WORD FREQUENCY EFFECTS IN L2 SPEAKERS: AN ERP STUDY

by

Akinjide A. Famoyegun

A Thesis Submitted to the Faculty of the
SPEECH, LANGUAGE & HEARING SCIENCES DEPARTMENT
In Partial Fulfillment of the Requirements
For the Degree of
MASTER OF SCIENCE
In the Graduate College
THE UNIVERSITY OF ARIZONA

2012
STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Akinjide Famoyegun

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

_________________________          May 8, 2012
Gayle DeDe
Date
Assistant Professor of Speech, Language & Hearing Sciences
# TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................. 4
LIST OF FIGURES ................................................................................................................ 5
ABSTRACT ............................................................................................................................ 6
1. INTRODUCTION ............................................................................................................... 7
   1a. ERP ............................................................................................................................. 12
   1b. N400 ........................................................................................................................ 16
   1c. Word frequency effects ............................................................................................. 18
   1d. Models of L2 word recognition ............................................................................... 27
   1e. Present Study ............................................................................................................ 30
2. METHODS ......................................................................................................................... 32
   2a. Participants .................................................................................................................. 32
   2b. Stimuli ......................................................................................................................... 32
   2c. Procedure ..................................................................................................................... 34
   2d. Individual Differences ............................................................................................... 36
   2e. ERP Data Collection ................................................................................................. 37
   2f. Analysis ......................................................................................................................... 39
3. RESULTS .......................................................................................................................... 42
   3a. ERP Results ............................................................................................................... 42
   3b. Correlation Results ................................................................................................... 48
4. DISCUSSION ...................................................................................................................... 52
APPENDIX A ......................................................................................................................... 60
REFERENCES ......................................................................................................................... 62
LIST OF TABLES

Table 1 ......................................................................................................................... 43
Table 2 ......................................................................................................................... 45
Table 3 ......................................................................................................................... 49
Table 4 ......................................................................................................................... 50
LIST OF FIGURES

Figure 1 ………………………………………………………………………………………………………………………………. 21
Figure 2 ………………………………………………………………………………………………………………………… 28
Figure 3 ………………………………………………………………………………………………………………………… 38
Figure 4 ………………………………………………………………………………………………………………………… 41
Figure 5 ………………………………………………………………………………………………………………………… 46
Figure 6 ………………………………………………………………………………………………………………………… 47
Figure 7 ………………………………………………………………………………………………………………………… 50
ABSTRACT

The brain’s neural responses to words of different frequencies provide information on lexical organization and the cognitive processes involved in word identification and retrieval of meaning. Monolingual research has shown that exposure to high frequency words yields less cognitive difficulty than low frequency words as demonstrated by smaller N400 waves within even-related potential (ERP) methodology. The purpose of the present study was to compare frequency effects in adult native (L1) and non-native (L2) speakers of English during a sentence reading task embedded with high and low frequency word-pairs. Both L1 and L2 groups produced N400 waves of larger amplitudes for high frequency words compared to low frequency words that peaked around the 400 ms time mark. Group comparison found no significant difference in N400 wave amplitude and peak latency between both groups. The results are discussed with respect to theories of L2 word learning and lexical organization.
1. INTRODUCTION

One aim of second language acquisition research has been to understand the unique features of word processing in an individual’s non-native language. Whether achieved naturally during childhood or explicitly in later years, second language (L2) learning provides the opportunity to examine the underlying basis of comprehension difficulty for late learning L2 speakers in order to reveal when and why comprehension breakdowns occur. This has a practical application to current language learning programs because possible teaching interventions can be created in order to alleviate such receptive language problems for L2 speakers and optimize their formal language-learning experience. For example, learning an L2 that utilizes a different orthographic system than one’s native language may present difficulty in both reading and writing as the individual has to be attuned to a rather foreign written system composed of unfamiliar symbols. Additionally, learning an L2 in a region in which the L2 is not commonly spoken will limit L2 input to the classroom environment. The input from the classroom will represent a large majority of an individual's exposure to that language and consequently be their only opportunity to develop receptive language skills. These previously mentioned problems and others have an effect on comprehension from the word to the discourse levels. Information gained from L2 studies is being used to further the knowledge in bilingual language development to bridge the gap between theory and pedagogical practice.

For a majority of adult L2 learners, achieving an advanced level of proficiency
is an active process often filled with comprehension breakdowns. At times, these breakdowns occur because of a lack of semantic (vocabulary) knowledge as unrecognized L2 words prevent the L2 speaker from comprehending the complete message. Although this phenomenon is not unique solely to L2 speakers, it warrants considerable attention because semantic knowledge is such an integral part of communication in general and especially in the L2. Difficulty eliciting meaning results in global errors that hinder successful communication. Since most introductory second-language classroom instruction focuses on rote word-learning, it is helpful to correlate semantic characteristics with underlying cognitive processes in order to gain insight into how L2 speakers process words and their meanings in the L2. Throughout the second-language literature, word recognition in L2 speakers has been investigated to explore how semantic knowledge affects language processing (e.g., Batterink & Neville, 2011; Kutas & Federmeier, 2011; van Heuven & Dijkstra, 2010; Kroll, Sumutka & Schwartz, 2005; Mueller, 2005; Larsen-Freeman, 2002; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008).

In order to analyze semantic knowledge, studies have used visual reading tasks to investigate the cognitive mechanisms involved in word recognition. During reading, semantic, orthographic, morphological and syntactic knowledge may all contribute to the recognition process (Nation, 2009; Vigneau et al., 2006). For native (L1) speakers, word recognition requires a search in one mental lexicon (Murray & Forster, 2004). However, for non-native (L2) speakers, additional or altogether different mechanisms may play a role in correctly searching for and identifying
words and accessing their meanings (Duyck et al., 2008; van Heuven & Dijkstra, 2010). How L2 speakers retrieve a target word’s meaning has been the subject of debate in research because of different views on the functional organization of two lexicons in one brain (Mueller, 2005; van Heuven & Dijkstra, 2010). It is unclear whether L2 speakers separate their L1 and L2 lexicon into structured divisions such as lists or if both lexicons intermingle without division. In addition, certain lexical factors such as homophony and multiple meanings influence semantic processing. These factors provide complex perceptual analyses that may have different impacts on accessing meaning in the L2. This thesis focuses on the influence of one such factor, word frequency, in L2 processing of whole sentences.

In the monolingual literature, there is a lot of evidence that lexical retrieval is slowed by relatively complex stimuli. Complexity can be defined as variables that can make it more difficult to access and interpret a word. Polysemous words, for example, are words that have more than one related sense and can elicit several definitions if the comprehender is not given the proper semantic context. Words like hot can be defined as: a) having a relatively high temperature; b) violent, stormy; c) attractive; d) emotionally exciting; e) newly made. When presented without adequate context, the lexical retrieval process of the word hot will leave an individual with several possible interpretations. Previous studies in the monolingual literature have examined behavioral responses during reading and the results have suggested possible cognitive mechanisms to explain observed behaviors such as eye blinks and increased reaction times (Forster & Chambers, 1973; Howes & Solomon,
1951; Murray and Forster, 2004; Juhasz et al., 2006). Some of the cognitive mechanisms include a serial search in the mental lexicon, detecting familiar letter combinations and a systematic identification-recognition process.

Especially robust in the monolingual literature are word frequency effects in semantic processing. Word frequency is defined as the number of times a word is used in one’s language. Words that are more familiar (e.g., boat) are known as high frequency and less familiar words (e.g., hovercraft) are known as low frequency. Previous studies have found that the meaning of high frequency words are accessed faster than their low frequency counterparts (Howes & Solomon, 1951; Forster & Chambers, 1973; Halgren et al., 2002; Ellis, 2002; Hauk & Pulvermüeller, 2003; Duyck et al., 2008) Thus, word frequency effects allow for an examination of lexical retrieval processes and possible neural correlates of lexical retrieval. Murray and Forster (2004) argued that the frequency effect is a tool can be utilized as a diagnostic indicator of the brain’s search and verification process during reading. Frequency effects have been documented using a variety of methods (for review, see Ellis, 2002) and the topic has been investigated in second language research to test hypotheses of the organization of mental lexicon in L1 and L2 speakers.

