential is a variation of the N400 effect and use it as example of how N400 scalp topography is variable (Voss & Federmeier, 2011; Kutas & Federmeier, 2011). Other researchers posit that the FN400 component is a different and distinct component from the N400 because the N400 component primarily indicates cognitive processes related to linguistic semantics and lexical properties, whereas the FN400 characterizes more generalized long-term memory retrieval processes

(Stróżak, Leynes, Wojtasiński, 2021; Stróżak, Abedzadeh, & Curran, 2016; Bridger, Bader, & Mecklinger, 2014; Bridger et al., 2012). In the thesis experiment, both groups displayed a more posteriorly located N400 effect than anticipated. Furthermore, there were significant differences between task condition in the central parietal electrodes, a region where the N400 component is typically largest. If a body of research continues to accumulate that supports the notion that differences in ERP scalp distribution can characterize specific different sets of cognitive processes, then it would be worthwhile to investigate if a similar effect (i.e. of a consistently posterior distribution of the N400 ERP) exists regarding a) language tasks involving a social component, b) type of social role/participatory involvement during a task, and c) naturalistic versus highly controlled experimental tasks.

It should be noted that there are a few limitations to the frequency conditions. The frequency effect is known to predominantly occur with words with extremely low frequency scores (Van Petten & Kutas, 1990) (Rugg, 1990). The thesis experiments (and most importantly the stimulus set scripts) started as a series of pilot experiments to determine the feasibility of EEG hyperscanning experiments as part of an undergraduate research project. Due to the considerable challenges of muscle noise from speech and social interaction, using a single-EEG system split between two individuals, and the lack of a reference electrode, it was not anticipated that the experimental data would yield usable EEG data. Another experiment that controlled for these factors and used more robust psycholinguistic measures was cancelled due to the Covid-19 pandemic. When designing the stimulus scripts, we used a wider low frequency range (and somewhat arbitrary) than many studies because if words were too infrequently used, our pilot trials found that participants believed these were non-sense words (e.g. kestrel). For context, in many studies the definition of "low frequency" is exceedingly low and will range from 1-99 (> log 0 to < log 2; note some studies report words than occur less than once (e.g. 0.25 occurrences per million)) occurrences per million for low frequency words and high frequency range words with 700-800 (~ log 2.8 – log 2.9) occurrence per million (Kutas & Van Petten, 1990; Rugg, 1990; Embick et al., 2001). For this study, low frequency words were any words below 1500 (< log 3.2) occurrences per million and high frequency words were words with greater than 1800 (> log 3.2) occurrences per million. As a result, the frequency categories used in these experiments are not in line with much of the past literature. That being said, studies have used larger corpuses and a larger range of word frequencies and detected a frequency effect. Dambacher and colleagues (2006) detected a frequency effect with stimuli ranging between greater than log 0 and log 4, with the greatest differences being in the range of words with frequency scores in the range of about > log 0 to log 2 and words with frequency scores of about log 2.5 and higher. A study that examined the N400 ERP during naturalistic listening to a recording of a story found an inverse frequency effect, insofar as N400 amplitude became more negative with increasing word frequency when using words that ranged in frequency from > log 0 to log 30 (Alday, Schlesewsky, & Bornkessel-Schlesewsky, 2017). One criticism of studies on the N400 lexical frequency effect is that studies use often stimuli with exceedingly low frequencies (sometimes even less than 1 occurrence per million)

(Hauk & Pulvermuller, 2004). Further, the lexical frequency effect has been thought to be non-linear as the difference is only detectable between words with exceedingly low frequencies with less than 30 occurrences per million and words of with more than 30 occurrences per million (Van Petten & Kutas, 1990). But, this non-linear effect of frequency may only exist in studies that use more constrained lexical frequency ranges as a regression analyses by Van Petten (2014) determined that lexical frequency as a continuous variable (ranging from 0.8 to 1350 occurrences per million) was a significant predictor of variance of the N400 ERP. Van Petten's findings (2014) considered alongside Dambacher and colleagues (2006) and Alday and colleagues (2017) findings, suggest that a more robust lexical frequency effect may exist when considering a greater range and distribution of word frequencies.

