WORD FREQUENCY EFFECTS ON LEXICAL SELECTION: EVIDENCE FROM A PICTURE–WORD INTERFERENCE (PWI) TASK

by

Sangeun Shin

Bachelor of Science, Ewha Woman's University, 2002

Master of Science, Ewha Woman's University, 2008

Submitted to the Graduate Faculty of

Communication Science and Disorders in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH & REHABILITATION SCIENCES

This dissertation was presented

by

Sangeun Shin

It was defended on

March 28, 2016

and approved by

Connie Tompkins, Ph.D., Professor, Communication Science and Disorders

Michael Dickey, Ph.D., Associate Professor, Communication Science and Disorders

Lauren Terhorst, Ph.D., Associate Professor, Occupational Therapy

Dissertation Co-Advisors: Katya Hill, Ph.D., Associate Professor, Communication Science and Disorders

Malcolm R. McNeil, Ph.D., Distinguished Service Professor and Chair, Communication Science and Disorders

WORD FREQUENCY EFFECTS ON LEXICAL SELECTION: EVIDENCE FROM A PICTURE–WORD INTERFERENCE (PWI) TASK

Sangeun Shin, PhD

University of Pittsburgh, 2016

Given that a high frequency (HF) word advantage exists in lexical processing, a question arises about the locus of that frequency effect. This locus may be important for AAC research and clinical practice to provide an empirical rationale for graphic symbol-based AAC interfaces.

This study's first specific aim was to identify whether word frequency affects lexical selection using a picture-word interference (PWI) task. Fifty healthy, monolingual, Native American English speakers, between 40 and 64 years of age, participated in the study. Response times (RT) for semantic, phonological, and mixed distractor conditions served as the dependent variable during a PWI task. Data were analyzed using generalized linear mixed models (GLMMs).

In all distractor conditions, participants named HF pictures significantly faster than low frequency (LF) pictures. This result revealed that the word frequency effect occurs not only with phonological encoding, but also with lexical selection and interactive processing between these two steps. This finding is at odds with the Discrete Two Stage (DTS) model that states that word frequency selectively affects phonological encoding.

The secondary aim was to determine whether the target item's frequency interacts with the distractor item's frequency. HF distractor words had a stronger interference effect on the

retrieval of target words than LF distractor words, resulting in a more delayed RT when naming the target pictures in semantic and mixed distractor conditions. However, an interaction effect was observed between target word frequency and distractor frequency only in the mixed distractor condition.

A third aim was to determine whether RT analyses provide a more sensitive measurement than error type analysis for healthy adults. No significant word-frequency effect or interaction was found for error type. A high rate of correct responses and the characteristics of the errors as the end product of inefficient word retrieval are considered as main reasons for this negative result. This finding supports the hypothesis that RT is a more sensitive measure than error type for indexing the inefficiencies that affect naming behavior in the PWI task for healthy adults.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AAC = augmentative and alternative communication

ABCD = Arizona Battery for Communication Disorders of Dementia (ABCD) (Bayles & Tomoeda, 1993)

ANOVA = analysis of variance

AoA = age of acquisition

CI = confidence interval

CPM = Coloured Progressive Matrices (Raven, 1962)

CRTT = Computerized Revised Token Test (McNeil, Pratt, Szuminsky, Sung, Fossett, Fassbinder, & Lim, 2015)

C-VIC = Computerized Visual Communication (Weinrich, McCall, Weber, Shoosmith, Thomas, & Katzenberger, 1993)

DTS = Discrete Two-Step (Levelt, 1989, 1992)

GLMMs = generalized linear mixed models

HF = high frequency

IA = Interactive-Activation (Dell, 1986; Dell & O'Seaghdha, 1991, 1992)

LAM = Language Activity Monitoring (Hill, 2010)

LF = low frequency

LRM = Language Representation Methods (Hill, 2010)

MAT = Matching Persons and Technology (Hill & Scherer, 2008; Hill, 2010)

MCDI = MacArthur Communicative Development Inventories (Fenson, Tomasello, Mervis, & Stiles, 1994)

MDES = minimum detectable effect size

ML = maximum likelihood

MRI = magnetic resonance imaging

N = number

NP = noun phrase

OD = Optimal Design (Raudenbush, Spybrook, Congdon, Liu, Martinez, Bloom, & Hill, 2011)

OR = odds ratio

PICA = Porch Index of Communicative Ability (Porch, 1981)(Porch, 1981)

PNT = Philadelphia Naming Test (Roach, Schwartz, Martin, & Grewal, 1996)

PWI = picture-word interference

PWA = people with aphasia

ReML = restricted maximum likelihood

RT = response time

SD = standard deviation

SE = standard errors

SNUG = spontaneous, novel utterance generation

SOA = stimulus onset asynchrony

VSDs = Visual Scene Displays (Dietz, McKelvey, & Beukelman, 2006)

WEAVER = Word-form Encoding by Activation and VERification (Levelt, Roelofs, & Meyer, 1999)

WFM = Word Fluency Measure (Borkowski, Benton, & Spreen, 1967)

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

A word retrieval deficit is one of the prominent symptoms of the disorder of aphasia (Goodglass & Wingfield, 1997; Laine & Martin, 2006; Raymer, 2005). For people with aphasia (PWA) who are suffering from chronic and severe language disorders, augmentative and alternative communication (AAC) technologies commonly are recommended in order to ameliorate the impact of the impairment and increase the effectiveness of communication in daily activity and social participation (Beukelman, Ball, & Fager, 2008; Fried-Oken & Granlund, 2012; Worrall, Rose, Howe, & McKenna, 2007). To assist many PWA's selection of words, graphic symbol-based systems are provided (Fox & Fried-Oken, 1996; Garrett, Beukelman, & Low-Morrow, 1989; Steele, Weinrich, Wertz, Kleczewska, & Carlson, 1989).

Despite the variety of graphic symbol-based AAC technologies available, AAC research has not explored the relatively preserved word retrieval ability for the high frequency (HF) words of PWA. Studies have reported that PWA showed higher performance on HF word retrieval despite their damaged retrieval process (e.g., Kittredge, Dell, Verkuilen, & Schwartz, 2008; Knobel, Finkbeiner, & Caramazza, 2008), which was similar to the performance of non-aphasic speakers (e.g., Jescheniak & Levelt, 1994; Navarrete, Basagni, Alario, & Costa, 2006). That HF words are produced faster and more accurately than low frequency (LF) words seems to support the notion that exploiting PWA's less impaired word retrieval for HF words could be used to improve PWA's AAC use rather than bypassing the function.

In order to determine whether the advantage for HF words will transfer to an AAC interface, understanding the locus where the word frequency effect occurs during lexical retrieval processing is essential. Since vocabulary on an AAC display is represented as graphic symbols, the approach is grounded in semantic representations. With the information about the locus of the word frequency effect, PWA's relatively less impaired lexical retrieval abilities for HF words can be predicted and used to support vocabulary selection for the AAC user interface.

However, an ongoing debate about the locus of frequency effect exists. The established view is that word frequency only influences the phonological encoding step, not the lexical selection step (Jescheniak & Levelt, 1994). On the contrary, the more recent view proposes that word frequency influences both phonological encoding and lexical selection in healthy adults (e.g., Caramazza, Costa, Miozzo, & Bi, 2001; Cuetos, Bonin, Alameda, & Caramazza, 2010; Gahl, 2008; Navarrete et al., 2006) and PWA (e.g., Kittredge et al., 2008; Knobel et al., 2008). Each of the viewpoints stemmed from the Discrete Two-Step (DTS) model (e.g., Levelt, 1989; Levelt, 1992) and the Interactive-Activation (IA) model (e.g., Dell & O'Seaghdha, 1991, 1992; Dell, 1986) respectively. Both models shared the notion that lexical selection and phonological encoding are distinct, serial ordered steps. However, DTS models argue that the steps are separate modules, while IA models propose that retrieval processing is achieved through bi-directional spreading of activation between the two representation levels. Thus, IA proposes that semantic information can affect phonological retrieval and phonological information can influence word retrieval.

Although some studies showed the word frequency effect on lexical selection for healthy adults (e.g., Caramazza et al., 2001; Cuetos et al., 2010; Jescheniak & Levelt, 1994; Navarrete et al., 2006), no study designed an experiment for examining the effect on whole lexical retrieval process including lexical selection, phonological encoding, and their interactive stage. This limitation makes it hard to extend the traditional view of the effect on phonological encoding to the effect on the higher step.

Another limitation of the previous studies involves the pathway of the effect. Few studies provided evidence on this question with one exception. Based on studies of error analysis for PWA, Kittredge et al. (2008) hypothesized that the effect on lexical selection might be due to the transmission of activated phonological nodes to the semantic representation level. However, an investigation focusing on mixed errors, that play a critical role in support of interactivity, was absent in their study. Thus, the evidence is inconclusive as to whether this indirect effect influences semantic encoding. If there is an indirect route as well as a direct route for lexical selection, a larger HF word advantage can be expected for PWA who mainly depend on graphic symbols due to concurrent impairment in phonological encoding.

Since the word frequency effect is closely related to the spread of activation as a function of time (Dell, 1986), using a sensitive measurement is important to examine the phenomenon. Some word frequency studies used reaction time to index naming homophones in healthy adults (e.g., Caramazza et al., 2001; Cuetos et al., 2010; Jescheniak & Levelt, 1994)(Caramazza et al., 2001; Cuetos et al., 2010; Jescheniak & Levelt, 1994). However, in the Kittredge et al. (2008) study, only accuracy was measured, which limits identification of the influence of word frequency.

Overcoming the limitations of previous studies, the current study aimed to identify the locus of the word frequency effect in order to provide an empirical justification for the use of a symbol-based AAC interface for PWA. For this purpose, the picture-word interference (PWI) paradigm was used with healthy adults to examine the effect of each retrieval process by including semantic, phonological, and mixed (semantically and phonologically related) distractors during a picture naming task. Findings of the current study are expected to extend the understanding of the nature of the word frequency effect for PWA based on the literature that both healthy adults and PWA share the same mechanisms of lexical access (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Rapp & Goldrick, 2000). In order to increase the sensitivity of measurement, both response time (RT) and error frequency were analyzed.

The specific aims were to investigate: 1) whether word frequency affects lexical selection during the PWI task; 2) whether the target item's frequency interacts with the distractors frequency during lexical selection; and 3) whether a difference exists between RT and response type when examining the frequency effect for healthy adults. In the following section, AAC technologies for PWA are compared to identify possible user interfaces that would enable PWA to fully use their advantage for HF words. Then, the characteristics of aphasia, types of naming errors, and the underlying mechanism are discussed under the framework of the two influential word retrieval models: DTS and the IA models. The focus will then shift to a debate on the locus of the frequency effect. Finally, studies related to the PWI task are reviewed.

2.0 BACKGROUND AND SIGNIFICANCE

2.1 BACKGROUND

2.1.1 AAC Treatment for People with Aphasia (PWA)

2.1.1.1 Primary, secondary, and tertiary features of AAC systems

The *Matching persons and AAC Technology* framework (Hill & Scherer, 2008; Hill, 2010) as shown in Figure 1, offers a principled approach to testing, evaluating, and selecting an AAC system. AAC components or features are organized or categorized as primary, secondary, and tertiary features of the technology or device under consideration. Hill (2010) emphasized that in order to design effective AAC systems and interventions, primary features should be considered first, thus giving importance to the language elements of a system.