Behaviors demonstrated during reading tasks provide information on lexical organization and subsequent cognitive actions that arise when encountering words of different frequencies. Studies about the cognitive processes involved in the identification and retrieval of words in both L1 and L2 speakers have primarily used behavioral measures to obtain data (e.g., Howes & Solomon, 1951; Murray &
Forster, 2004; Juhasz et al., 2006; Forster & Chambers, 1973; Mandler, 1980; Frenck-Mestre & Pynte, 1997). Results from these frequency effect studies suggest that a word’s frequency of occurrence affects how long it takes to process the orthographic form and find its meaning. However, behavioral studies of word frequency do not reveal the neural bases of such behaviors.

One of the issues with using behavioral measures is that they are not capable of revealing events occurring in the brain as reading is taking place. That is, the information gathered from reaction times is insufficient in tracking underlying mechanisms happening at the moment. In order to do so, techniques that reflect what is happening during reading comprehension must be used. This is important because measurements of brain response can help bridge the gap between behavior and brain function. One such neuroimaging technique, event-related potentials (ERP), provides real-time measurements of semantic processing during reading tasks with a resolution in the millisecond range. Since language comprehension takes place with extraordinary speed, ERPs are an appropriate tool for capturing electrophysiological responses during the short period between signal (word) encoding and its comprehension (Mueller, 2005).

The present study used ERP to investigate the effect of English word frequency in native (L1) and non-native (L2) sentence processing. One specific ERP waveform, the N400, was used to compare the mean amplitude and peak latency for both groups for critical high and low frequency words embedded in a sentence-reading task. The study also explored a correlation between language proficiency
and magnitude of the N400 wave. The next sections will discuss the purpose and application of ERPs, the N400 wave and previous research exploring the word frequency effect using ERP.

1a. ERP

Measuring neural reactions to the presentation of words and/or sentences requires simultaneously recording electrical brain activity during stimuli exposure. Event-related potentials (ERP) and functional magnetic resonance imaging (fMRI) are frequently used to measure different aspects of brain function. Both are safe, noninvasive techniques that collect data on neural processes given a specific cognitive task. However, the measurement of sentence processing is time-sensitive and requires a technique that has good temporal resolution. fMRI provides good spatial resolution; that is, it is accurate (within a millimeter) in identifying areas of the brain that are activated during a given task. In terms of temporal resolution, fMRI is a slow measure that takes up to several seconds to record the response to the visual or auditory signal. ERPs, on the other hand, have a temporal resolution of 1 millisecond (ms) or better and thus provide excellent measurements of comprehension as they occur during real-time or “online”. While ERPs do not have as fine-grained spatial resolution as fMRI, ERP components do have general anatomical landmarks associated with them (i.e., components are often specific to the left or right hemisphere and/or the anterior or posterior regions) (Luck, 2005).

In general, ERP methodology measures electrophysiological brain response
to visual or auditory stimuli. ERPs reflect voltage variation of electrical signals emitted by neurons in the brain over time (known as electroencephalography or EEG). Electrodes placed on the surface of the scalp capture the signal through a differential amplifier. Specific, time-locked responses to a particular stimulus event cause a voltage change that constitutes the ERP (Coles & Rugg, 1996).

EEG measures raw brain activity from hundreds of different neural sources. It picks up a large amount of electrical activity associated with sensory, cognitive and motor actions (Luck, 2005). Electrical activity in the EEG is measured as a function of amplitude, which can vary with minimum and maximum values of approximately between -100 and +100 microvolts (µV) (Coles & Rugg, 1996). In order to isolate the ERP associated with certain stimuli from spontaneous brain activity unrelated the stimulus (noise), the EEG signal from the time-specific epoch relative to the onset of the stimuli of interest is segmented from the continuous EEG signal and then averaged. For example, an experiment presents a word on the computer screen for 500 ms to an individual. Reading the word will elicit voltage fluctuations directly related to the brain's response upon visualizing the word. In order to derive these functional measures associated with reading the word, time windows of interest are averaged to eliminate noise and, thus, derive the ERP.

An approach called monopolar recording is used in most ERP work to quantify and average the voltage difference between electrodes. Specific to this current study, one active electrode labeled Cz (located at the midline of the scalp) is used as a recording reference. Electrical activity recorded by all the electrodes in the
entire scalp is compared to the activity recorded by Cz. Identical activity picked up by Cz and any other electrode is filtered out by the computer software and only the difference is obtained in order to extract the ERP (Garnsey, 1993). Artifacts like muscle movement constitute a majority of the identical background “noise” filtered out during data collection. Muscle contractions in the form of eye blinks or heartbeat are often regarded as noise because they are not related to the voluntary cognitive operations caused by the stimulus event and thus mask the voltage fluctuations that stem from language processing (Kutas et al., 2006).

ERP data are represented as waveforms with high and low peaks whose components depict the features of the voluntary cognitive operations caused by the stimulus event. Four main components of the ERP are of particular interest for analysis in language research: amplitude, latency, polarity and scalp distribution. The typical data plot of the ERP waveform measures amplitude in the y-axis and latency in the x-axis. Amplitude refers to the microvolt (µV) difference of a peak from the 0 µV baseline. Latency, measured in milliseconds (ms), depicts the timing and duration of the highest amplitude wave following the stimulus presentation (Coles & Rugg, 1996). Polarity describes whether a wave is positive (labeled P) or negative (labeled N) with respect to baseline. A positive wave means an electrode measured electrical activity more positive than baseline and a negative wave was measured as more negative than baseline. Finally, scalp distribution identifies specific electrodes that recognize the greatest increases in voltage (Kutas et al., 2006). Luck (2005) explained that scalp distribution reveals the general area in
which specific ERP waves were originally generated. However with the poor spatial resolution of ERPs, it is difficult to accurately determine the site of exact generation because the electrical signal emitted from the cortex must travel through the brain, skull, meninges and skin before eventually reaching the scalp electrode.

ERP waveforms consist of a series of positive and negative peaks that are plotted relative to stimulus onset time (in milliseconds) and labeled according to their polarity and their latency. For example, the P600 is a positive-going wave that reaches its highest peak 600 ms after the stimulus onset. Other ERP waveforms are labeled according to their functional description (e.g., MMN: mismatch negativity), neural generator (e.g., ABR: auditory brainstem response) or scalp location (e.g., LAN: left anterior negativity). This mix of peaks and components allow for in-depth analysis of ERPs and the cerebral processes and locations involved in real-time cognitive tasks (Kutas et al., 2006).

ERP waveforms manifest as a function of specific cognitive processes. For example, one waveform is specific to comprehension difficulty while another is specific to unfamiliarity of words. One example is mismatch negativity (MMN), which manifests when a subject who is exposed to a series of identical stimuli suddenly encounters a significantly different stimulus (Luck, 2005). Take for example an experiment in which the word *star* is flashed repetitively on a screen to a subject and occasionally the word *octopus* flashes during the series. Presentation of the relatively unfamiliar stimuli *octopus* will elicit the MMN. The MMN along with other ERP components have been identified to correlate with language processing
difficulty (Luck, 2005). Of central importance to this study is one particular waveform: the N400.

1b. N400

The N400 is a negative-going waveform that peaks around 400 milliseconds after stimulus presentation. The component begins around 250 ms after the stimulus onset and then peaks at about 400 ms. It was discovered in 1980 by Kutas and Hillyard who in their experiment utilized ERP measures during a visual reading task in which unexpected words appeared at the end of sentences. Kutas and Hillyard (1980) presented monolingual English subjects with 160 different seven-word sentences that appeared one word at a time on a screen. As the subjects were reading, the experimenters collected electrical activity from an electrode net placed on the scalp. The critical 40 sentences contained semantic violations, ranging from “moderate” (e.g., She walked through the highway) to “strong” (e.g., She walked through the papaya). 120 semantically congruent sentences were used as controls (e.g., I shaved off my mustache and beard). The ERP measures showed a large negative wave (maximal around the parietal area) peaking around 400 ms. It was present in both “moderate” and “strong” conditions but largest for the strong mismatches relative to the control. This wave was called the N400.

Since Kutas and Hillyard (1980) first described the N400, a wide range of studies have reported N400 effects in response to semantic violations (e.g., Garnsey, Tanenhaus & Chapman, 1989; Van Petten & Kutas, 1990; Hahne, 2001; Hahne &
Friederici, 2001; Halgren et al., 2002). However, the N400 is not a catch-all ERP effect that manifests solely when processing semantic violations. In fact, semantic violations are neither required nor sufficient for the occurrence of an N400. In their review of the N400 research within the past 30 years, Kutas and Federmeier (2011) discussed that the N400 is not a discrete neural entity that can be attributed to a single cognitive process. Instead, it is simply an electrophysiological, event-related response present between 200-600 ms post onset of auditory or visual stimuli.