As well, in attempts to keep the dialogue as naturalistic as possible, we did not control for sentence position. The lexical frequency effect is greatest for words occurring earlier in sentences, and dissipates at the end of sentences (Kutas & Van Petten, 1990). Although outside the scope of this thesis, if we reanalyzed the data with sentence position as a factor, we would predict to only see words towards the beginning and middle of sentences to elicit the N400 ERP effects. However, the sentence position modulation of the N400 frequency effect is thought to be closely tied to context and predictability (Kutas & Van Petten, 1990; Dambacher et al., 2006). The sentences within the dialogue have not been rated for predictability or related scores, so these factors are not known. As the scripts were constructed to contain a general plot and a variety of unusual words, they may not be as predictable as other psycholinguistic stimuli. For example, one example would be the sentence: “Thanks for putting the decorative *gourd* out. It really helped me find you.”, where gourd is the target word. Or in response to the question “When do you want to stop and set up camp?”, the other participant responds with the line “We can stop soon. I’m glad we have a *tent*.”, where the target word is tent. Although these words are not semantic anomalies, they may not necessarily be the most predicted conclusions to these sentences (or combinations of sentences). Conducting a study on completion norms of the stimuli sentence alongside an analysis of sentence position would help identify if either played a role in N400 ERP modulation (or lack thereof) in these experiments.

In terms of future directions, first, we hope that the demonstration that it is possible to detect the N400 ERP in conversational social interactions will open the door to further psycholinguistic research that implements naturalistic testing paradigms. An obvious next step is to attempt to replicate these findings by first, running studies that use established low and high lexical frequency categories and control for sentence position. Second, would be to run studies using measures other than lexical frequency that modulate the N400 ERP, like close probability and surprisal. And finally, to attempt to run studies that examine ERP responses during less controlled conversation (i.e. asking participants to self-generate words missing within a dialogue or on specific topics) or even free conversation.

REFERENCES:

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle", in Petrov, B. N.; Csáki, F. (eds.), *2nd International Symposium on Information Theory*, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akadémiai Kiadó, pp. 267–281. Republished in Kotz, S.; Johnson, N. L., eds. (1992), B*reakthroughs in Statistics*, I, Springer-Verlag, pp. 610–624.

Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2017). Electrophysiology Reveals the Neural Dynamics of Naturalistic Auditory Language Processing: Event-Related Potentials Reflect Continuous Model Updates. *eNeuro*, *4*(6), ENEURO.0311-16.2017. https://doi.org/10.1523/ENEURO.0311-16.2017.

Babiloni, F., Cincotti, F., Mattia, D., Mattiocco, M., De Vico Fallani, F., Tocci, A., Bianchi, L., Marciani, M. G., & Astolfi, L. (2006). Hypermethods for EEG hyperscanning. *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2006*, 3666–3669.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior research methods*, *39*(3), 445–459. https://doi.org/10.3758/bf03193014.

Barber, H. A., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: Evidence from ERPs. *Neuroreport, 15,* 545–548. doi:10.1097/00001756-200403010-00032.

Barraza, P., Dumas, G., Liu, H., Blanco-Gomez, G., van den Heuvel, M.I., Baart, M., & Pérez, A. (2019). Implementing EEG hyperscanning setups. *MethodsX*, *26*(6):428-436. doi: 10.1016/j.mex.2019.02.021.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Brennan, J. (2016). Naturalistic Sentence Comprehension in the Brain. Language and Linguistics Compass, 10(7), 299-313.

Boersma, P. & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.51, retrieved 22 July 2021 from http://www.praat.org/.

Cespón, J., & Carreiras, M. (2020). Is there electrophysiological evidence for a bilingual advantage in neural processes related to executive functions? *Neuroscience and Biobehavioral Reviews, 118*, 315-330.

Chauncey, K., Holcomb, P. J., & Grainger, J. (2008). Effects of stimulus font and size on masked repetition priming: An event-related potentials (ERP) investigation. *Language and cognitive processes*, *23*(1), 183–200. https://doi.org/10.1080/01690960701579839.

Cohen, Mike X. (2014). *Analyzing Neural Time Series Data*. Cambridge: MIT Press.

Collura, T. (1993). History and Evolution of Electroencephalographic Instruments and Techniques. *Journal of Clinical Neurophysiology, 10*(4), 476-504.

Czeszumski, A., Eustergerling S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z.Z., and König, P. (2020). Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Frontiers in Human Neuroscience*, *14*(39). doi: 10.3389/fnhum.2020.00039.

Davies, Mark. (2008) *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Debruille, J. B., Brodeur, M. B., & Franco Porras, C. (2012). N300 and social affordances: a study with a real person and a dummy as stimuli. *PloS one*, *7*(10), e47922. https://doi.org/10.1371/journal.pone.0047922.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain research*, *1084*(1), 89–103. https://doi.org/10.1016/j.brainres.2006.02.010.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Galelese, L, & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91***,** 176–180.

Duane, T. D., & Behrendt, T. (1965). Extrasensory electroencephalographic induction between identical twins. *Science, 150*(3694), 367. https://doi.org/10.1126/science.150.3694.367.

Dumas G, Nadel J, Soussignan R, Martinerie J, Garnero L (2010). Inter-Brain Synchronization during Social Interaction. *PLoS ONE 5*(8): e12166. https://doi.org/10.1371/journal.pone.0012166

Fischler, I., Jin, Y., Boaz, T., Perry, N., & Childers, D. (1987). Brain potentials related to seeing one's own name. *Brain & Language, 30*(2), 245-262.