Primary features relate to the system's functions to represent vocabulary and generate spontaneous, novel utterances. Language representation methods (LRMs) are the ways to represent words or messages on AAC systems by using a set of symbols or lexemes. Depending on the type of graphics, LRMs can be grouped into three types: single-meaning pictures, multi-meaning icons (semantic compaction), and alphabet-based methods. Single and multi-meaning may be photos, line drawings, color graphics, or animation; thus, literacy skills are not necessarily required. Since these symbols represent semantic features (e.g., RED, FRUIT, ROUND) of a target word (e.g., “apple”), lexical selection in the mind can be carried out in this type of AAC interface. To be specific, single-meaning pictures employ use of graphic symbols to represent one word or message. In other words, each word in this language method is represented by a single symbol. Thus, the most concrete and transparent symbols are commonly depicted by AAC developers for accurate identification of the corresponding target words. For this reason, many pages of symbols need to be organized to represent a large vocabulary, resulting in a demand of page navigation for selecting a target word. This representation method tends to have symbols organized in a semantically hierarchical structure or key words associated with a topic or activity (e.g., a target word ‘apple’ is under a symbol of ‘house’ → ‘kitchen’ → ‘refrigerator’ → ‘fruit’ in *Lingraphica*). AAC systems that are dedicated to PWA tend to be based on single-meaning pictures (Aftonomous, Steele, & Wertz, 1997; Shelton, Weinrich, McCall, & Cox, 1996; Steele et al., 1989; van de Sandt-Koenderman, Wiegers, & Hardy, 2005).

Multi-meaning icons or semantic compaction also uses symbols; but unlike single-meaning pictures, symbols on a single page can be used in a prescribed sequence to access a large vocabulary and each symbol can represent multiple words by using the metaphors that several symbols convey. For example, if a symbol of “apple” is sequenced with a symbol of “action”, a verb “eat” can be selected. When “apple” is followed by a symbol that represents an “adjective”, a word “hungry” can be selected. This LRM avoids the need for a large number of symbols to represent a large vocabulary and can thus avoid searching through several screen displays to select or retrieve words. However, the symbols sequences must be taught. The interface of such a system is designed to generate HF words as much as possible with a limited number of symbols.

Matching Persons and AAC Technology:

Primary, Secondary, and Tertiary Considerations of the Assessment

(Start at the top and work down.)

Primary Components		
Language Representation Methods	**Vocabulary**	**Methods of Utterance Generation**
Single Meaning Pictures Alphabet-Based Methods Semantic Compaction	Core Extended	SNUG (spontaneous novel utterance generation) Pre-stored sentences

Secondary Components		
User Interface	**Control Interface – Selection Methods**	**Outputs**
Symbols Navigation Automaticity Human Factors	Direct Selection *Keyboard, head pointing,* *Eye gaze* Scanning *Switches* *Physiological (EMG, BCI, etc.)* Morse Code	Speech Display Electronic / Infrared / Radio Frequency Data logging

Tertiary Components		
Peripheral and Integrated features	Training and Support	Telerehabilitation

Figure 1. Augmentative and alternative communication (AAC) primary, secondary, and tertiary components during the Matching Persons and Technology (MPT) process (from Hill, 2010, p.45).

Communication can be achieved by using alphabet-based methods that use orthography (letter-by-letter) and key stroke saving techniques such as using a keyboard or alphabet array. In particular, a high level of spelling is required for a user to formulate messages. Thus, the majority of PWA are not expected to benefit from this type of methods due to their impaired spelling (van de Sandt-Koenderman, 2004). A word prediction feature may be available with spelling but still good literacy skills are required. Word prediction presents the candidate with possible word choices based on selected letters. Predicted words may be grammatically related HF words, commonly used HF words, or personal words added to the dictionary once a user types in an initial letter of a target word.

Word frequency has received a lot of attention from AAC researchers for designing systems that optimize vocabulary selection and organization (e.g., Balandin & Iacono, 1999; Beukelman, Yorkston, Poblete, & Naranjo, 1984; Beukelman, Jones, & Rowan, 1989; Hill, 2001; Stuart, Beukelman, & King, 1997; Trembath, Balandin, & Togher, 2007). Some AAC literature uses the term “core” vocabulary to indicate highly-used words in daily communication, and the term “extended” vocabulary to indicate less frequently used words. The definition of the terms remains debatable, but is outside the scope of this paper. Studies show that HF words, consisting of about 450 to 500 words, make up approximately 80% to 85% of the words used in language samples, regardless of the conversational topic in the samples (Hill, 2010). Along with findings of high speed and accuracy of HF word retrieval by researchers in psycholinguistics (Alario, Costa, & Caramazza, 2002; Cuetos, Aguado, Izura, & Ellis, 2002; Feyereisen, Van der Borght, & Seron, 1988; Nickels & Howard, 1994; Oldfield & Wingfield, 1965; Schwartz, Wilshire, Gagnon, & Polansky, 2004; Wilshire, 2002), the HF advantage needs to be considered when developing AAC interventions. However, HF words are not necessarily organized well on currently marketed AAC systems despite the fact that the HF words used by speakers are consistent across cohorts and are a relatively small pool of words (e.g., Balandin & Iacono, 1998; Beukelman, McGinnis, & Morrow, 1991; Hill, 2010; Mein & O'Connor, 1960).

The third primary feature to consider is the methods of utterance generation that include of spontaneous, novel utterance generation (SNUG) and pre-stored messages. SNUG reflects the natural language process of being able to say exactly what you intend to say. Thus, this utterance generation tool plays a role of human being's phonological encoding that is required to produce selected lexical nodes in the mind. Accordingly, for AAC speakers who rely on graphic symbol-based interface, phonological encoding skills are not necessary. With only appropriate lexical selection through the AAC graphic symbols, a target word can be generated and integrated with other selected words to produce an utterance. Pre-stored messages are stored in advance based on what you think you want to say and are later retrieved. Thus, in an expected communication situation, for example expressing greetings or making an introduction at a meeting, pre-stored messages can be used effectively and efficiently. However, users easily encounter challenges in expressing exactly what they want to say or replying to unpredictable questions using pre-stored content (van de Sandt-Koenderman, 2004). Thus, the interactive nature of communication and social participation require the flexibility of SNUG or access to individual words to formulate a response.

The secondary components focus on the AAC hardware and access technology that consist of the user interface, control interface or selection methods, and the outputs. The user interface refers to the display where interaction between the user and the AAC system occurs. Today's user interfaces for PWA typically are a visual touch screen with graphic symbols or pictures used to represent words or message for the user to select. If a symbol's location on the system does not change and/or navigation among a variety of displays is not required to locate words/symbols, the degree to which users can become more automatic in word selection increases. However, most of AAC systems for PWA use single-meaning pictures with a limited number of locations, but a large number of words. Thus, navigation is inevitable and automaticity is difficult to achieve. For example, a small vocabulary of only 200 words would require 10 pages using a 20-location display.

Lastly, the tertiary components include issues of peripheral and integrated features, training and support, and tele-rehabilitation. These are critical components to achieving effective communication using an AAC system, since without training and support expected patient and family outcomes will not be reached. These issues are beyond the scope of this research study, thus, for further discussion, see Hill (2010).

2.1.1.2 AAC technologies for PWA and their limitations

Since the 1960s, graphic pictures and symbols have been used in AAC interventions from manual communication boards to today's computer-based technology (Jacobs, R., Ogletree, & Pierce, 2004; Kraat, 1990; Wallace & Bradshaw, 2011). Many studies reported the effect of symbol-based AAC interventions on PWA's communication (e.g., Aftonomous et al., 1997; Fox & Fried-Oken, 1996; Garrett et al., 1989; Koul & Harding, 1998; Steele et al., 1989; Weinrich et al., 1989). Symbol-based user-interfaces are occasionally recommended to PWA, because of the perceived difficulties with reading and spelling that occur with aphasia (Nicholas, Sinotte, & Helm-Estabrooks, 2011). However, despite a variety of AAC technology available, AAC treatment effects are still questionable (Jacobs et al., 2004). Some AAC programs designed specifically for PWA are described as follows with their limitations.

Portable Communication Assistant for people with Dysphasia (PCAD^{TM}) and the revised version of the PCAD, TouchSpeak (available at http://www.touchspeak.nl) have been developed especially for PWA (van de Sandt-Koenderman, 2004; van de Sandt-Koenderman et al., 2005; van de Sandt-Koenderman, Wiegers, Wielaert, Duivenvoorden, & Ribbers, 2007). This handheld device incorporates symbols, pictures, and text. Speech output is produced by using either digitized speech or synthesized speech. One of the emphases of the TouchSpeak is that this system allows the users to fill the system with the vocabulary that is personally relevant to each individual. Since photos, pictures, symbols, words, and sentences are organized in a hierarchical manner, users need to generate messages by navigating the hierarchical system. Individualization of AAC vocabulary and hierarchical vocabulary organization can give an impression of an effective and systematic AAC interface. However, these approaches are not necessarily based on theoretical and empirical evidence to support their effectiveness in daily communication settings. Further studies are required to test their effectiveness. In addition, there has been no consideration of the superior HF word retrieval in PWA in the process of AAC vocabulary selection and organization.

Visual Scene Displays (VSDs) are designed to complement the residual cognitive and linguistic ability of PWA by utilizing their intact episodic memory (Dietz, McKelvey, & Beukelman, 2006; Dietz, Thiessen, Griffith, & Peterson, 2013) (see Figure 2). Images such as photographs or pictures that have personal relevance to a PWA are incorporated on the user interface. This approach is believed to provide a context for a user and to give his/her communication partner(s) information to support communicative interaction (Beukelman, Hux, Dietz, McKelvey, & Weissling, 2015; McKelvey, Hux, Beukelman, & Dietz, 2008). However, use of VSDs avoids addressing the primary components of an AAC system by failing to consider the LRMs, HF and LF vocabulary and availability of SNUG. Rather VSD focuses on manipulating the user interface of the technology.

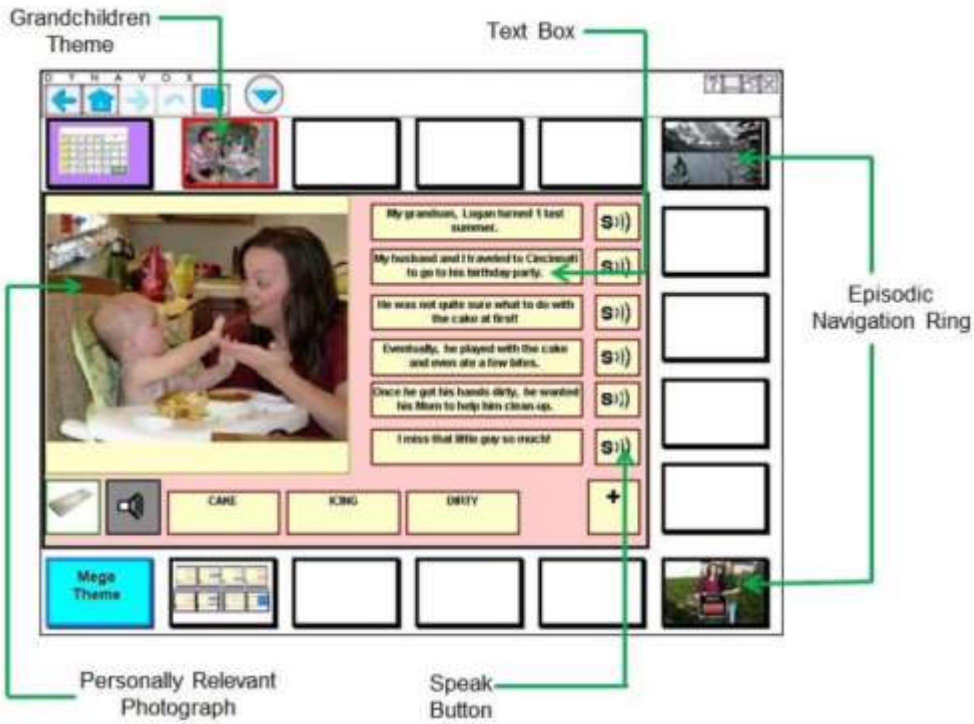

Figure 2. Display of Visual Scene Display (VSD) (from Dietz et al., 2013).
Note: Six pre-stored sentences relevant to the picture are shown. Other pictures and their corresponding sentences can be extracted by navigating the pictures.