Throughout the literature, the N400 has been elicited by a diverse set of anomalies, such as prepositional violations, pictures of faces and irregular letter shapes, just to name a few (for full review, see Kutas & Federmeier, 2011).

Even though multiple types of stimuli elicit the N400, it has been most commonly reported with experiments containing semantic violations within sentences. The N400 is also observed in word-level processes, which was evidenced when Rugg (1990) analyzed the N400 effect in stand-alone words without syntactic influence. With the effects of sentence ambiguity eliminated, lexical decision tasks were used to test the electrophysiological difference between processing high and low frequency words (Rugg, 1990). Level of frequency was measured by the work from Kucera and Francis (1967) that ranked roughly 1 million English words based on their frequency of occurrence within the language. The high frequency words used by Rugg from the Kucera and Francis corpus had counts of 100 or more (e.g., day, because, people) and low frequency words had counts of 1 (e.g., abrasion, unanswered, jab). In the lexical decision task, individuals were presented with a
series of letter strings in which they were to determine as real or not in the target language (in this experiment, English). EEG data was collected from 16 monolingual English adults who performed the lexical decision task on a total of 150 high frequency words, 150 low frequency words and 100 nonwords. Rugg found that low frequency words produced larger N400 amplitudes than high frequency words. ERP data on nonwords were not reported; however, nonwords produced higher error rates compared to high and low frequency words in the lexical decision task.

1c. Word frequency effects

Rugg’s ERP results on the N400 were consistent with the existing literature about high and low frequency words. Howes and Solomon (1951) and Forster and Chambers (1973) provided evidence that monolinguals for example, recognize high frequency words faster than low frequency words. Both studies used reaction times from a word recognition task to determine a frequency effect. Howes and Solomon (1951) tested 20 monolingual English subjects on 60 words that flashed on a screen with increasing speed from 10 to 1,000 ms. The words differed in frequency according to the Thorndike-Lorge word count, an English word frequency list developed in 1944. The subjects were instructed to read each word aloud as soon as they recognized it. Howes and Solomon found a strong linear relationship between visual duration of the word and word frequency. High frequency words required
shorter exposure durations for accurate recognition and the opposite effect was observed for low frequency words. Along the same lines, Forster and Chambers’ (1973) study involved showing 75 words of high and low frequency and nonwords (e.g., *bice, adit, obol*) to three groups of subjects who performed a lexical decision task. Frequency of occurrence for experimental stimuli was also determined by the Thorndike-Lorge word count. Forster and Chambers reported reaction times of 608 ms for high, 804 ms for low and 763 ms for nonwords with error percentages of 1.0, 14.3 and 2.0 respectively. Low frequency words elicited a more delayed reaction (196 ms) than high frequency words. This information suggested that high frequency words were processed faster in the mental lexicon relative to low frequency.

These behavioral findings led to the development of theories on the underlying cognitive processes for lexical retrieval in monolinguals. For example, Mandler (1980) outlined a two-fold process involved in the recognition of words: familiarity then identification. First, the brain determines the familiarity value of a word upon presentation. Second, if the word is deemed familiar then a retrieval process is engaged to find the word’s meaning. The retrieval process is in essence a serial search through the vast vocabulary stored in the human brain. The studies performed by Howes and Solomon (1951) and Forster and Chambers (1973) showed faster serial searches for high frequency words, as evidenced by faster reaction times. High frequency words appear more commonly in a language’s corpus compared to low frequency words, which allows for more opportunities of
exposure. Following the two-fold process, high exposure makes the connection between familiarity and identification stronger, eventually leading to quick recognition and access of the word’s meaning. Simply put, when we read a word, our brains go through a series of steps in order to conjure up that word’s meaning within the context in which it was presented. This period of time is affected by the target word’s frequency of occurrence. A word that one encounters frequently will be more familiar and thus, retrieving its meaning will be easier.

The idea that increased exposure results in better familiarity may seem intuitive, but a theoretical account of this phenomenon explains why high frequency words are easier to recognize than low frequency words. McCusker (1977, as cited by Duyck et al., 2008) provided such an explanation through the asymptotic learning model. In asymptotic learning, the more times an individual is exposed implicitly or explicitly to a word, the faster they will be able to access its meaning. The process becomes more automatic because the cognitive demands are decreased, making the word more familiar within the lexicon.

Visually, this theory is represented by an asymptote. Because of the nonlinear, logarithmic relation between frequency and reaction time, low frequency words gain more in recognition effects compared to high frequency words. Large recognition effects are seen in the low end of the word frequency spectrum and significantly smaller effects are seen in the higher end of the frequency spectrum. This relationship is illustrated in Fig. 1.
Asymptotic learning model of lexical retrieval. The more an individual is exposed to a word, the less recognition effects is required in order to access its meaning. The decreasing slope depicts decreasing cognitive demand as an individual becomes more familiar with the word.

For example, a high frequency word like *boat* is fairly common within one’s lexicon while a low frequency word like *hovercraft* is not. The 600th time an individual comes across the word *boat* will not elicit a large recognition effect because of the numerous exposures they previously had with it. *Boat* has already been systematically encoded in their lexicon so the 601st exposure in comparison will not make familiarity of the word that much stronger. In contrast, the 3rd time an individual comes across *hovercraft*, they gain a significant amount in recognition because its concept is much newer and fairly unfamiliar. This 3rd exposure strongly encodes *hovercraft* and has a significant effect on its familiarity. This example demonstrates that frequency does not yield a uniform effect at all times. Fig. 1
illustrates that increments in the high end of the frequency scale has far less impact than increments in the low end. Figure 1 also shows that there comes a point in which one is exposed to a word so many times that they become efficient in understanding the word’s meaning. Simply put, they become familiar with the word because of the strengthened memory traces that make up the lexical representation of the word (Murray & Forster, 2004).

With regard to L2 learners, Larsen-Freeman (2002) suggested that exposure to a word does not necessarily improve recognition because frequency does not guarantee attention. One must attend to the word and possess a mental framework or schema into which the word fits in order to incorporate it into his/her lexicon. Without this type of schema, the word may be lost in all the linguistic stimuli individuals are exposed to, and become noise. This idea is especially important for L2 speakers since during the beginning stages of L2 learning, most L2 words have such low frequency of occurrence that encoding them requires explicit learning, focusing on their form and meaning. Early explicit vocabulary learning of L2 speakers allows for a knowledge base that creates the framework or schema that later helps derive vocabulary acquisition implicitly through reading. Transitioning from explicit to implicit word learning requires different modes of instruction. A review on L2 vocabulary acquisition conducted by Huckin and Coady (1999) concluded that the first few thousand most common L2 words are learned through direct instruction. After this period of explicit learning, L2 speakers shift to an “incidental” learning phase in which they acquire the meaning of new L2 words.
Translating the behavioral theory of asymptotic learning to neural bases is facilitated by ERPs. Since ERPs measure online electrical brain activity during reading tasks, it is useful in testing cognitive effects of frequency. The theories of Mandler and McCusker suggest that ERP waveforms like the N400 would be associated with word frequency. Cognitive difficulty should be evidenced by larger N400 peaks for low frequency words because the retrieval process is more labored due to low familiarity. In fact, word frequency effects on ERP have been reported on some ERP measures. Hauk and Pulvermüller (2004) claimed that amplitude differences in the ERP data associated with word frequency reflect plasticity of the neural networks in the brain. They tested 12 native English speakers on a lexical decision task using short and long words of different frequencies (low, medium, high). Each subject read 414 high and low frequency words. The results showed significantly lower ERP amplitudes for high frequency words compared to low frequency words only in the latency ranges of 150-190 ms and 320-360 ms. Although an N400 wave was not reported, Hauk and Pulvermüller provided evidence of significant electrophysiological differences between high and low frequency words. They hypothesized that as the exposure of word increases the synaptic connections of the brain’s neurons become more efficient for the representation of the word and thus, less activation is required to retrieve it. Additional support of Hauk and Pulvermüller’s hypothesis is outlined in Binder’s (2008) explanation of visual word perception. He reported that during reading,
neurons in a region of the brain called the visual word form area are activated and function more efficiently with high frequency words. High frequency words have familiar letter combination and are easily detected by the neurons that in turn, develop proficiency in processing multi-letter strings and even whole words in parallel.