Fjaellingsdal, T., Schwenke, D., Scherbaum, S., Khulen, A., Bögels, S., Meekes, J., & Bleichner, M. (2020). Expectancy effects in the EEG during joint and spontaneous word-by-word sentence production in German. *Scientific Reports, 10*, 5460. https://doi.org/10.1038/s41598-020-62155-z.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience*, *7*, 267. https://doi.org/10.3389/fnins.2013.00267.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67-81.

Guajardo, L. F., & Wicha, N. Y. (2014). Morphosyntax can modulate the N400 component: event related potentials to gender-marked post-nominal adjectives. *NeuroImage*, *91*, 262–272. https://doi.org/10.1016/j.neuroimage.2013.09.077.

Hari, R., Henriksson, L., Malinen, S. & Parkkonen, L. (2015). Centrality of Social Interaction in Human Brain Function. *Neuron, 88*(1), 181-193.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology, 115*(5), 1090–1103. https://doi.org/10.1016/j.clinph.2003.12.020.

Hayward, R. (2001). The tortoise and the love-machine: grey walter and the politics of electroencephalography. *Sci. Context* 14, 615–641. doi: 10.1017/S0269889701000278.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology, 19*(4), 286–300.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks, 10(3), 626-634. doi: 10.1109/72.761722.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., & Feldman, R. (2017). Brain-to-Brain Synchrony during Naturalistic Social Interactions. *Nature Scientific reports*, *7*(1), 17060. https://doi.org/10.1038/s41598-017-17339-5.

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction?. *Frontiers in human neuroscience*, *6*, 215.

Kroll, J., Dussias, P., Bice, K., & Perrotti, L. (2015). Bilingualism, Mind, and Brain. *Annual Review of Linguistics, 1*(1), 377-394.

Kuhlen, A., Allefeld, C., & Haynes, J. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience, 6*, 266.

Kuhlen, A., Allefeld, C., Anders, S., & Haynes, J. (2015). *Towards a multi-brain perspective on communication in dialogue.* In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use* (p. 182–200). Cambridge, United Kingdom: Cambridge University Press.

Kutas, M., & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*(1), 621-647.

Kutas, M., & Hillyard, S. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science (American Association for the Advancement of Science), 207*(4427), 203-205.

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, *57*(7), e13351. https://doi.org/10.1111/psyp.13351.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin, 144*(4), 394–425.

Linck, J., Hoshino, N., & Kroll, J. (2008). Cross-language lexical processes and inhibitory control. *The mental lexicon*, *3*(3), 349–374.

Linderberger, U., Li, SC., Gruber, W., & Müller, V. (2009).Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neuroscience, 10*(22). https://doi.org/10.1186/1471-2202-10-22

Liu, D., Liu, S., Liu, X., Zhang, C., Li, A., Jin, C., Chen, Y., Wang, H., & Zhang, X. (2018). Interactive Brain Activity: Review and Progress on EEG-Based Hyperscanning in Social Interactions. *Frontiers in psychology*, *9*, 1862.

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of cognitive neuroscience*, *25*(7), 1049–1061. https://doi.org/10.1162/jocn_a_00388.

Lopes da Silva, F. (2013). EEG and MEG: Relevance to Neuroscience, *Neuron, 80*(5), 1112-1128.

Luck, S. (2014). *An introduction to the event-related potential technique (Second ed.).* Cambridge, Massachusetts: The MIT Press.

Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2015). Brain responses to world knowledge violations: A comparison of stimulus- and fixation-triggered event-related potentials and neural oscillations. *Journal of Cognitive Neuroscience, 27*(5), 1017–1028. https://doi.org/10.1162/jocn_a_00731.

Muise-Hennessey, A., Tremblay, A., Tagarelli, K., Petley, L., and Newman, A.J. (2016). The relationship between statistical learning and speech perception under noisy listening conditions. Poster presented at the 2016 Organization for Human Brain Mapping Annual Meeting, June 2016, Geneva, Switzerland.

Müller, H., & Kutas, M. (1996). What's in a name? Electrophysiological differences between spoken nouns, proper names and one's own name. *Neuroreport, 8*(1), 221-225.

Newman, A.J. (2019). Research Methods for Cognitive Neuroscience. London: SAGE.

Nguyen, T., Schleihauf, H., Kayhan, E., Matthes, D., Vrtička, P., & Hoehl, S. (2021). Neural synchrony in mother–child conversation: Exploring the role of conversation patterns, Social Cognitive and Affective Neuroscience, *16*(1-2), 93–102. https://doi.org/10.1093/scan/nsaa079.