Several application programs (Sutton, 2012) similar to VSDs are available online for PWA. Since personal pictures are used primarily, pre-stored utterances become context-governed. Current commercial applications that provide customizable visual scenes include Scene Speak (available at http://www.goodkarmaapplications.com) and Scene & Heard (available at http://www.tboxapps.com). Digital voice and popup text on the screen can be used for meeting functional communication goals. Pictello (available at http://www.assistiveware.com) provides a photo album that can be customized to help PWA share their experiences and stories with others by adding text and recorded sounds.

Applications such as VSDs, Scene Speak, Scene & Heard, and Pictello are so context-governed that they seem to be an effective communication tool by delivering personalized messages and life-experience content to a communication partner. Conversely, when the communication topics are unplanned, PWA face challenges due to the difficulties in finding an appropriate word or message on the system. In addition, most of the programs are based on pre-stored messages, so, interactive word-by-word formulation of messages fitting the situation, topic or preferences may be limited. Furthermore, this pre-stored approach often requires reading skills if one picture contains multiple written messages as in VSDs (see the "speak button" in Figure 2). These systems do not account for word frequency in their vocabulary representation since the unit of the utterance tends to be a phrase, a sentence, or even longer discourse to describe the picture. In addition, this approach fails to address the underlying language disorders of PWA.

Some less-context governed tools are applicable to many communication situations. Computerized Visual Communication (C-VIC) (Weinrich et al., 1993) allows a user to select picture symbols to form sentences. The spoken output is generated via SNUG or a recorded voice. The efficacy of the program has been reported by several studies (Aftonomous et al., 1997; Shelton et al., 1996; R. D. Steele, Kleczewska, Carlson, & Weinrich, 1992; M Weinrich, Shelton, McCall, & Cox, 1997; M. Weinrich et al., 1989, 1993; Michael Weinrich, McCall, & Weber, 1995). The sophisticated version of C-VIC is Lingraphica (The Aphasia Company™, Princeton, NJ) (available at http://www.aphasia.com). Some HF words are located on the main page but most words are organized in a semantic hierarchical order with many visual cues.

Other tablet based programs available as downloads from the internet such as Proloquo2Go (available at http://www.assistiveware.com) and TalkTablet (available at http://www.talktablet.com) also use single-meaning symbols to represent words. Thus, PWA can create messages by combining words or expressions to generate utterances. A few highly re-usable words such pronouns as “I” and “your” and verbs as “go,” “have,” and “see” are located in the display for easy access. Although a few HF words are available on a display for ready access, a PWA may still be required to search through displays to find a word in order to say what they want to say. Lingraphica, Proloquo2Go and TalkTablet were not designed to enable PWA to fully use their superior HF word retrieval.

Taken together, single-meaning pictures organized on the user interface providing the function of SNUG seem to be a more effective approach for PWA than the pre-stored message approach to accomplish interactive communication in real world conversational turn-taking. However, among the software programs that support SNUG, the word frequency issue is still in question for several reasons. Firstly, a trend in clinical service in the U.S. limiting AAC vocabulary/messages to addressing the user’s basic needs or medical necessity to meet the goal of functional communication is pervasive (Hill, 2010). Low frequency words seem to be emphasized in AAC systems due to high symbol transparency and perceptions of ease of use. In addition, low frequency word availability can purport individualization or customization of selected vocabulary without any theoretical and empirical evidence to support the clinical decision. However, the most serious reason would be the lack of understanding of the nature of aphasia in the area of AAC intervention and industry development. The characteristics of PWA’s word retrieval deficit, the mechanism of the word retrieval process, and the robustness of the word frequency effect on word retrieval have not been targeted in AAC research. In the following subsections, these topics are discussed in detail.

2.1.2 Aphasia and Error Patterns in Naming

2.1.2.1 Characteristics of word retrieval deficits in aphasia

McNeil and Pratt's definition (2001) provides some insight into the general mechanisms causing the defining characteristics of aphasia:

> *Aphasia is a multimodality physiological inefficiency with, greater than loss of, verbal symbolic manipulations (e.g. association, storage, retrieval, and rule implementation). In isolated form, it is caused by focal damage to cortical and/or subcortical structures of the hemisphere(s) dominant for such symbolic manipulations. It is affected by and affects other physiological information processes to the degree that they support, interact with, or are supported by the symbolic deficits.*

This definition highlights that aphasia is a *processing disorder of language*, not a disorder of linguistic knowledge (McNeil & Kimelman, 2001). In other words, there is no absolute loss of function. Instead, "inefficiency" across multiple language modalities is regarded as a cause of the deficits, which can selectively affect verbal symbolic manipulations (e.g. association, storage, retrieval, and rule implementation). For example, a patient with anomic deficits can be intact in identifying an object by pointing to it when failing to retrieve the name (Goodglass & Wingfield, 1997). Thus, word-finding difficulties in PWA are clearly distinguished from people with Alzheimer's disease whose failures are attributed to a loss of lexicon or profoundly inaccessible from semantic memory. Rather the problem for a person with aphasia is inefficient access to the target lexicon, but not that the target lexicon is erased or profoundly inaccessible from semantic memory (Chertkow & Bub, 1990).

2.1.2.2 Two steps of lexical access

Naturally occurring errors in normal speech and various error patterns in PWA have long been interpreted as a window into the mechanisms of lexical access (Blumstein, 1973; Dell et al., 1997; Goodglass & Wingfield, 1997; Laine & Martin, 2006; Lecours & Lhermitte, 1969). Impairments of lexical access is separable into two distinct steps (Harley, 1984; Jescheniak & Levelt, 1994; Kittredge et al., 2008): *lexical selection* and *phonological encoding*.

There are many terms to indicate the first step, *lexical selection* (Foygel & Dell, 2000; Levelt, 1992), which is synonymous with *word retrieval* (Schwartz, Dell, Martin, Gahl, & Sobel, 2006), *lemma access* (Levelt, Roelofs, & Meyer, 1999), *lemma retrieval* (Indefrey & Levelt, 2004), or *L-retrieval* (Rapp & Goldrick, 2000). Here lemma refers to holistic lexical representation with which grammatical information and syntactic frames are associated (Bock & Levelt, 1994; Dell & O'Seaghdha, 1991, 1992; Dell et al., 1997; Dell, 1986; Levelt et al., 1999; Levelt, 1989; Levelt, Schriefers, Vorberg, & Meyer, 1991). Thus, if some lexical factors have influence on the subject's performance in the grammar-relevant task, it may be inferred that the locus of the particular effect is at the lexical selection step, where the syntactic information is accessed and grammatical features are encoded (Jescheniak & Levelt, 1994; Navarrete et al., 2006). During lexical selection, semantic features of an intended word are activated at the semantic representation level and spread throughout the network, so that the lexical representation with the highest activation level is selected.

The next step is *phonological encoding* (Foygel & Dell, 2000; Levelt, 1992), which is synonymous with word form, lexeme retrieval (e.g., Butterworth, 1989; Dell, 1986; Fromkin, 1971; Garrett, 1975, 1976; Kempen & Huijbers, 1983; Levelt, 1989) or *phonological retrieval* (Kittredge et al., 2008; Schwartz et al., 2006). Phonological encoding starts with activation of the word unit that spreads to the phonological representation level through the network to select the highest activated phonemes. Note that some articles regard phonological encoding not as the second step of lexical access lexical but as sublexical access (e.g., Schwartz et al., 2006) or under a different concept of two steps (i.e., lexical and sublexical steps). However, the phonological errors that have word forms are categorized as lexical errors. Because most word frequency effect studies treated these errors as a phonological encoding step during lexical access, the two step assumption that is derived from lexical-sublexical levels will not be explored further in the current work.

Errors can occur during either step, because non-target units can become highly active at semantic and/or phonological representation levels if connected to activated target units. If a breakdown occurs during lexical selection, it may result in semantic errors. Semantic errors have the form of word substitution when two words share semantic features (e.g., *elbow* → “knee”, *green bean* → “asparagus”). If a breakdown occurs during phonological encoding, errors that are phonologically related to the target may be produced (e.g., e.g., *ankle* → “apple”, *train* → “tree”). Mixed errors that share both semantic and phonological features of the target word (e.g., snail → “snake”, *penguin* → “pelican”) is another error type. These errors are conceptualized as an interaction between the two steps serving as primary evidence to support the notion that these two levels or stages of word prediction interact directly with each other. This is at the core of the interactive activation model (Dell et al., 1997; Goodglass & Wingfield, 1997; Laine & Martin, 2006). If a word unit that does not share semantic or phonological features with the target word is highly activated, an unrelated error may occur (e.g., banana → drum).

There are other error types such as *non-word* errors that share phonemes with the target word (ghost → /goθ/); or neologisms, which are combinations of phonemes that differs greatly from the target sound (cane → /tΛ/) (Dell et al., 1997; Miller & Ellis, 1987; Schwartz et al., 2006). These errors can be categorized as sublexical errors. Unlike lexical errors which result from the selection of an incorrect word during lexical access, sublexical errors occur when non-target phonemes are selected, and those errors are more likely to be non-words. The existence of lexical-syntactic information supports the distinction between lexical and sublexical errors (Garrett, 1975, 1976). That is, lexical errors involve words of the same grammatical class. Thus, nouns replace nouns and verbs replace verbs, but not other syntactic category words (e.g., “Does it HEAR different?” to the target utterance of “Does it SOUND different?”). Sound exchanges tend to occur in words of different grammatical classes. For example, an error may be shown as “BLACK BLOX” for “BLACK BOX”), where a consonant from an adjective moves to a different grammatical category (noun). It indicates that word exchanges are constrained by grammatical features, which are a kind of lexical selection error, whereas sound exchanges are constrained by only phonological properties, not semantic or syntactic properties, which are a type of sub-lexical error. In the current study, these sub-lexical errors will not be dealt with in detail because the main interest lies in the word frequency effect on lexical retrieval processing.

Not only the evidence of errors, but also, the time course of activation for lexical representations provides empirical support for the existence of two steps in lexical retrieval. Schriefers et al. (1990) tested the serial processing to prove that access to a word involves an early step of exclusively semantic activation and a later step of exclusively phonological activation. They used a picture-word interference paradigm in which interfering concrete nouns were presented auditorily. Semantically related words and phonologically related words were presented to healthy adults during picture naming by manipulating the interval between picture and distractor onset (stimulus onset asynchrony [SOA]). The study revealed an inhibition effect with delayed RT compared to unrelated words for the semantically related words on picture naming latencies at an early SOA (- 150 ms). A facilitation effect was also seen, with reduced RT compared to unrelated words for phonologically related words at later SOAs (0 ms, + 150 ms). They concluded that there is a step of word retrieval where only its meaning is activated and that it is followed by a step where only its form is activated.

2.1.3 Word Retrieval Models

Several word retrieval models were proposed in the 1970s and 1980s. Those models are often classified as functional or connectionist models (Laine & Martin, 2006). Although they share a common assumption about the overall functional architecture of the two steps of lexical access, there is a discrepancy in the temporal coordination and the possible interaction between these two steps (e.g., Dell & O'Seaghdha, 1991, 1992; Dell, 1986; Levelt et al., 1991; Levelt, 1992; Schriefers, Meyer, & Levelt, 1990).

Many functional models (e.g., Butterworth, 1989; Fromkin, 1971; Garrett, 1975, 1976; Levelt, 1989) represent lexical and phonological components of a word as "boxes," which are retrieved independently of each other. Since there is no interaction between the two boxes, they argue that activation is transmitted from the preceding level to the next level, which is indicated by "arrows." Accordingly, functional models are called the "box and arrow" models. Alternatively, connectionist models have been more interested in the information flow between the “boxes” (Laine & Martin, 2006) than in the “box” itself. They argue that there is an interaction between the steps by providing evidence of the higher-than-chance occurrence of mixed errors and phonological coactivation of semantic competitors (e.g., Dell & O’Seaghdha, 1992; Dell, 1986; Harley, 1984; Martin, Dell, Saffran, & Schwartz, 1994; Plaut, 1996).