Other studies have found N400 effects between high and low frequency words (e.g., Rugg, 1990, Van Petten & Kutas, 1990). For example, Van Petten and Kutas (1990) performed two experiments in which participants read sentences that appeared one word at a time on a screen. Experiment 1 gathered data from 43 native English speakers who read 338 unrelated sentences with target high or low frequency words embedded in the first, middle and final position. The results showed an N400 effect for low frequency words in the first and middle positions, but not for low frequency words in the last sentence position. Further analysis of the data showed that the N400 peak was greater for low frequency words in the first position than the middle position. One problem with the experiment was that words in the three positions were not matched by length. Experiment 2 closely mirrored the design of Experiment 1 but controlled for word length. Native English speakers (n=31) read 240 unrelated sentences with high and low frequency words in only the first and middle positions of the sentences. A significant N400 effect was found only for low frequency words in the first position. Perhaps the reason for an N400 effect only in one position is that words in position 1 are not preceded by other words that may provide syntactic clues to the target word’s meaning.
The results from the Van Petten (1990) experiment along with Rugg (1990) are relevant to L2 word recognition because they give a basis for comparison of the word frequency effect with L2 speakers. If low frequency words generally elicit slower reaction times and greater N400 amplitudes in L1 speakers, will these same results manifest in L2 speakers? If the amplitudes were found to be different then analyzing the significance and degree of difference might reflect how the cognitive processes in L2 speakers handle word frequency.

Kerkhofs, Dijkstra, Chwillaa and de Bruijn (2005) reported frequency effects in L2 speakers in their ERP study of priming effects for interlingual homographs (IH), which are orthographically identical words in two languages that have different meanings. This study was important because it was one of the earliest ERP studies that explored frequency effects in L2 speakers. The experiment tested 17 native Dutch university students with high proficiency in English with an average of 13 years of English language classroom instruction. The study utilized English-Dutch pairs that were presented to the participants one word at a time. The participants were instructed to decide whether the second word was an English word or not. Half of the pairs were interlingual homographs and the other half were semantically unrelated. The homograph targets were split into four categories of varying word frequency in English and Dutch. The ERP data showed that low frequency L2 English words elicited the largest N400 amplitude. It also showed that the N400 peak was affected by word frequency in both languages.

Duyck et al. (2008) also explored the frequency effect in a study utilizing
bilingual Dutch-English speakers (L1:Dutch, L2:English) and monolingual English speakers during an English lexical decision task. In experiment 1, 18 bilingual Dutch university students with at least 5 years of English language instruction performed lexical decision tasks in their L1 and L2. The results showed significantly larger frequency effects in English than Dutch. Average reaction time for English low frequency words was 129 ms greater than reaction time for low frequency words in Dutch. In comparison, high frequency difference in both languages was 72 ms. The 129 ms difference in reaction time for low frequency words was statistically significant while the 72 ms difference for high frequency words was not. In experiment 2, 20 monolingual American university students performed a lexical decision task in English. The data showed a frequency effect for the American students in English similar to the frequency effect shown by the Dutch students in their native Dutch language (native Dutch frequency effect: 46 – native English frequency effect: 52). Although this study did not utilize ERP methodology to measure an N400 effect, it gives support to the idea that the magnitude of word frequency effects will be greater in L2 than L1 speakers. This also supports the asymptotic learning model proposed by McCusker (1977) in that low frequency English words were encountered much less often for the Dutch-English bilinguals as compared to low frequency words in the their L1 of Dutch.
1d. Models of L2 word recognition

There exist a variety of theoretical models for L2 word recognition and each one uniquely characterizes the process in which L2 speakers recognize words in their native and non-native languages. Duyck et al. (2008) discussed two theoretical models of word recognition and their implications on L2 lexical access: implicit learning and rank hypothesis. Implicit learning states that repeated exposure causes a word’s recognition threshold to be lowered. The rank hypothesis (Murray & Forster, 2004) states that one’s vocabulary is organized in frequency-ordered bins (lists) that are searched serially during reading tasks. The lexicon is divided into bins that contain word with similar orthographic properties. Within each bin, words are ranked according to their relative frequency of occurrence in the language and high frequency words are listed first (see Fig. 2). High frequency words are searched faster than low because they are situated at the top of the list.
Murray and Forster (2004) discussed that each bin has its own unique hash-code function, which operates similarly to that of an address. When a word is read, its orthographic properties are assessed and the mental search leads directly to the bin location that closely matches the input letter string. Murray and Forster argued that dividing the lexicon into bins serves a useful purpose. When trying to access a word’s meaning, a serial search throughout all encoded entries in one’s lexicon is essentially impractical. This search is exhaustive and carelessly consumes cognitive effort. In a bin structure, only the contents of relevant bins are searched. This unique property of the rank hypothesis shows that access time is a function of relative frequency of a word’s occurrence within their specific bin, not its absolute frequency in an individual’s entire mental lexicon. The bins themselves are organized according to orthographic form. Letter strings that occur more frequently within an individual’s mental lexicon are situated closer to the top and letter strings that occur less frequently form bins in the lower end.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Frequency</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>534654</td>
<td>the</td>
</tr>
<tr>
<td>2</td>
<td>168580</td>
<td>thing</td>
</tr>
<tr>
<td>3</td>
<td>21321</td>
<td>that</td>
</tr>
<tr>
<td>4</td>
<td>10005</td>
<td>this</td>
</tr>
<tr>
<td>5</td>
<td>998</td>
<td>thought</td>
</tr>
<tr>
<td>6</td>
<td>500</td>
<td>then</td>
</tr>
<tr>
<td>7</td>
<td>106</td>
<td>they’ve</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>thigh</td>
</tr>
</tbody>
</table>

Figure 2: Adapted from Murray and Forster (2004) All entries have similar orthographic form: all one syllable words that begin with “th” and followed by a vowel.
Duyck and colleagues (2008) discussed a possible extension of the rank model. On their account, the lexicon of both L1 and L2 are ordered in the same list within the same bins. Since L2 words are encountered in fewer opportunities than L1 words, they would occupy a lower rank in the list. This is assuming that the L1 and L2 item share a similar orthographic form. For example, air and **aire** are high frequency words in English and Spanish respectively. For an L2 speaker of Spanish, **aire** would occupy a lower rank in their vocabulary list because **aire** is used far fewer than a majority of their English vocabulary words. During a word recognition task, the processor would scan the entire list from the top until the target word is found. This is evidenced by the results of the study in which the L2 lexical decision task produced a larger frequency effect than L1.

However it is uncertain what arises when an L1 item and its L2 equivalent do not share similar orthographic forms. Examples are illustrated in the following English and Spanish translations: **cat/gato; stapler/engrapadora; rollercoaster/montaña rusa.** When adhering to Murray and Forster’s (2004) bin model, the English and Spanish items would occupy separate bins because of their different orthographic forms. The Spanish words would still be located in bins with other orthographically similar English words but not in the same bin as their English translation. Further, Duyck et al. (2008) proposed a situation in which an individual’s L1 and L2 have different orthographic systems, such as that of Chinese and English. In this situation, lexical items in the L1 and L2 will have no possibility of sharing any frequency-ordered bin because the alphabet systems have no
It is unclear how orthographically dissimilar L1 and L2 words fit into Murray and Forster's (2008) bin model and it is also unclear how their organization in the mental lexicon will affect the retrieval process. Since the current study looks at the word frequency effects in a group of Chinese-English second-language learners, a discussion of their results will touch on the organization of the L1 and L2 and their assumptions.

1e. Present Study

The purpose of the present study was to use ERP methodology to explore the effects of high and low frequency words during sentence reading. Specifically, this study will compare N400 waveforms between native and non-native English speakers during a sentence reading task. As mentioned before, there is robust evidence of word frequency effects in monolingual language processing (Forster & Chambers, 1973; Rugg, 1990; Van Petten & Kutas, 1990; Ellis, 2002). N400 peaks have manifested in both the single-word and sentence processing levels, with low frequency words generally eliciting higher N400 peaks. The present study compared performance of both the L1 and L2 groups to determine if there is a significant difference in the averaged N400 of both groups in the two frequency conditions. Further, the study also examined whether proficiency in L2 is correlated with magnitude of the N400.

Three possible outcomes are proposed with varying assumptions:
1. Both groups will show frequency effects. Frequency effects will appear as higher N400 amplitude and later peak latencies for low versus high frequency words but the L2 group will show greater frequency effects than the L1 group. This would mean that the L2 group was more sensitive to the low frequency words than high frequency because accessing meaning required a much higher recognition threshold as compared to the L1 group. This hypothesis assumes separate frequency-ordered bins for L1 and L2 entries and also that L2 speakers experience a delay in mapping on to their L2 bins.