Nieuwland, M., & Van Berkum, J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience*, *18*(7), 1098–1111.

Ning, S., Hayakawa, S., Bartolotti, J., & Marian, V. (2020). On language and thought: Bilingual experience influences semantic associations. *Journal of Neurolinguistics, 56*, 100932.

O'Connell, D., & Kowal, C. (2011). Sources of History for "A Psychology of Verbal Communication". *Journal of Psycholinguistic Research*, *40*(1), 29-47.

Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, *52*(11), 1456–1469. https://doi.org/10.1111/psyp.12515.

Rugg, M.D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition, 18***,** 367–379. https://doi.org/10.3758/BF03197126.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, *36*(4), 393-414.

Schoot, L., Hagoort, P., & Segaert, K. (2016). What can we learn from a two-brain approach to verbal interaction? *Neuroscience and Biobehavioral Reviews, 68*, 454-459.

Tardif, A., Chau-Morris, A., Wang, Z. Y., Takahara, E., Hadjis, T., Debruille, J., & Debruille, J. B. (2018). How to Find Effects of Stimulus Processing on Event Related Brain Potentials of Close Others when Hyperscanning Partners. *Journal of visualized experiments: JoVE*, (135), 56120. https://doi.org/10.3791/56120.

Van Petten C. (2014). Examining the N400 semantic context effect item-by-item: relationship to corpus-based measures of word co-occurrence. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *94*(3), 407–419. https://doi.org/10.1016/j.ijpsycho.2014.10.012.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18*(4), 380–393. https://doi.org/10.3758/BF03197127.

Vergara-Martínez, M., Comesaña, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, affective & behavioral neuroscience*, *17*(3), 461–474. https://doi.org/10.3758/s13415-016-0491-7.

Virtanen, P., Gommers, R., Oliphant, T.E. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17,** 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*(1), 192–196. https://doi.org/10.3758/bf03206482.

White, K. R., Crites, S. L., Jr, Taylor, J. H., & Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social cognitive and affective neuroscience*, *4*(2), 191–198. https://doi.org/10.1093/scan/nsp004.

Wood, S.N. (2011), Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73: 3-36. https://doi.org/10.1111/j.1467-9868.2010.00749.

Woodman G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, perception & psychophysics*, *72*(8), 2031–2046. https://doi.org/10.3758/APP.72.8.2031.

Woods, D. (1990). *The physiological basis of selective attention: Implications of event-related potential studies.* In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (p. 178–209). Oxford, United Kingdom: Oxford University Press.

Yang, P., Fan, C., Wang, M., & Li, L. (2017). A Comparative Study of Average, Linked Mastoid, and REST References for ERP Components Acquired during fMRI. *Frontiers in neuroscience*, *11*, 247. https://doi.org/10.3389/fnins.2017.00247.

APPENDIX A: A SAMPLE OF A SINGLE PARTICIPANT'S SCRIPT

Script 1: Garden Script

P1: Hi! It's great to see you. Thanks for coming to help out.
P2:
P1: Do you think it goes with the flowers I have out here?
P2:
P1: I also have tulips planted but they aren't growing right now. You'll have to come and see them another time.
P2:
P1: Follow me! It's right behind my house.
P2:
P1: Yes, unfortunately I only have one tree. I wish it was more sheltered. I'm planning to plant a sapling.
P2:
P1: Yes, I know. I had one a few years back but it withered away until it only had one leaf left. Ok, so today our plan is to get the plot ready to seed. Did you bring your pitchfork? I thought you had it with you when you came in?
P2:
P1: Great, thank you. Honestly, I don't need much help in term of gardening. My main problem is that I only own one trowel so I need to borrow everything else.
P2:
P1: Yeah, the neighbourhood cat kept digging everything up. Last year, it created so much damage that I only had a carrot that grew successfully so this time I planted something to distract it.
P2:
P1: Yeah, I have big plans this year. I'm hoping to be able to plant some beetroot and chard because those are my kid's favorites.
P2:
P1: Well, in the corner over there I'm going to put some berries.
P2:
P1: No. Why are they a favorite of yours?
P2:
P1: Yes, that sounds delicious. So today, we're just going to add in some new dirt. Is your friend still able to help us with the next stage tomorrow.
P2:
P1: Great, but why does he own one of those? That's a little unusual.
P2:
P1: That makes sense. Well, adding the furrows will make planting much easier.
P2:
P1: Maybe. Does he have turnips? I would also love to plant some more lettuce. Do you think they will grow well together?
P2:

P1: I thought that since your wife ran a gardening business that you would know this stuff.
P2:
P1: But you have that beautiful growing trellis in your yard that your wife said you were responsible for.
P2:
P1: Wait, what about the kiwi you gave me last year that you said you grew?
P2:
P1: I can't believe that I asked you for help. I should have asked your wife instead.