Among the many proposed functional and connectionist models, the focus for this research is on two currently influential models of the discrete steps (e.g., Levelt et al., 1999; Levelt et al., 1991; Levelt, 1992) and their interactivity (Dell & O’Seaghdha, 1991, 1992; Dell, 1986). These two models provide different perspectives on mixed errors and phonological coactivation of semantic competitors that can be contrasted to account for different findings regarding the locus of the word frequency effect.

2.1.3.1 Discrete two steps (DTS) model

Levelt (1989) proposed a blueprint for the speaker, which consists of a conceptualizer, a formulator, an articulator in a sequential order (see Figure 3). During the conceptualizer’s processing, preverbal messages (i.e., conceptual information) are generated. Then, a lexical node matched with the preverbal message is activated and makes its syntax available. Focusing on the formulator process, Levelt (1989, 1992) illustrated that at the lexical selection step, grammatical encoding takes a message as input and retrieves lexical nodes from the mental lexicon. Then a surface structure, a hierarchical organization of syntactic phrases consisting of *lemmas* (lexical representation), is delivered as output. Thus, the appropriate thematic role for all lexical nodes is retrieved.

The second subcomponent of the formulator, the phonological encoder, creates a phonetic plan for the selected lexical node. The major source of the information to be accessed by the phonological encoder is the lexicon’s information containing its morphology (e.g., two morphemes of a lexicon *dangerous*: a root *danger* and a suffix *-ous*) and phonology (e.g., the first of three syllables is *dan* /deɪn/). Finally, the articulators execute the phonetic plan by giving a series of neuromuscular instructions to motor systems such as the lips, jaw, tongue, and pharynx. Levelt (1989) claimed that through the connection between the production and the comprehension system, self-produced internal and overt speech could be monitored by the speaker.

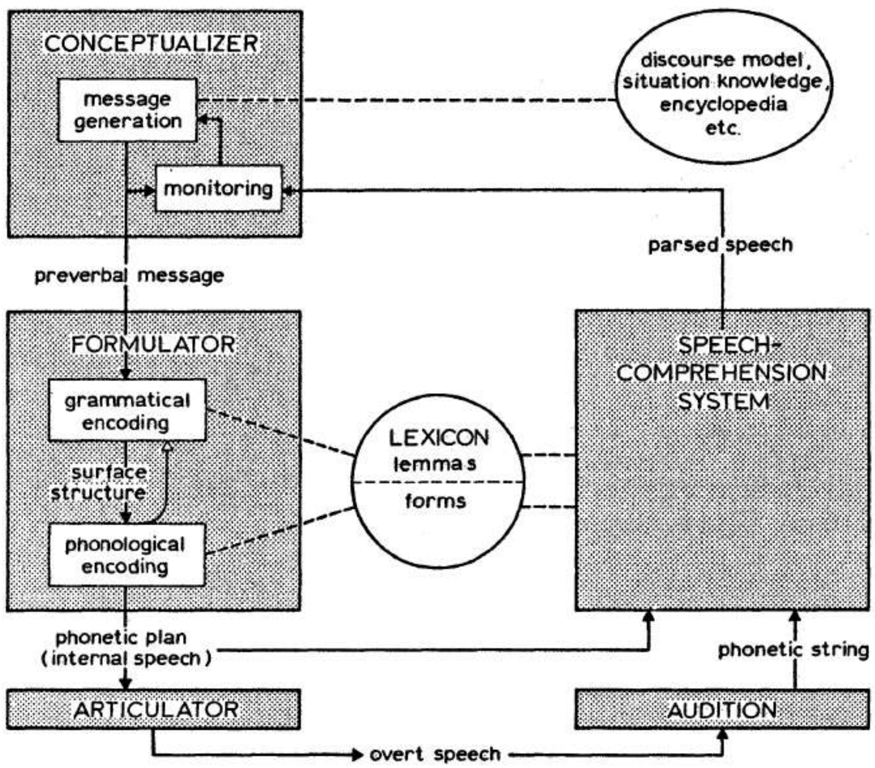

Figure 3. “A blueprint for the speaker” (from Levelt, 1989, p.9).

The basic idea about the two discrete and serial steps in Levelt’s model (1989, 1992) has been further elaborated in computational approaches in Levelt, Roelofs, and Meyer’s (1999) WEAVER ++ model (word-from encoding by activation and verification two plus model). In this model, the production of words follows a path from conceptual preparation to lexical selection, morphological encoding, phonological encoding and syllabification, phonetic encoding, and ends with articulation steps. Each step produces its own characteristic output representation such as lexical concept, lemma, morpheme, phonological word, phonetic gestural score (which are executed during articulation), and sound waves respectively (see Figure 4).

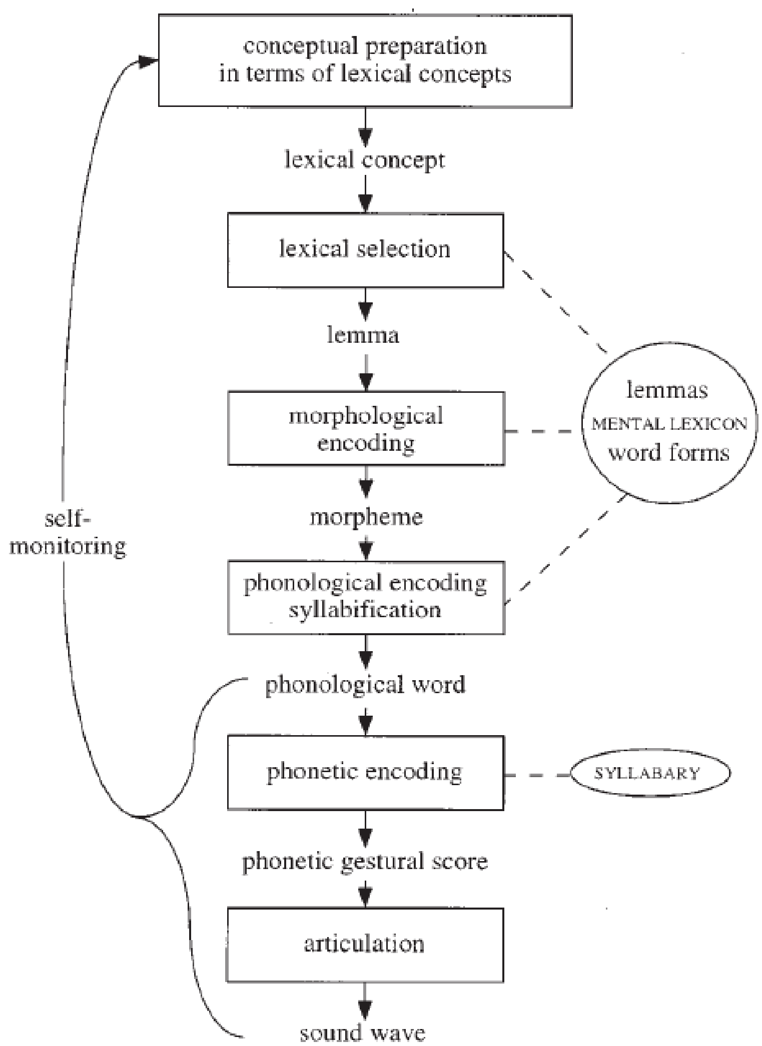

Figure 4. "The theory in outline" (from Levelt et al., 1999, p.3).

Levelt et al. (1999) describes three strata of nodes in the word retrieval network: the conceptual stratum representing lexical concepts, the lemma stratum representing lemmas and their lexical-syntactic information, and the form stratum representing morphemes, their phonemic segments, and syllable nodes as well (see Figure 5). For example, a concept node, ESCORT(X, Y), which stands for the meaning of the verb *escort*, is activated at the conceptual stratum. ESCORT(X, Y) is represented with a link between a concept and its superordinates, such as IS-TO-ACCOMPANY (X, Y). Here, IS-TO indicates the character of the connection: i.e., ESCORT (X, Y) IS-TO-ACCOMPANY(X, Y). The activation at the conceptual stratum spreads to the lemma stratum and the lexical node *escort* becomes activated. It is possible that other semantically related lexical nodes are activated at this level. However, Levelt et al. (1999)

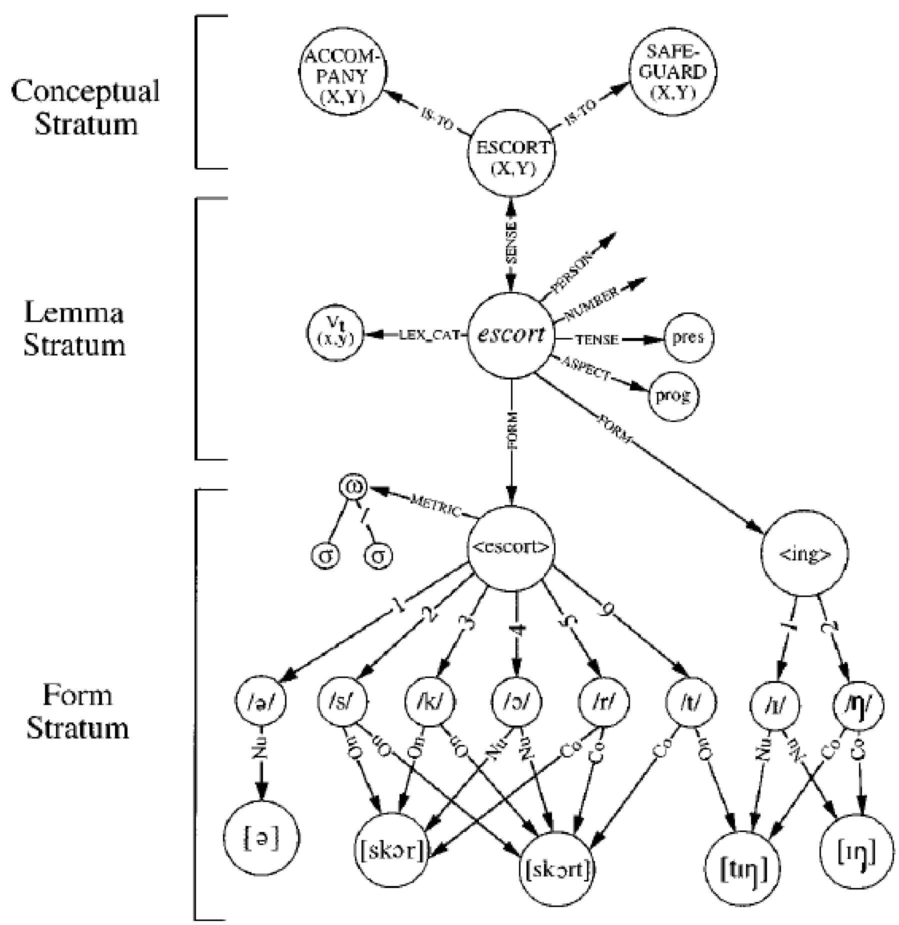

Figure 5. "Fragment of the lexical network underlying lexical access" (from Levelt et al., 1999, p. 4).

argued that the *escort* node receives the full proportion of ESCORT (X, Y)'s activation and thus the most highly activated lexical node of *escort* is selected. After lexical selection, its syntax becomes available for grammatical encoding. For example, the two argument positions (x, y), corresponding to the transitive verb *escort*, are valued. Other features are also valued during the grammatical encoding; verbs have features such as number, person, and tense, and nouns have number in the English language. Then its activation spreads to the third stratum, where the word-form stratum of the node *escort* activates the two morphemes, <escort> and <ing>. The segmental properties of these morphemes are "spelled out" as /ə/, /s/, /k/, /ɔ/ (sic /o/), /r/, /t/ for <escort> and /I/, /η/ for <ing>. Then, the links between segments and syllable program nodes specify possible syllabifications such as /ə/, /skɔr/ (sic /skor/), /tiη/.