2. The L1 group will show a frequency effect, but the L2 group will not. That is, the waveform peaks (amplitude and latency) of the L2 group for high and low frequency items will not differ significantly. This assumes that L1 and L2 entries for the L2 speaker are not frequency-ordered and are perhaps organized by another factor.

3. The L1 group will show a frequency effect, but the L2 group will show higher amplitude and later peak latencies for high versus low frequency words, a reverse frequency effect. Waveform peaks of the L2 group for high and low frequency items will differ significantly. This pattern of results would suggest that low frequency items for the L2 group would show little to insignificant activation because they will be treated as nonwords, suggesting that no word retrieval mechanisms are taking place. This assumes that only high-frequency words are entered into the L2 speakers' bins.
2. METHODS

2a. Participants

Thirty-nine individuals aged 18-32 were recruited from the University of Arizona and Tucson, Arizona community. Nineteen participants were native speakers of English and twenty participants were non-native English speakers from China and Taiwan who began formal study of English around age 12 (at age 9 for one participant). All non-native English speakers were undergraduate and graduate student at the University of Arizona. They had all achieved an intermediate to advanced level of proficiency prior to the experiment based on self-report and their state minimum Test of English as a Foreign Language (TOEFL) score for University of Arizona admission. All participants had completed at least 13 years of formal education. Each participant was paid or received academic credit for participating in the study. All participants reported that they had normal or corrected-to-normal vision, were right-handed, had no history of cognitive impairment or a learning disability and had never suffered a head injury causing a loss of consciousness. At the time of the experiment all participants reported that they were not taking medication that altered natural brain activity (anti-psychotics, anti-depressants, psychostimulants, etc.).

2b. Stimuli

The experimental stimuli consisted of 30 pairs of sentences. One member of the pair contained a high frequency word, and the other contained a low
frequency word. Distracter sentences were included to add variety and consisted of various syntactic structures. The experimental sentence pairs were 9-12 words long and contained matched high and low frequency nouns. The critical nouns were 7-10 letters long and matched exactly on word length. Low frequency nouns (e.g. _catacomb, camouflage, trellis_) had a mean frequency of 1.35 words per million on the CELEX (Linguistic Data Consortium, Philadelphia, PA) database and high frequency words (e.g. _trouble, village, machine_) had a mean frequency of 143 words per million.

The experimental sentences were adapted from stimuli developed by Juhasz, Liversedge, White, and Rayner (2006) for their study of word frequency effects. Each sentence pair was exactly the same except for word position 5, which was a high or low frequency word. To reduce priming effects, sentence structure did not allow for anticipation of the fifth word. Priming refers to the exposure of an irrelevant stimulus influencing the response to a later stimulus.

Each high and low frequency pair was separated into two lists along with distracters. One list was presented in the first session and the other in the second session. Below, A and B are examples of the experimental sentences with the target word underlined in position 5.

<table>
<thead>
<tr>
<th>A (high frequency)</th>
<th>John repaired the old <strong>machine</strong> very quickly</th>
</tr>
</thead>
<tbody>
<tr>
<td>B (low frequency)</td>
<td>John repaired the old <strong>trellis</strong> very quickly</td>
</tr>
</tbody>
</table>
Four lists were utilized for the experiment. Each participant was randomly assigned a set of two lists. Lists 1 & 2 and 3 & 4 contained the different members of the frequency pair (i.e., the high and low frequency versions). Lists 1 & 3 and 2 & 4 contained the same items in a different order. Participants either read lists 1 and 2 in the two testing sessions, or lists 3 and 4. Each list contained 120 sentences: 30 critical sentences (15 high frequency, 15 low frequency) and 90 fillers.

The distractor sentences were for unrelated experiments and consisted of plausible and implausible sentences that did not contain the critical nouns. C and D below are examples of the distractors.

<table>
<thead>
<tr>
<th>C (plausible)</th>
<th>D (implausible)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The umpire asked which player the coach threatened before the game.</td>
<td>The referee asked whether the team threatened the football before the event.</td>
</tr>
</tbody>
</table>

At the end of the second session, each participant had read a total of 240 sentences: 60 experimental and 180 distractors.

2c. Procedure

The testing was completed over two sessions, which were scheduled at least one week apart. The sessions lasted approximately 2 hours each. The sentence reading experiment was performed with the participant sitting comfortably in a chair inside a sound-proof booth facing a computer monitor with the participant’s
dominant hand resting on a button box. Each participant was fitted with a 128-electrode EEG net prior to entering the booth. Before the experiment began, the experimenters ensured adequate net-scalp contact by checking the resistance level between the skin and electrode (impedance, measured in kΩ) and making adjustments as necessary (e.g., repositioning electrodes or adding additional electrolyte solution; See below for further details on the EEG methodology). Instructions for the experiment were presented on the computer monitor as an experimenter read them out loud to the participant. Each participant was instructed to keep still and not blink during the sentence presentation. Prior to recording of the experimental procedure, the participants completed 10 practice trials to ensure they understood the task. They were also provided the opportunity to ask the experimenter questions regarding execution of the procedure.

Stimuli were presented in the serial visual presentation paradigm. Sentences were presented one word at a time for 350 ms followed by a blank screen for 350 ms, for a total duration of 700 ms. A fixation cross appeared for 1000 ms in the center of the monitor to indicate the start of the presentation followed by a blank screen for 300 ms and then the first word of the sentence. After the final word of each sentence, the participant was prompted to make a decision about the plausibility of the sentence with a button press. Button 1 corresponded to a sentence that made sense and button 2 for a sentence that did not make sense.

Each EEG recording session was divided into four blocks. 30 sentences were presented in each block. At the end of a block experimenters, entered the booth to
recheck impedances and made sure they were kept under 50 kΩ. A five-minute break occurred at the end of the second block. Stimuli were presented electronically using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). E-Prime controlled presentation rate and collected accuracy for responses.

2d. Individual Differences

In order to measure English proficiency, all participants completed two vocabulary tests and one lexical decision task. The non-native English speakers also completed a proficiency questionnaire that included self-report measures of speaking, listening, reading and writing in English on a 5-point Likert scale. The two vocabulary tests used were the Shipley Vocabulary Test (SVT) and the WAIS Vocabulary Subtest. The vocabulary tests were administered randomly either before or after the experimental procedure. The lexical decision task was always administered after the experimental procedure at the end of the second session.

The SVT is one half of the Shipley Institute of Living Scale (SILS, Shipley, 1940), a standardized assessment of mental impairment. The SVT required the participant to silently read a capitalized target vocabulary word and choose its closest synonym from a set of four different words within the same line. A total of 40 vocabulary words were tested ranging from more familiar words (e.g., talk, couch) to less familiar words (e.g., querulous, temerity). The WAIS Vocabulary Subtest is part of the Wechsler Adult Intelligence Scale (Wechsler, 1955). It is a measure of expressive vocabulary knowledge that required the participant to verbally provide a
definition to a word read out loud by an experimenter. A total of 35 words of increasing difficulty (e.g., word #4: winter, word #35: tirade) were tested. The participant was instructed to guess definitions of words unfamiliar to them. Responses were transcribed by an experimenter and scored on a scale of 0-2. 2 indicated a correct, full definition, 1 indicated a partially correct definition and 0 indicated an incorrect definition.

The lexical decision task comprised 160 stimuli words that flashed on a computer screen for 500 ms with a 500 ms interval between words. The stimuli were 80 real English words and 80 nonwords (e.g., clapsum). The participant was seated in front of a screen in a soundproof booth and rested their dominant hand on a button box. They were instructed to select button 1 if the letter string was a real English word and button 2 if it was a nonword. Since accuracy and response time were measured through this task, the participant was instructed to respond as quickly and as accurately as possible.

2e. ERP Data Collection

EEG data was recorded using an Electrical Geodesics System 300, with a high density 128-channel Hydrocel Geodesic Sensor Net composed of carbon-filled plastic electrodes with a silver/silver-chloride layer (Ag/AgCl) in electrolyte-wetted sponges. Net Station acquisition software (Electrical Geodesics, Incorporated, Eugene, OR) was used to run EEG operations through a portable Mac OS X 10.4
PowerPC. The signal was amplified with an Electrical Geodesics Net Amps 300 amplifier. Electrode positions are shown below in Fig. 2. EEG was continuously recorded with a sampling rate of 500 Hz and rereferenced to the whole brain to establish the grand average waveform. Electrode impedances were kept below 50 kΩ and a 30Hz low pass filter and 0.1 Hz high pass filter were applied to the EEG.