DTS's perspectives on the phonological coactivation and mixed errors through Levelt and his colleagues' time-course experiments (Levelt et al., 1991; Schriefers et al., 1990) are crucial to supporting the fundamental features (i.e., the discrete, serial, and no interactive information processing) of the DTS models. Levelt and colleagues (1991) conducted a series of time-course experiments to determine whether phonological coactivation of semantic competitors exists. In their study, healthy adults participated in an acoustic lexical decision task, which also involved object naming (the dual naming-lexical decision task) in different SOA conditions. Participants named a line drawing as quickly as possible. They were also required to perform the lexical decision task with one of the four types of probes acoustically presented shortly after the presentation of a picture and before the naming response. The probe word could be identical to the picture name (e.g., sheep), semantically related (e.g., goat), phonologically related to the target (e.g., sheet), phonologically related to a semantic competitor (e.g., goal), or unrelated (e.g., knife).

The hypothesis was that during lexical selection, lexical decisions for the semantically related probe word should be delayed. Similarly, during the process of phonological encoding, lexical decisions for the phonologically related probe word should be delayed. If the semantic competitor was phonologically active, lexical decision latencies on the word “goal” should be delayed. The results showed that lexical decisions for probes either semantically or phonologically related to the target were delayed when the probes were presented shortly after picture onset. However, when the probe was phonologically related to a semantic competitor, no significant delay in response was found. Levelt and colleagues interpreted these findings as that the phonological activation occurred for a single target node only, so semantic competitors were activated at the semantic level, but their corresponding phonological forms were not activated. According to their logic, the mixed errors, which require the activation of a phonological node for the semantic competitors, cannot be explained under this theoretical framework.

The argument of the non-existence of phonological coactivation is illustrated in Figure 6 (a), where lexical selection is followed by phonological encoding only for the selected node. That is, during the first step, only the target node is selected for the concept among a set of one or more meaning-related nodes. During the second step, the selected target node becomes phonologically encoded, and then the articulatory plan is executed. Figure 6 (b) shows semantic activation during two types of phonological activation. For the first step, activation of the target node increases until the moment of selection, but drops to zero as soon as the second step begins for the selected node. There is no overlapping between semantic and phonological activation for the target node during the first and second steps. On the other hand, the phonological activation for the semantic competitor stays at zero. Under this framework, the Levelt (1989, 1992) and the Levelt et al. (1999) models posit that phonological encoding begins only after lexical selection.

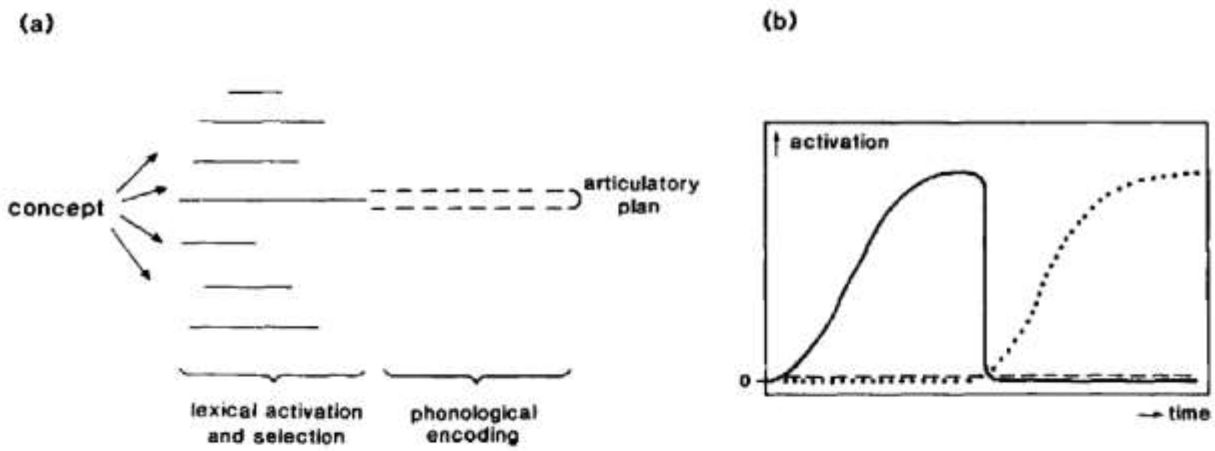

Figure 6. "The discrete two-step theory of the word retrieval process.
Note: (a) Steps showed that lexical activation and selection of one node is followed by phonological encoding step only for the selected node in the prior step. (b) Schematic diagram of the semantic and phonological activation levels for the target and other lexical node across the time course. Semantic activation of the target lexical node is indicated as [solid line] and the phonological activation level for the selected lexical node is indicated as [dotted line] activation. For the phonological activation for the semantic alternatives, they are all marked as [dashed line].)" (from Levelt et al., 1991, p. 124).

Regarding the mixed errors, DTS model views that the probability of a mixed error, *Pm*, is predicted as the product of the probability of a semantic error, *Ps*, multiplied by the probability of a phonological error, *Pp* (i.e., $Pm = Ps \times Pp$) because each error type occurs independently at different steps without interaction. Thus, it is postulated that the chances of a mixed error's occurrence are few in naturally occurring speech errors (Levelt et al., 1999). The DTS model handles these error patterns by adopting the post-lexical editing mechanism proposed by Butterworth (1989). The post-lexical editing mechanism uses lexical criteria for filtering outputs before the production steps of speech. It is assumed that the editing mechanism recognizes an error sensitively when the word error is only semantically related, only phonologically related, or non-related to the target word. Thus, if the erroneous word is both semantically and phonologically related to the target word, the editor tends to fail to recognize the error during the word retrieval.

2.1.3.2 Interactive-activation (IA) model

Like the DTS model, the IA model assumes that two steps are involved in word retrieval processing. However, the IA models' major difference from the DTS models is that bidirectional spread of activation between nodes at adjacent levels is allowed. This subsection delineates the characteristics and the word retrieval process in the IA models in detail and then discusses the phonological coactivation and uses mixed errors as key evidence to support the IA models' perspective on the interactivity.

Ignoring the articulatory/motor aspect of the IA model, the three levels that are involved in the lexical selection and the phonological encoding steps is the focus here (Dell & O'Seaghdha, 1991; Foygel & Dell, 2000): *semantic layer, lexical layer* and *phonological layer* (Figure 7). Dell (1986) argued that the nodes in each layer are connected to the nodes in the adjacent layers and their connections have certain weights. Depending on how strongly two nodes are connected, the value of their weight is determined. Thus, if the nodes have strong connection to each other with high weights, the activation of one node will strongly affect the activation of the other node. The transmission of the activation can also run in a forward (to the upper) direction or backward (to the lower representation level) direction. For example, an activated word *cat* spreads activation down to the corresponding phonological nodes /k/, /a/, /t/ and up to the connected semantic features for the *cat*. Since several semantically related lexical nodes such as *rat* and *dog* share the semantic features, those semantic competitors are activated at the lexical representation level as well. Then, the corresponding phonological node for each semantic competitor also becomes activated: e.g., the activation of /r/, /a/, /t/ for rat, and /d/, /o/, /g/ for dog. These activated phonological nodes also spread feedback to the preceding level to transmit the activation level to the corresponding lexical node. Because of this two-way spread

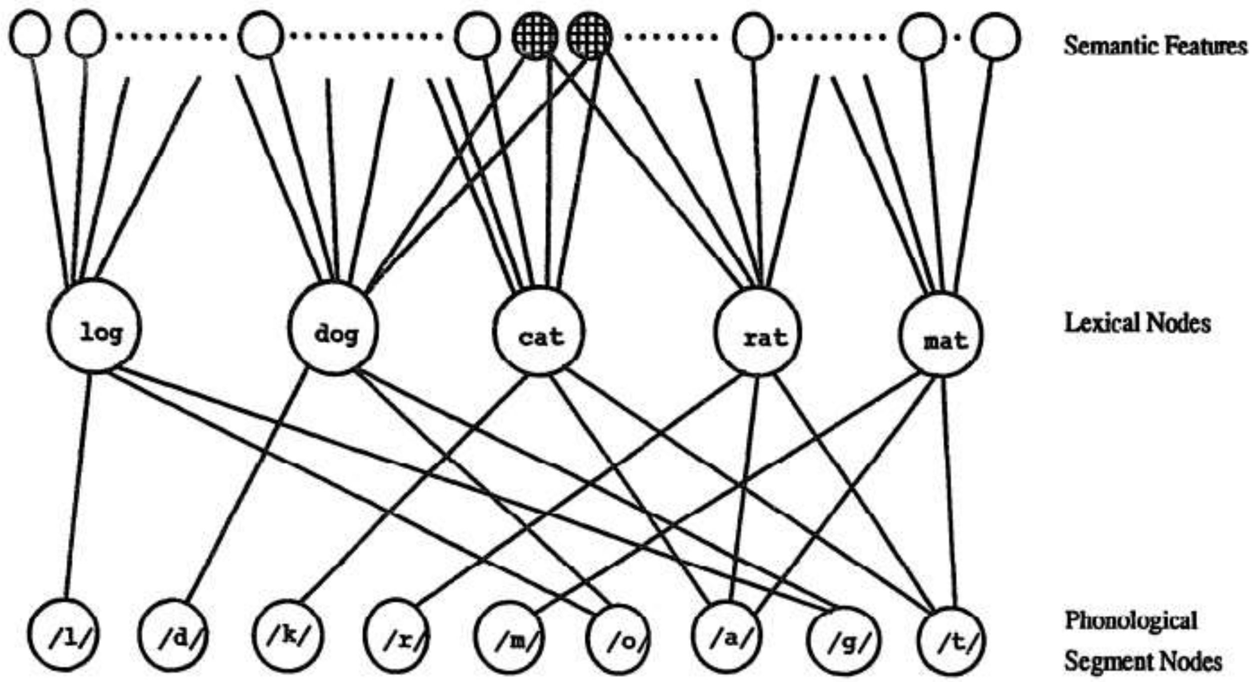

Figure 7. "Lexical network structure in the spreading-activation production model (from Dell & O'Seaghdha, 1991, p.605).
Note: In this figure, two semantic features (highlighted) are shared by the three lexical nodes "dog", "cat" and "rat."

of activation, two steps (i.e., lexical selection and phonological encoding) are not independent. In other words, each step influences the other, making the network highly interactive. Finally, the most active node at the lexical layer is selected and its grammatical encoding is conducted, such as the grammatical categorization (e.g., noun, verb), gender, number for nouns, etc.

The spread of activation has three components, *spreading*, *summation,* and *decay*, which are applied to all of the nodes (target or not) in the lexical network all the time (Dell, 1986). Unless the activation level of a node is zero, some portion of its activation level is sent to the connected nodes, which is called the *spreading* operation. The spreading activation levels of all connected nodes are added at the final destination node, which is the *summation* operation. The activated node does not hold that activation level. Instead, it decreases exponentially over time toward zero, which is the *decay* operation. These operations are expressed in an equation as follows.

$$A(j, t_i) = [A(j, t_{i-1}) + \sum_{k=1}^{n} p_k A(c_k, t_{i-1})](1 - q)$$

Here, A (j, t_i) stands for the activation level of node j at a particular time t_i and A (j, t_{i-1}) stands for the activation level of the same node j at a preceding time t_{i-1}. The c_1, c_2, c_3, … c_k are all nodes directly connected to node j, and the p_1, p_2, p_3, … p_k are spreading rates associated with the connections between c_1, c_2, c_3, … c_k and node j. Lastly, q means the decay rate and is always greater than zero but less than 1. If it is assumed that node j is the current node being translated into other nodes at the adjunct levels, the activation level of the current node can be calculated as the activation level of node j at a preceding time, plus the sum of all activated levels of connected nodes c_k with node j with spreading rate p_k, and multiplied by $(1 – q)$. Thus, in the case of node j at time t_i, if the decay rate q is near zero, the activation level of node j will be great. However, as the time t_i increases, the decay rate will become bigger and will result in decreasing the activation level of node j. Dell highlighted that the activation level depends on the values of p and q, and the structure of the network.