Figure 3: high density 128-channel Hydrocel Geodesic Sensor Net
2f. Analysis

Data were segmented 100 ms before and 900 ms after the presentation of the critical word. Artifact detection was programmed to mark bad channels and segments due to extraneous muscle contractions like eye movements and eye blinks. ERP data for each experimental sentence were manually reviewed to further mark bad channels and segments not identified by the software. Postprocessing of the data consisted of four processes: bad channel replacement, averaging, rereferencing and baseline correction. Bad channel replacement used an algorithm to interpolate for missing data in order to replace bad channels with an approximation of a waveform of the signal that was presented at that location on the scalp during recording. The computer software calculated the average waveforms of the electrical activity measured from all 30 high frequency sentences and 30 low frequency sentences. This allowed for two grand average waveforms (one for high frequency and one for low frequency) per participant. The software also identified what net was being used and electrodes were rereferenced using an average of all electrodes. Finally, baseline correction set the voltage measured 100 ms before word 5 as close to 0 µV as possible in order to create a baseline.

In order to statistically analyze ERP effects, electrodes on the high-density montage net were divided into two regions. Data were analyzed to compare clusters of anterior and posterior electrodes in their measurement of the N400 waves. This technique, as suggested by Dien and Santuzzi (2005), allows for several advantages. It makes for easier interpretation of experimental effects by organizing factors into
descriptive units. This helps reduce unwanted variability when performing the statistical analysis. This clustering technique has been replicated in other ERP research with positive results (e.g., Hestvik, Bradley & Bradley, 2012). For this study, posterior centroparietal electrodes, upon visual inspection, were utilized because of their robustness to the N400. Fig. 3 below illustrates the electrodes that comprised of the anterior and posterior divisions. Anterior electrodes comprised of channels 4, 10, 11, 18 and 19. Posterior electrodes comprised of channels 31, 37, REF, 54, 55, 61, 62, 78, 80 and 87.
Figure 4: The cluster of blue circles represents the anterior electrode region and the red represents the posterior electrode region.
3. RESULTS

The purpose of this study was twofold: (1) to compare the mean amplitude and latency of the N400 wave elicited by two word frequency conditions between groups of L1 and L2 English-speaking adults; and (2) to examine whether language proficiency is correlated with magnitude of the N400.

3a. ERP Results

ERP data from the electrodes were averaged for each group and frequency condition. The 300-450 ms time window was analyzed in order to capture the N400 effect for critical words because it was the time period in which the N400 was most robust. Two values were extracted: the mean amplitude of the waveform within the time window, and the latency of the waveform peak. Each value was analyzed with 2 (group: L1 vs. L2) by 2 (frequency: high vs. low) by 2 (distribution: anterior vs. posterior) mixed ANOVAs. Group was the between participants variable and distribution and frequency were the within participants variables. The dependent variables were the mean amplitude and latency in the 300-450 ms time window.

Table 1 presents the full results of the ANOVAs for the mean amplitude and latency values across all variables and variable interactions.
Table 1: Results of mean amplitude and latency values across all variables and variable interactions. * marks significance (p < 0.05)

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Mean Amplitude</th>
<th>Latency</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F-value</td>
<td>p-value</td>
<td>F-value</td>
<td>p-value</td>
</tr>
<tr>
<td>Group</td>
<td>1,37</td>
<td>0.47</td>
<td>0.50</td>
<td>0.47</td>
<td>0.50</td>
</tr>
<tr>
<td>Frequency</td>
<td>1,37</td>
<td>0.35</td>
<td>0.55</td>
<td>0.62</td>
<td>0.44</td>
</tr>
<tr>
<td>Distribution</td>
<td>1,37</td>
<td>4.54</td>
<td>0.04*</td>
<td>3.45</td>
<td>0.07</td>
</tr>
<tr>
<td>Freq x Dist</td>
<td>1,37</td>
<td>5.72</td>
<td>0.02*</td>
<td>0.29</td>
<td>0.59</td>
</tr>
<tr>
<td>Group x Dist</td>
<td>1,37</td>
<td>0.20</td>
<td>0.65</td>
<td>0.28</td>
<td>0.60</td>
</tr>
<tr>
<td>Group x Freq</td>
<td>1,37</td>
<td>0.35</td>
<td>0.46</td>
<td>0.30</td>
<td>0.59</td>
</tr>
<tr>
<td>Group x Freq x Dist</td>
<td>1,37</td>
<td>1.72</td>
<td>0.20</td>
<td>0.45</td>
<td>0.51</td>
</tr>
</tbody>
</table>

The mean amplitudes within the 300-450 ms time window revealed an N400 effect across both L1 and L2 groups. Within the latency values, there was no main effect observed across all variables and variable interactions. The peak of the wave was marginally delayed for the anterior electrodes (mean = 370 ms) relative to the posterior electrodes (mean = 356 ms) but this difference was not significant, F(1,37)=3.45, p = 0.07.

With respect to mean amplitude, the main effects of group and frequency were not significant. There was a main effect of distribution, F(1,37) = 4.54, p = 0.04. Mean voltage of the cluster of anterior electrodes was -0.42 μV and posterior electrodes were -1.02 μV. The difference in voltage of the two regions was 0.60 μV with negativity more than double in the posterior region as compared to the
The interaction between frequency and distribution was also significant, $F(1,37) = 5.72$, $p = 0.02$. The effect of distribution did not significantly interact with group and frequency. For the interaction of the frequency condition and distribution, low frequency words in the posterior regions revealed the most negativity. Low frequency-posterior: -1.26 $\mu$V; low frequency-anterior: -0.26 $\mu$V; high frequency-posterior: -0.79 $\mu$V; high frequency-anterior: -0.58 $\mu$V. Further analysis revealed that critical high and low frequency words in the anterior region did not reveal significance, $p = 0.17$, while high and low frequency words in the posterior region did reveal significance, $p = 0.05$. There was marginal significance in the N400 as being more robust in the posterior region.

Table 2 summarizes the means and standard deviations of the different conditions. Since there was no significant difference in the mean amplitude of the N400 between groups, discrete L1 and L2 values are not reported.
<table>
<thead>
<tr>
<th>Frequency</th>
<th>Distribution</th>
<th>Group</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>High freq</td>
<td>Anterior</td>
<td></td>
<td>-0.58</td>
<td>1.77</td>
</tr>
<tr>
<td>High freq</td>
<td>Posterior</td>
<td></td>
<td>-0.79</td>
<td>1.13</td>
</tr>
<tr>
<td>Low freq</td>
<td>Anterior</td>
<td></td>
<td>-0.26</td>
<td>1.73</td>
</tr>
<tr>
<td>Low freq</td>
<td>Posterior</td>
<td></td>
<td>-1.26</td>
<td>1.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Group</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>High freq</td>
<td>L1</td>
<td>-0.85</td>
<td>1.40</td>
</tr>
<tr>
<td>High freq</td>
<td>L2</td>
<td>-0.53</td>
<td>1.55</td>
</tr>
<tr>
<td>Low freq</td>
<td>L1</td>
<td>-0.83</td>
<td>1.51</td>
</tr>
<tr>
<td>Low freq</td>
<td>L2</td>
<td>-0.69</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Table 2: On the top half, the average voltage recordings (in µV) during interaction of word frequency and distribution. On the bottom half, average recordings of both groups in high and low frequency conditions.

Figures 5 and 6 present topoplots of mean N400 wave amplitudes for the L1 and L2 group in both frequency conditions. All target electrodes are depicted with their corresponding average N400 wave for all experimental stimuli. The grand average of all electrodes is depicted in the REF electrode. A clear peak around the 400 ms time mark is illustrated in the REF electrode with the low frequency stimuli (red line) eliciting a higher amplitude than the high frequency (black line). Both figures illustrate stronger N400 recordings in the posterior region of the scalp compared to the anterior.
Figure 5: Anterior and posterior electrode recordings of N400 waves for the L1 group. The low frequency wave is represented by the red line and the high frequency by the blue line. The grand average waveform of all electrodes is depicted by the REF electrode.
Figure 6: Anterior and posterior electrode recordings of N400 waves for the L2 group. The low frequency wave is represented by the red line and the high frequency by the blue line. The grand average waveform of all electrodes is depicted by the REF electrode.
Overall, the ERP data revealed that there were no significant differences in the N400 waves of both groups for the high and low frequency words. Significance was found only in the mean amplitude values of anterior and posterior electrode regions as posterior regions recorded the more robust N400 wave. Interaction between frequency and condition also revealed significance as low frequency words in the posterior electrodes elicited the most negativity.

3b. Correlation Results

Correlational analyses of the data examined whether individual differences in proficiency were related to the magnitude of measured N400 waves. Proficiency values were calculated from scores of the lexical decision task and Shipley Vocabulary Test. The WAIS Vocabulary Subtest score was not included because it reflects productive ability and not vocabulary knowledge. The proficiency value was the averaged z-scores of the lexical decision task and Shipley.

The N400 wave used in this correlation analysis was the difference wave of mean amplitude of high frequency and low frequency conditions for the average of the posterior electrodes.

Table 3 summarizes the proficiency scores and mean N400 amplitudes of both groups.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L1</strong></td>
<td>Proficiency</td>
<td>19</td>
<td>0.35</td>
<td>0.64</td>
<td>6.75</td>
<td>-0.87</td>
<td>1.38</td>
<td>p = 0.73</td>
</tr>
<tr>
<td></td>
<td>N400</td>
<td>19</td>
<td>0.60</td>
<td>1.20</td>
<td>11.44</td>
<td>-1.15</td>
<td>3.07</td>
<td>r = -0.08</td>
</tr>
<tr>
<td><strong>L2</strong></td>
<td>Proficiency</td>
<td>20</td>
<td>-0.25</td>
<td>0.70</td>
<td>-4.96</td>
<td>-2.27</td>
<td>0.96</td>
<td>p = 0.03</td>
</tr>
<tr>
<td></td>
<td>N400</td>
<td>20</td>
<td>0.34</td>
<td>0.94</td>
<td>6.87</td>
<td>-1.55</td>
<td>1.86</td>
<td>r = -0.48</td>
</tr>
</tbody>
</table>

Table 3: Proficiency scores and mean amplitude (µV) of the difference waves of L1 and L2

The L1 group achieved a mean proficiency score of 0.35 with an average N400 difference wave of 0.60 µV. The mean proficiency score for the L2 group was -0.25 with an average N400 difference wave of 0.34. The analysis revealed no significant correlation between proficiency and N400 wave for the L1 group, r = -0.08, p = 0.74. However, proficiency and magnitude of the N400 difference wave were significantly correlated in the L2 group, r = -0.48, p = 0.03. This suggests that a negative correlation exists between proficiency and N400 difference wave in the L2 speakers. As proficiency increases in the L2, the gap between high and low frequency N400 peaks decreases. However, it is unknown if both or one of those waves are changing. One of three possible events is occurring: 1) the high frequency wave amplitude stays the same as the low frequency amplitude is decreasing; 2) the high frequency wave is increasing as the low frequency is decreasing; and 3) the
A high frequency wave is increasing as the low frequency amplitude stays the same.

No further statistical analysis was conducted on the correlation measure; however, an observation of the patterns of proficiency and N400 effect size is made with the scatterplot depicted in Fig. 7 and Table 4.

![Figure 7: Scatterplot of proficiency score (x-axis) and N400 effect size in posterior electrodes (y-axis) of all 19 L2 speakers](image)

<table>
<thead>
<tr>
<th></th>
<th>High Proficiency (&gt; -0.1)</th>
<th>Low Proficiency (&lt; -0.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Freq</td>
<td>-0.57</td>
<td>-0.76</td>
</tr>
<tr>
<td>Low Freq</td>
<td>-0.77</td>
<td>-1.27</td>
</tr>
<tr>
<td>Effect Size</td>
<td>0.20</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 4: Average high and low frequency N400 amplitudes of high and low proficiency groups of L2 speakers
Fig. 7 represents the plot of the relationship between proficiency z-score and the N400 effect size. In order to understand the effect, the L2 speakers were divided into high and low proficiency groups based on a median split. That is, the L2 learners with the 10 highest scores were placed in the high proficiency group (> -0.1) and the 10 lowest scores were placed in the low proficiency group (< -0.1). The low proficiency L2 speakers show a numerically larger frequency effect (0.51 vs. 0.20), which seems to reflect more negativity in the low frequency items. This is illustrated in Table 4. It is also interesting to note that the low proficiency group is more negative overall. The mean proficiency score for the high proficiency group was .28 and the mean proficiency score for the low proficiency group was -.78.
4. DISCUSSION

This experiment examined the N400 effect in native and non-native English speakers during a reading task using lexical stimuli with different frequencies of occurrence. ERP results of both groups were compared to identify differences in the N400 mean amplitude and latency relative to the groups themselves, regions on the electrode scalp net and frequency of occurrence of the experimental stimuli. The study also examined the correlation between proficiency and N400 difference wave amplitude in the L2 group.

Analyses of the data revealed no significant N400 difference in mean amplitude or latency between the L1 and L2 groups. Essentially, both groups elicited a larger N400 wave for low frequency words relative to high frequency words. The N400 wave of both frequency conditions also peaked around 400 ms after the onset of the critical word (see Figures 5 & 6). This is consistent with the ERP results of previous studies from the monolingual literature on word frequency effects (Rugg, 1990; Van Petten, 1990; Kerkhofs et al., 2005). The interaction between stimuli frequency and scalp region revealed that low frequency words recorded in posterior electrodes produced an N400 with the most negativity, relative to the anterior region.

The results did not support the second and third hypothesis outlined in section 1e; however, they were partially consistent with the first hypothesis. The first hypothesis predicted that both groups would show an N400 effect characterized by a higher mean amplitude and later latency peak for low frequency
words relative to high frequency words. Additionally, the first hypothesis predicted that the L2 group would show a larger frequency effect than the L1 group. The results found that both groups did not differ. The data suggest that processing low frequency words for the L2 group was similar to the L1 group. Assuming that large N400 waves represent cognitive difficulty and longer search times, the results imply that the time course of lexical retrieval was similar in both groups, especially since no significant latency differences were found. When applying the model of asymptotic learning (McCusker, 1977, as cited by Duyck at el., 2008), recognition effects for low frequency words are represented as larger N400 waves relative to the waves of high frequency words. The larger waves reflect increased cognitive effort because entries of low frequency in the mental lexicon are not as strongly encoded as their high frequency counterparts. The smaller amplitude of the high frequency words is a consequence of automaticity. Visually, this is depicted on the higher end of the slope of the logarithmic learning function where word exposures are increasing while recognition effects are decreasing (see Figure 1).

This automaticity in the lexical retrieval process is interesting because one might assume that the L2 group would access word meaning less efficiently than the L1 group. Since all non-native speakers in the current study had begun formal English language instruction around age 12 (one participant began at age 9), the L2 group underwent a long period of limited exposure to English, which makes for a large learning curve once formal language instruction begun. It is assumed that the L2 speakers would be less familiar with English vocabulary, resulting in greater
cognitive demand during lexical retrieval tasks. Despite this, there comes a point during L2 learning in which lexical retrieval starts to become easier and more automatic. This is represented in the lower end of the asymptotic learning slope. Although the L2 group may not have had an equal number of exposures to experimental stimuli throughout their language-learning experience as the L1 group, the actual occurrences of exposure does not have a strong impact in the higher end of the slope. For example, exposure number 100 of the word *hovercraft* for an L2 speaker may invoke similar recognition effects as exposure number 500 of *hovercraft* for an L1 speaker.

The results of the current study support this asymptotic learning slope model because there was no significant difference in N400 amplitude and latency between the L1 and L2 groups. The late language learning of the L2 group did not result in added cognitive difficulty in processing the experimental stimuli because their measured electrophysiological activity was similar to the native-speakers in the L1 group. The L2 group demonstrated that they reached the point in which lexical retrieval becomes more automatic through native-like N400 wave data. Further examination of the L2 group also showed a negative correlation with N400 effect size and proficiency. The scatterplot in Figure 5 illustrated that the low proficiency L2 learners elicited a larger frequency effect as compared to the high proficient L2 learners. The low proficient group also showed more negativity in low frequency items.