Regarding the time course of semantic and phonological activation, Dell and O’Seaghdha (1992) demonstrated that there is a distinct time course of activations between these levels, which is consistent with the results of Levelt et al. (1991). This finding supports the notion that the time course is globally modular and the processing is serial and forward. However, one important feature is that activation at a certain representation level is accompanied by top-down and bottom-up spreading activation with the pre- and the post-level’s representation. This feedforward and feedback contributes to some local interaction. Thus, as Figure 8 shows, when there is activation at the phonological representation level, some activation occurs on at the levels of prior and next representation levels. Dell & O'Seaghdha (1992) argued that this finding supports the contention that activation is exchanged between nodes at adjacent levels. This interactive activation between adjacent levels accounts for the mixed errors.

Thus, the perspective of the IA model on mixed errors differs from that of the DTS model. Activation of semantic and phonological information at the same time as well as the spreading interactive activation is the key mechanisms of the mixed effect as accounted for by the IA model. For example, the competitor word *rat* obtains activation directly from shared semantics of the target word *cat*, and from feedback from shared phonemes of the /cat/ and /rat/. It is possible that the top-down (i.e., feedforward) and the bottom-up (i.e., feedback) information processing give more activation value to the word *rat* and result in a much better chance of being selected and articulated as a mixed error rather than a purely semantic or phonological lexical error.

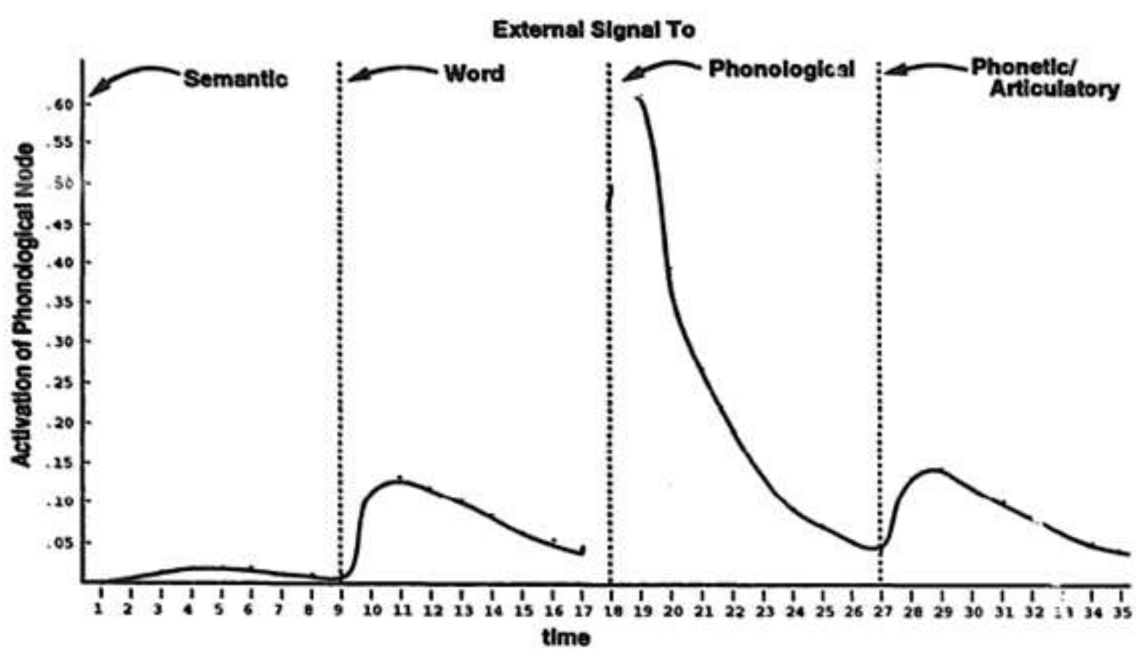

Figure 8. The activation of the phonological nodes as a function of time (from Dell & O'Seaghdha, 1991, p.612).

Since the IA model assumes that the semantic and phonological influences are not independent (Dell et al., 1997), the probability of mixed errors' occurrence becomes different from that in the DTS model where the chances are very slim. This was supported by several studies about normal speech errors (Dell & Reich, 1981; Harley, 1984) and experimental studies of picture naming tasks for non-aphasic speakers (Martin, Weisberg, & Saffran, 1989). In addition, Martin et al. (1989) found that mixed errors occurred more than those that were either just semantically or just phonologically related. In essence, these differences between IA and DTS models in terms of the phonological coactivation and chance of mixed errors result in a different perspective on the interactivity. Furthermore, these make a distinct argument on the locus of word frequency effect in the two steps that will be discussed in the following sections.

2.1.4 Effect of Word Frequency

2.1.4.1 Mechanism of word frequency in lexical access

Lexical nodes or phonological nodes are selected depending on the the highest activated node. Activation then spreads to the corresponding node in the next (e.g., Levelt et al., 1999; Levelt, 1992) or adjunct levels through interactive spreading activation (e.g., Dell & O'Seaghdha, 1991; Dell, 1986).

From the literature, Miozzo & Caramazza (2003) summarized two hypotheses for the psycholinguistic mechanism behind word frequency in lexical access: 1) the activation level hypothesis and 2) the selection threshold hypothesis. First, the activation level hypothesis suggests that HF words are processed *faster* because they have a higher level of resting activation, thus less activation is required to reach a selection threshold than LF words

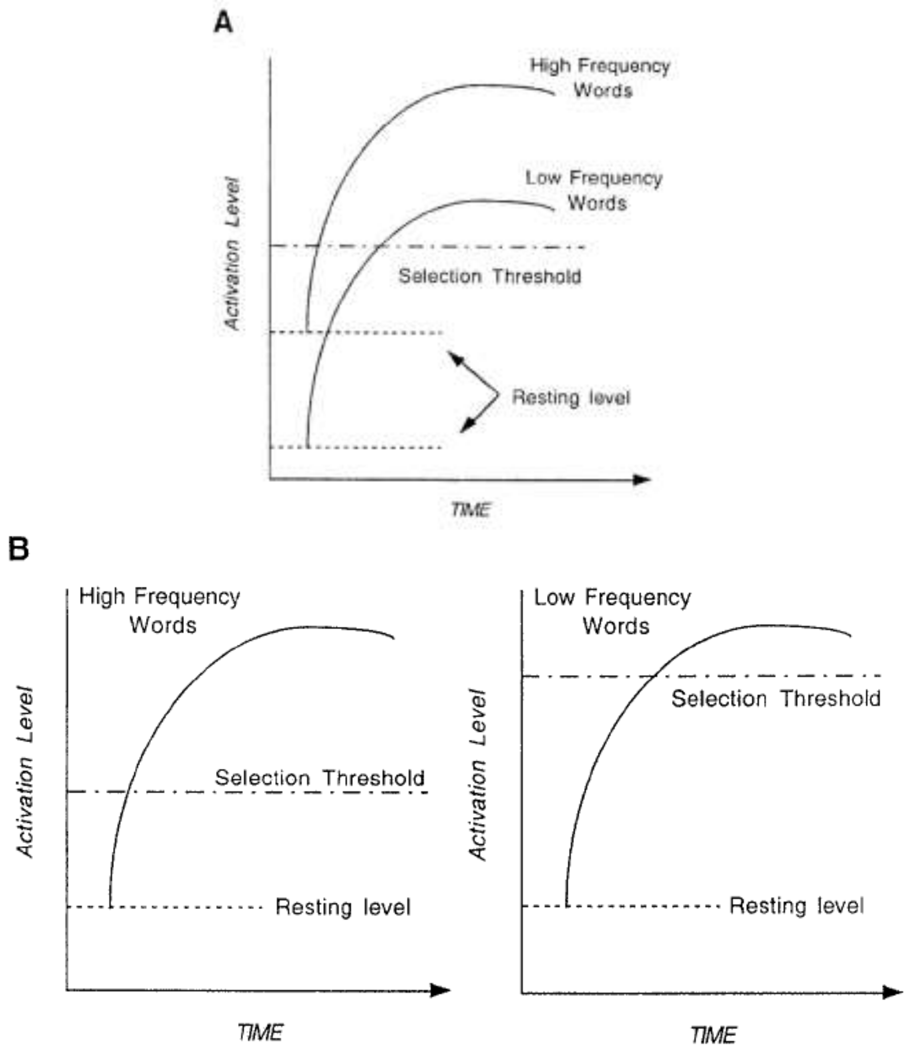

Figure 9. "Schematic illustration of the activation level and the selection threshold assumed by the activation level hypothesis (A) and by the selection threshold hypothesis (B) for high- and low-frequency words" (From Miozzo & Caramazza, 2003, p. 230).

(Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Goodglass & Wingfield, 1997; McClelland & Rumelhart, 1981) (see Figure 9A). The selection threshold hypothesis proposes that HF and LF words have the same resting activation level, but the selection threshold of lexical nodes differs depending on the word frequency. HF words have lower selection threshold than LF words (e.g., Jescheniak & Levelt, 1994; Morton, 1969) (see Figure 9B). Thus, with equal amounts of activation as input, levels of activation will be increased identically for both HF and

LF nodes. However, due to the lower threshold of HF words, those words reach the selection threshold faster than LF words, which will then have a high probability to be selected.

Although there is ongoing debate about which hypothesis accounts best for the word frequency effect, either hypothesis provides insight into the non-target words' frequency effect during naming behavior (Miozzo & Caramazza, 2003). According to the activation level hypothesis, if the non-target words are HF, they will interfere the target word's retrieval more than LF non-target words due to their highly activated level. Whereas according to the selection threshold hypothesis, there would not be a significant effect on word retrieval between two frequency types of non-target words. This is because the resting activation levels as well as the amount of activation change of HF and LF non-target words are equal. In the case of LF non-target words interfering more than HF non-target words, other hypotheses may be required. This will be discussed further in section 2.1.5.2.

Derived from these hypotheses, time is an important aspect of specific node selection. No matter how likely word frequency is to affect the activation level in reaching a selection threshold under the two different hypotheses, the dynamics of activation level are time-dependent. Under both hypotheses, HF words are selected faster than LF words. In this sense, RT is a more sensitive measurement than accuracy in word frequency effect studies, which may increase statistical power accordingly.

2.1.4.2 Evidence of word frequency effect

There is considerable evidence supporting the effects of word frequency on naming performance. Most studies have used a picture-naming task to explore the frequency effect in single word production. Oldfield and Wingfield (1965) first reported this effect. They conducted a study to examine whether the latency in naming was related to relative frequencies of the objects named in healthy adults. The names of thirty-six objects were collected from the Thorndike-Lorge Word List, which contained a frequency count derived from written English (Thorndike & Lorge, 1944). Participants were asked to name the objects as quickly as possible, as they were presented with simple black outline drawings. The onset of the spoken verbal responses (i.e., naming latencies) was recorded and measured. When analyzing the correlation between logarithm frequency and naming latencies, a linear relationship was found between them. That is, the higher the frequency occurrence, the shorter the latency.

The word frequency effect has been detected not only in non-aphasic speakers but also in PWA through picture-naming tasks (e.g., Cuetos et al., 2002; Feyereisen et al., 1988; Gordon, 2002; Nickels & Howard, 1994; Wilshire, 2002). For example, Cuetos et al. (2002) found that word frequency was one of the prominent lexical variables shown by the greatest number of Spanish aphasic patients who were having problems in word retrieval across diverse aphasic types. Seven different properties of target words were analyzed as predictors including visual complexity, object familiarity, imageability (i.e., the ease of generating an image when a given word is presented), animacy, age of acquisition, word frequency, and word length (i.e., in number of syllables in the object name). The word frequency data were derived from a corpus of written texts with two million words. In multiple regression analyses conducted on the naming accuracy scores, seven independent variables significantly predicted participants' naming, which accounted for 62% of the variance in naming accuracy. Along with age of acquisition, object familiarity, and visual complexity, word frequency made a significant independent effect on naming performance of the group as a whole and of many individual patients. The independent contribution of the word frequency factor was consistent with previous studies (e.g., Barry, Morrison, & Ellis, 1997; Ellis & Morrison, 1998). Cuetos et al. (2002) showed that several lexical properties such as imageability, animacy, age of acquisition, word frequency, and word length can affect word retrieval. Although word frequency had a significant independent effect on naming performance, an appropriate statistical way of examining the lexical variables' effects as covariates is necessary to increase the statistical power and alleviate confounding effects.