Since the N400 is sensitive to an individual’s language proficiency, it is a
useful tool for examining competence in L2 acquisition. The N400 data of the L2 group showed that they processed words similarly to the L1 group, suggesting that adult L2 learners are capable of very fast word learning. It was previously thought that development of adult L2 vocabulary knowledge is slow and native-like word processing cannot be achieved until well after a long period of direct instruction (Ellis, 2005). McLaughlin, Osterhout and Kim (2004) showed that the emergence of native-like lexicality judgments in L2 learners is achieved much earlier than previously assumed. They found that a group of French nonlearners (individuals without previous French exposure) started to distinguish French words from pseudowords with as little as 14 hours of French lessons. During a lexical decision task the nonlearners elicited higher amplitude N400 waves for pseudowords relative to real French words.

The remarkable speed of word learning in L2 speakers provides additional support for the asymptotic learning model since large recognition effects are seen within the first few exposures of a novel word. Further, it sheds light on the amount of L2 exposure needed for L2 learners to cross the threshold in which one becomes efficient in understanding a word’s meaning/concept. Although the exact number is still unclear, the results of McLaughlin, Osterhout and Kim (2004) suggest that near-native processing of known L2 words may occur within the first year of formal language instruction. Since the non-native participants in this present study have well over a year of English classroom instruction, they crossed the word learning threshold by showing similar processing of high and low frequency words.
compared to the native participants.

In terms of organization of the mental lexicon, Murray and Forster’s (2004) rank model is partially supported by the results of the current experiment because the L1 and L2 groups did not differ significantly in the latency of N400 waveforms. Since the L2 speakers utilize a different orthography in their native language, it is believed that the L2 group has separate lexical bins of Chinese and English items, posing a possible extension of the rank model as discussed in Duyck et al. (2008). Duyck et al. (2008) showed large frequency effects for low frequency words in their L2 group and argued that bins with both Dutch and English contents accounted for the large effects. Murray and Forster’s (2004) model states that lexical entries in each bin share similar orthographic properties and entries are then ranked on a list in their frequency of occurrence within the language. Because of the top-down nature of the search process in a frequency-ordered list, low frequency entries are positioned in the bottom of the list and are encountered later. When L1 and L2 words of similar orthographic structure are contained within the same list, L2 words will be most affected by frequency effects because they will hold much lower ranks in the list as compared to L1 words.

Unlike the findings of Duyck et al. (2008), the current study found equivalent frequency effects in both groups; high frequency words elicited lower N400 amplitudes than low frequency words and both native and non-native groups recorded statistically similar N400 measures. Duyck et al. (2008) hypothesized that Chinese-English L2 learners would not produce large frequency effects because
within the rank hypothesis, they would have two separate bin systems since written Chinese is drastically different from the Latin-based writing system of English. When ranked according to familiarity, the English-designated bins of the L2 group would have to occupy much lower ranks than their native Chinese bins because entries in both languages cannot be integrated in any one bin. If this were true then word recognition of English words would yield greater cognitive difficulty leading to a longer search process and larger N400 peaks. These results were not found within the current study.

A possible explanation of this study's results is that visual presentation of English words to the non-native speakers may have triggered a shift to a search solely in the ranked English bins. Binder (2008) explained that neurons in the brain’s visual word form area become attuned to a word’s orthographic structure and develop an expertise in efficient lexical retrieval by eventually processing whole words given a high frequency of exposure. The marginal latency differences between both groups were not significant and show that neural processing of words in both frequency conditions was very similar in speed. This information leads one to believe that the L2 group was automatically able to map on to words in their English bins once they were exposed to the experimental stimuli.

A closer look at the L2 group shows that the short period between stimuli exposure and a search in the English bins may be affected by proficiency. The correlations suggested that as proficiency increases, the scanning period decreases. Within the L2 group, the N400 effect size was smaller with the more proficient L2
speakers. As explained by Binder (2008), it is assumed that the visual word form areas of the more proficient L2 speakers were more familiar with the orthographic structures of the experimental stimuli. N400 effect size of the higher proficiency L2 speakers was more similar to the L1 group than the lower proficiency L2 speakers. This trend towards native-like processing shows that the L2 group can overcome the challenge posed by different orthographic systems with higher proficiency in the L2.

Another possibility for the similar performance of native and non-native groups in the experiment may have been a consequence of sentential context inadvertently providing meaning. Although this is cannot be verified by this ERP study, the L2 group may have benefitted from sentential context to access meanings of critical words. A different ERP study could be designed to address this question. However since the design of the current study was neither one of lexical or syntactic comprehension, it is unknown whether the participants fully understood all critical words and the sentences in which they were embedded. Further information about comprehension could be collected if the lexical decision task administered at the end of current experiment included all experimental high and low frequency words along with distractors.

As for future studies, recruiting an L2 group with the same alphabet system as English to perform tasks of the current study would yield ERP results that can be compared to the behavioral results of the Duyck et al. (2008) study. Further, the proposed future study would also shed light on the rank hypothesis of Murray and
Forster (2004). If L1 and L2 words of similar orthographic structure are ranked accordingly in the same bin, then this should be reflected in the ERP measures as statistically significant N400 waves of larger amplitude and later latency peak for L2 words relative to L1 words.

In conclusion, the results showed that N400 effects of word frequency are similar in both L1 and L2 groups. The data supports McCusker's (1977) asymptotic learning model and Murray and Forster's (2004) rank model of lexical organization. The findings showed that semantic processing of native English speakers is similar to that of non-native English speakers.
APPENDIX A

Experimental Sentences

High Frequency Words

1. Mary loved her little brother so much that she spoiled him.
2. Mark went to the department for help with his problem.
3. Sue said that the service is bad at that restaurant.
4. The neighbor’s very loud telephone really bothered Barbara.
5. Ralph rested in the village before he started on his trip.
6. I had too much success very quickly and could not handle it.
7. I toured the famous building while I was on holiday.
8. The scientists created new technology for the military after years of research.
9. Tom said that the argument caused him much distress.
10. Rick opened up the chapter and looked for the information he needed.
11. Bob had a horrible experience that left him very weak.
12. John repaired the old machine very quickly.
13. Please bring me the material and a needle right away.
14. Melanie attended the lengthy practice yesterday afternoon.
15. Paul asked whether the process would take a long time.
16. Mark’s son received his education from a prestigious university.
17. We couldn’t repair the marriage even though we tried.
18. Clive doesn’t like studying science because he finds it very hard.
19. Sara rushed her sister’s husband to the doctor after he hurt himself.
20. Mark asked for some support when he was not feeling well.
21. Take money from the account and pay the debt.
22. Sandy got the star’s attention after the play was over.
23. My father loved his business and always enjoyed working.
24. John is a strange character and is not very trustworthy.
25. Beth wanted to study history at a college in Canada.
26. Pam picked up the picture off the dirty floor.
27. Kathy disliked the snobby president and refused to say hello.
28. Michael’s uncomfortable and difficult situation was the topic of many conversations.
29. Dr. Smith said that Dad’s trouble was only temporary.
30. Please clean the dirty surface before you put food on it.

Low Frequency Words

1. Mary loved her little terrier so much that she spoiled him.
2. Mark went to the pharmacist for help with his problem.
3. Sue said that the cuisine is bad at that restaurant.
4. The neighbor’s very loud accordion really bothered Barbara.
5. Ralph rested in the hammock before he started on his trip.
6. I had too much tequila very quickly and could not handle it.
I toured the famous catacomb while I was on holiday.
The scientists created new camouflage for the military after years of research.
Tom said that the splinter caused him much distress.
Rick opened up the tabloid and looked for the information he needed.
Bob had a horrible amputation that left him very weak.
John repaired the old trellis very quickly.
Please bring me the scissors and a needle right away.
Melanie attended the lengthy tutorial yesterday afternoon.
Paul asked whether the autopsy would take a long time.
Mark’s son received his doctorate from a prestigious university.
We couldn’t repair the ligament even though we tried.
Clive doesn’t like studying algebra because he finds it very hard.
Sara rushed her sister’s toddler to the doctor after he hurt himself.
Mark asked for some aspirin when he was not feeling well.
Take money from the satchel and pay the debt.
Sandy got the star’s autograph after the play was over.
My father loved his vocation and always enjoyed working.
John is a strange scoundrel and is not very trustworthy.
Beth wanted to study zoology at a college in Canada.
Pam picked up the mascara off the dirty floor.
Kathy disliked the snobby ballerina and refused to say hello.
Michael’s uncomfortable and difficult deformity was the topic of many conversations.
Dr. Smith said that Dad’s amnesia was only temporary.
Please clean the dirty platter before you put food on it.
REFERENCES


Hahne, A. (2001). What’s different in second-language processing? Evidence from