Beyond word-level production, very few studies have demonstrated the frequency effect in the production of multi-word utterances such as phrases or sentences. However, consistent findings of a robust effect of word frequency have been reported (Alario et al., 2002; Goral, Levy, Swann-Sternberg, & Obler, 2010). Alario et al. (2002) investigated the word frequency effect on adjective and noun retrieval during the production of phrases through two experiments. In Experiment 1, healthy adults were asked to name the pictures of common objects in the form of adjectival noun phrases (NPs) which consist of determiner + adjective + noun (e.g., "the blue kite") as fast and accurately as possible. Each NP was composed of HF or LF adjectives and nouns. Among 32 pictures, 16 were HF words (average: 174 occurrences per million in Francis & Kucera, 1982; range: 58–662) and 16 were LF words (average: 13 occurrences per million; range: 1–36). Each picture was presented in eight different colors, among which four had HF names (average: 136 occurrences per million; range: 85–169) and four had LF names (average frequency: 30 occurrences per million; range: 8–52) (Alario et al., 2002, p.306). To assess the reliability of the study, they replicated Experiment 1 in Experiment 2 where the procedures were identical, but the participant population and the picture stimuli did not overlap those of Experiment 1. In Experiment 2, 25 HF words (average: 148 occurrences per million; range: 46–662) and 25 LF words (average: 9 occurrences per million; range: 1–20) (Alario et al., 2002, p.309) were used. Participants' response latencies and error rates were analyzed and it was found that both noun and adjective frequencies significantly affected naming latencies. The latencies were shorter for the HF nouns and adjectives. There was no significant interaction between the two word categories. They interpreted these results as the *additivity* of the frequency effect of nouns and adjectives. The same results were found in Experiment 2 (see Figure 10). This experiment offered an important finding: the frequency effect for the noun of the NP was observed not only when the noun was in the first position of the utterance but also when it was in the last position of the utterance.

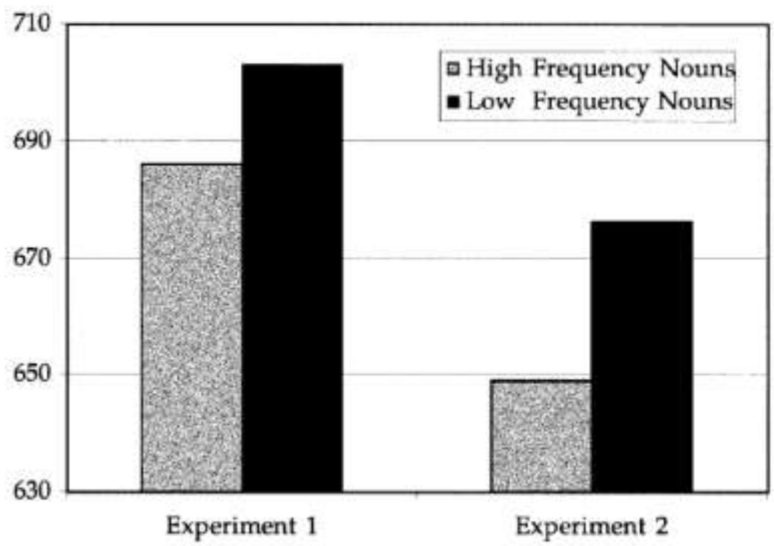

Figure 10. Naming latencies (ms) for the production of adjectival NPs in Experiments 1 and 2 as a function of the frequency of the noun. The data is collapsed for adjective frequency (from Alario et al., 2002, p.311).

However, caution is required when interpreting the results of the previous studies. They did not provide rationales for determining HF and LF words, which undermines the validity of setting up the independent variable—frequency. In particular, when words are categorized into HF and LF groups, there should be a valid principle that those words are discrete in terms of frequency. However, most of the studies did not address whether these two groups were discrete. In addition, the standard for defining HF and LF words varies across studies. For example, in the first experiment of Alameda and Cuetos (1995), the average of the HF nouns was 174 occurrences per million with a range of 58–662, and the average of the LF nouns was 13 with range of 1–36. It is not explained why a word with 58‰ is grouped as one of the HF words and a word with 36‰ is grouped as one of the LF words.

2.1.4.3 Debate on the locus of a frequency effect

Although there is general agreement that there is a strong word frequency effect on the phonological encoding step, it is still debated whether the lexical selection step is also affected by word frequency. This section will review the different findings and views on the loci of word frequency effects.

Jescheniak and Levelt (1994) hypothesized that the access to a lexical node's syntactic information, which includes its grammatical gender, is independent of phonological node activation. Thus, the activation of a word's grammatical gender can be used to determine the frequency effect on the lexical selection step. Before they tested the hypothesis, they first examined if robust word frequency effects existed during the noun retrieval process. Twelve native Dutch speakers participated in a picture-naming task. Twenty-four pictures represented HF words (those occurring more than 60 in 1 million words) and the 24 others represented LF words (those that occur less than 12 in 1 million words). All of these 48 experimental items were *de* words (masculine words in Dutch) and intermixed with 48-filler items which are *het* words (feminine words in Dutch). Each of the 96 items was presented three times to examine the word frequency effect through a repetition task and to examine the locus of word frequency effect through a gender decision task. The results showed that naming latencies for pictures with LF words was longer than naming latencies for pictures with HF words. Even after participants named the pictures three times, the size of the frequency effect remained consistent (see Figure 11). This finding supports the notion that relatively few repetitions do not influence the frequency effect observed in picture naming tasks with healthy adults.

Jescheniak and Levelt (1994) also examined the performances of Dutch speakers on a gender decision task to examine whether word frequency affects the lexical selection step, which required grammatical gender retrieval for the target pictures. They assumed that if there is a robust frequency effect on the grammatical encoding in the lexical selection step, HF words would reduce the response latency on the grammatical gender decision task. The Dutch-speaking participants were instructed to decide on the singular definite article that the name of the presented picture object should take. They pushed a *de* button for masculine and feminine words or a *het* button for neuter words. As a result, faster responses were observed for HF words than for LF ones the first time the pictures were presented. However, the effect quickly reduced in subsequent presentations (see Figure 12). The authors interpreted that this transient frequency effect is at odds with the nature of the robust frequency effect. They concluded that lexical selection step is insensitive to word frequency.

In order to investigate whether word frequency affects *only* phonological encoding, Jescheniak and Levelt (1994) adopted a homophone production task. Homophones have different word meanings, but have the same pronunciation. Theoretically, they are different at the concept level and at the lexical representation level, but share the word form at the phonological representation level. The authors hypothesized that if word frequency affects only the phonological representation level the LF homophone would be accessed just as quickly as its HF twin because the phonological representation is shared. However, if word frequency affects the lexical representation level, the LF homophone should be retrieved more slowly than the HF homophone. To test this hypothesis, Jescheniak and Levelt (1994) used the English-Dutch translation paradigm. Twenty native Dutch speakers who had good English skills were given a high frequency (e.g., *forest*) and a low frequency (e.g., *bunch*) English written word that

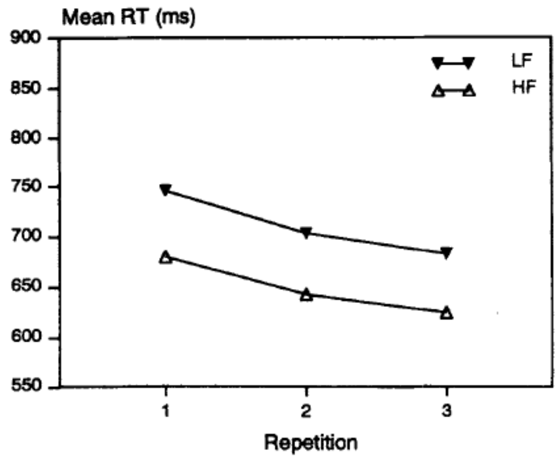

Figure 11. Picture naming latencies (from Jescheniak & Levelt 1994, p.829).

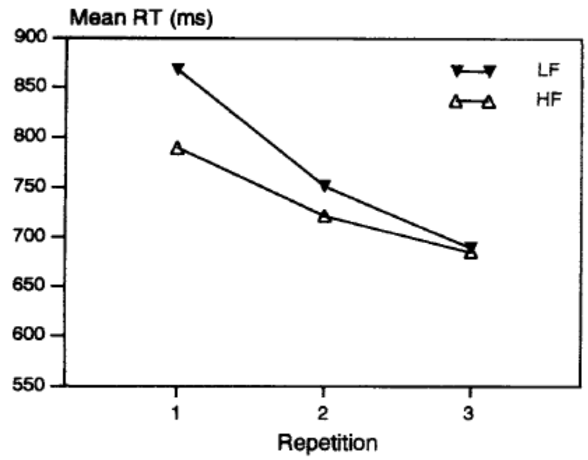

Figure 12. Gender decision latencies (from Jescheniak & Levelt, 1994, p.832).

appeared on the screen and they produced the Dutch translation (e.g., Dutch homophone *bos* for both *forest* and *bunch*). During the translation task, the phonological representation nodes of Dutch were accessed. Frequency-matched LF controls and HF controls were used to compare the response times. Note that the HF control was determined by its lemma frequency being roughly the sum of the lemma frequencies of the homophones (e.g., sum of frequencies of *bos* [bunch] and *bos* [forest]). If the participants translated the LF homophone as quickly as the HF control, this finding would support the hypothesis that word frequency operates in the phonological representation level instead of the lexical representation level. Alternatively, if the participants translated the LF homophone as slowly as LF control, it would support that lexical selection depends on word frequency. The results showed that the LF homophones were as fast as the HF controls and faster than the LF controls (see Figure 13). The authors termed this phenomenon as the homophone effect and argued that this finding of the LF homophone behaving like the HF homophone twin supports that the locus of frequency effect is at the phonological encoding step rather than at the lexical selection step.

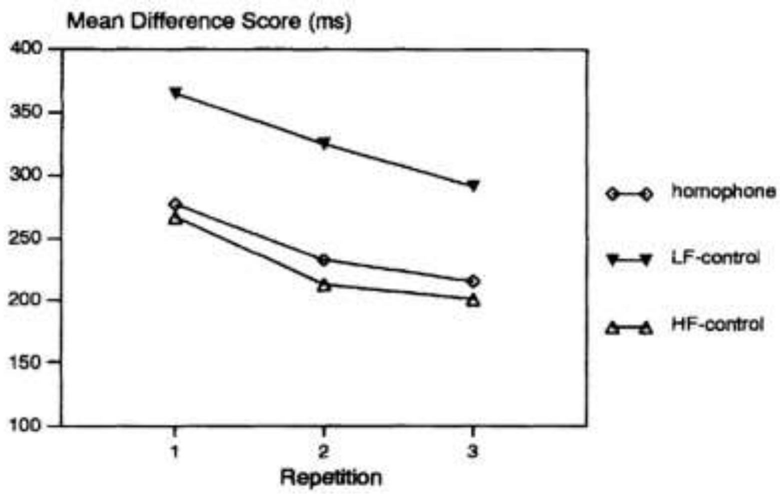

Figure 13. Difference scores from Experiment 6 (from Jescheniak & Levelt (1994), p. 638).
Note: LF = low frequency; HF = high frequency.

The findings of Jescheniak and Levelt (1994) have been challenged by Gahl (2008). Gahl investigated the homophone frequency effect on the response durations by analyzing roughly 90,000 tokens of homophones in the Switchboard corpus of American English telephone conversations. She found that retrieval times for HF words (e.g., time) were significantly shorter than those for the LF homophone twins (e.g., thyme). Caramazza et al. (2001) investigated whether naming latencies for homophonic words (e.g., *nun)* are influenced by the frequency of a specific word (i.e., that of *nun)* or a cumulative-homophone frequency (i.e., the sum of the frequencies of *nun* and *none)*. Thirty native English speakers named three sets of pictures and their naming latencies were measured through: 1) homophone-name pictures; 2) pictures matched to specific word frequency; and 3) pictures matched to cumulative-homophone frequency. The results showed that there was no difference in naming latencies between the homophone-name pictures and the pictures matched to specific word frequency. Participants' naming latencies were slower than the pictures matched to cumulative-homophone frequency. This finding is at odds with the homophone effects previously found by Jescheniak and Levelt. Rather, naming latencies for homophonic words turned out to be determined by their specific-word frequencies. In other words, the naming latencies are determined by the frequency of the pairing of a specific semantic representation and a word form, rather than the surface frequency of that particular string of sounds. The same results were found in a different language using Spanish-English translation task that replicated the Jescheniak and Levelt's (1994) experiment.

In addition, Cuetos et al. (2010) examined whether picture-naming latencies of homophones were predicted better in the picture-naming task based on cumulative homophone frequency than on specific-word frequency. Their finding was consistent with Caramazza et al. (2001), Shatzmana and Schiller (2004) and Gahl (2008), which supports that naming of homophones does not benefit from the shared phonological representation of HF homophone twins. Taken together, these findings support the interpretation that the frequency effect has its locus at the level where the different members of a homophone set are distinguished from each other—at the lexical selection step.

Jescheniak and Levelt's (1994) vanishing frequency effect has been also challenged. Navarrete et al. (2006) conducted two experiments to determine whether lexical selection is affected by word frequency. In Experiment 1, sixteen native Spanish speakers were asked to generate an utterance as fast and accurately as possible by using the sentence structure of "Este es nuevo" (This$_{masc}$ is new$_{masc}$) or "Aquella es vieja" (That$_{fem}$ is old$_{fem}$) when a picture was presented. To avoid the massive repetition of the same utterance pattern, two semantic features of the objects were adopted: distance (close/far) and appearance (new/old). Figure 14a depicts a car that is far from the participant's point of view and new. Figure 14b depicts a car that is close to the participant's point of view and old with a blurred image. In Experiment 2, participants undertook a gender decision task for the same pictures. They decided whether the picture's name was masculine or feminine, while ignoring the object's distance and appearance. In both

Experiment 1 and 2, retrieving phonological representation was not required, but retrieving lexical representation was necessary to process the grammatical encoding for the target nouns in order to decide the gender for the nouns.

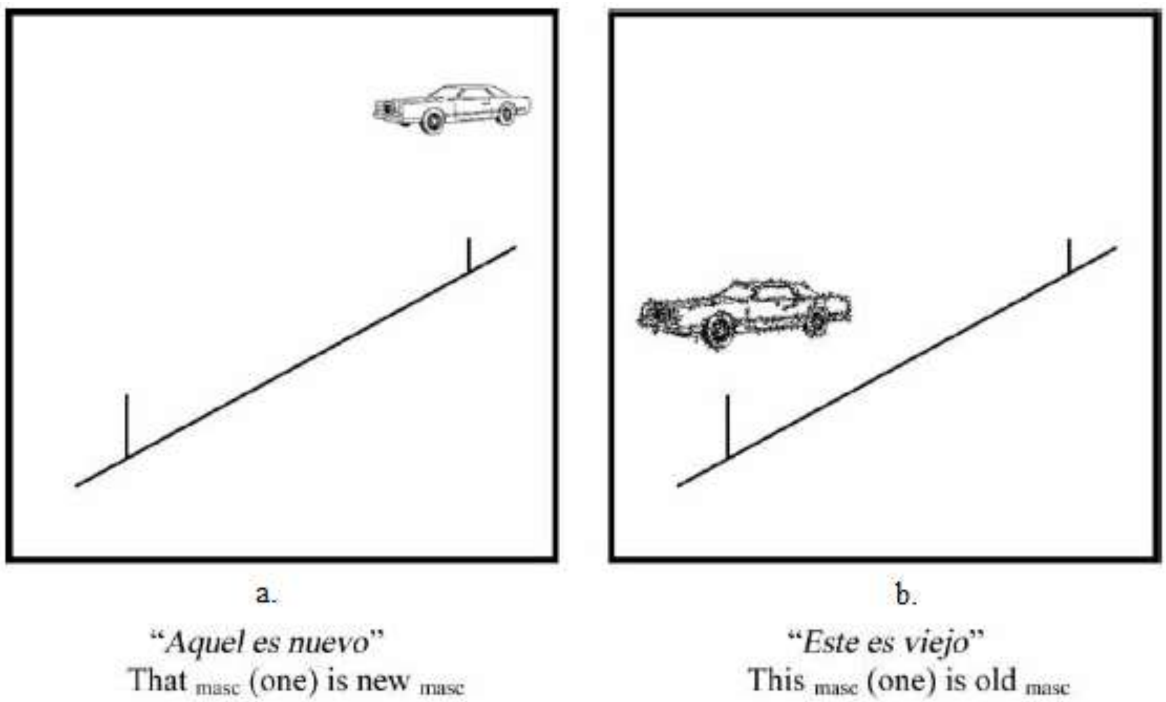

Figure 14. Two examples of the picture stimuli in Experiment 1 (from Navarrete et al., 2006, p. 1684).

Word frequency repetition effects were significant in terms of the error rates and response latency. That is, lower error rates and shorter response latencies were found for the HF pictures than for the LF pictures. Frequency and repetition did not significantly interact with each other. The results of Experiment 2 revealed that there was a robust word frequency effect in the gender decision task over repetition (see Figure 15). Counter to Jescheniak and Levelt's (1994) interpretation, this finding indicates that the grammatical encoding is influenced by word frequency, which implies that lexical selection is sensitive to the frequency effect.

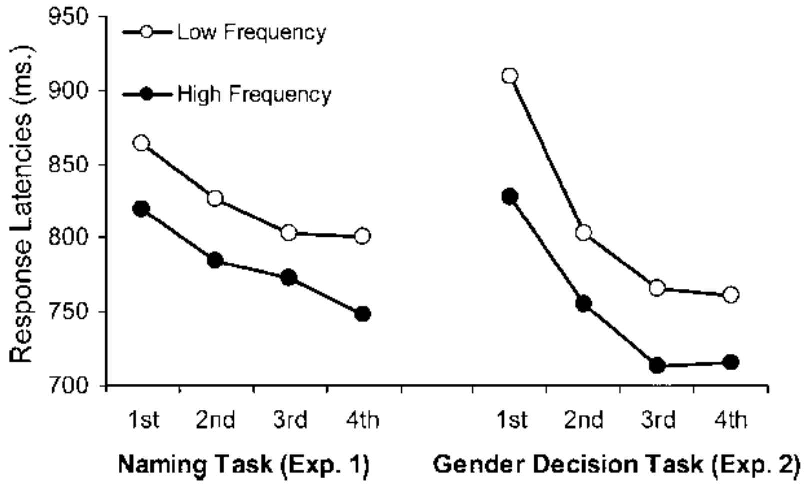

Figure 15. "Average response latencies for high-frequency and low-frequency sets broken by repetition in Experiments 1 and 2" (from Navarrete et al., 2006, p. 1686).

Unlike other studies where healthy adults were recruited to determine the locus of the frequency effect (Caramazza et al., 2001; Cuetos et al., 2010; Gahl, 2008; Jescheniak & Levelt, 1994; Navarrete et al., 2006), Kittredge, Dell, Verkuilen, and Schwartz (2008) investigated the locus of the frequency effect through the error analysis of fifty individuals with aphasia using a picture-naming task. Drawing on Foygel and Dell's (2000) model where the frequency of the target word is responsible for the weights of either its lexical-semantic connections or its lexical-phonological connections, Kittredge et al. (2008) hypothesized that if frequency only affects phonological encoding, it should be responsible for the incidence of phonological errors. They also argued that if frequency also affects lexical selection, it should be responsible for the incidence of semantic errors, and if it affects both steps of word retrieval, both error types would be predicted. The analyzed error types and error categories are described in detail in Table 1. Note that Kittredge and colleagues focused on three main error categories: semantic, phonological (collapsing across formal and non-word responses), and omission (collapsing across "no response" and "description" omissions). Mixed, unrelated, and miscellaneous errors, categorized as "other", were not analyzed statistically.

Table 1. Error types for observation and error category for data analysis in Kittredge et al. (2008)

Error type	Description	Example (for *pineapple*)	Error category for data analysis
Semantic errors	Synonym of the target or coordinate, superordinate or subordinate member of its category	*apricot*	"Semantic errors"
Formal errors	Any word response that meets the PNT's phonological similarity criterion	*pillow*	"Phonological errors"
Mixed errors	Response that meets both semantic and phonological similarity criteria	*banana*	"Other"
Unrelated	Response that meets neither semantic nor phonological similarity criteria and is not visually related to the target	*gun*	"Other"
Non-word	Neologism that is not also a blend, which either did or did not meet the PNT's phonological similarity criterion	*pineme or fepe*	"Phonological errors"
Description	Description of the target, often with semantic or phonologically relevant information	*you eat it*	"Omission"
No response, omission,	Null or semantically empty response	*I don't know*	"Omission"
Miscellaneous error	Other types of errors including visual errors		"Other"

The results revealed that HF words resulted in fewer word retrieval errors. Although the frequency effect was strong in the phonological errors as demonstrated in previous studies (e.g., Feyereisen et al., 1988; Goldrick & Rapp, 2007; Schwartz et al., 2004), the lower-frequency words resulted in more semantic errors and errors of omission as well. This indicates that word frequency influences multiple steps of the word retrieval process, including lexical selection and phonological encoding.

There are several notable points in the Kittredge et al. (2008) study. First, a large sample of patients was analyzed, which increased the external validity. Second, a multinomial logistic regression model enabled a statistical analysis of error data from heterogeneous patients. Third, this study confirms that the error analysis can be used to determine the locus of the frequency effect. As reviewed, one of the primary sources used in developing the two steps of the word retrieval models (i.e., lexical selection and phonological encoding) stemmed from speech errors of PWA. In summary, by tracking the errors, the damaged step(s) affected by word frequency in the word retrieval process can be determined.

However, the Kittredge et al. (2008) experiment was not designed to answer how word frequency affects the first step. They assumed that the frequency effect on the first step was due to the spreading activation along the bidirectional connections linking lexical representation retrieval and phonological representation retrieval. In order to attribute the reason to the interactive activation between two steps, they would have had to include mixed errors in their error analysis or use alternative methods. If there are enough mixed errors obtained to compute statistical analyses, the interactivity can be tested since mixed errors are the primary evidence to support the IA model, as previously reviewed.

It may be hard to detect a frequency effect on lexical selection by measuring only the number of semantic or phonological errors as in Kittredge et al., (2008). In order to identify the word frequency's influence on the retrieval process as a temporal measure, examining RTs is necessary in naming tasks where the duration to reach the selection threshold level and/or the resting activation level is involved.

2.1.5 Picture-Word Interference (PWI) Paradigm

2.1.5.1 Rationales to use PWI task

Based on the prominent characteristics of aphasia addressed in the definition of aphasia (McNeil & Pratt, 2001), inefficiency in accessing the appropriate representation node in the lexical retrieval process is expected to manifest in more errors and delayed response time for PWA. If inefficient access affects the duration of retrieving words, measuring RT will provide evidence concerning the loci of language processing deficits with temporal sensitivity (Rogers, Redmond, & Alarcon, 1999). Also, the time-dependent nature of the word frequency effect mechanism supports the importance of measuring RT.

The PWI paradigm can be applied in order to detect the frequency effect at different lexical processing steps. The PWI, a variant of the Stroop task, is a widely used technique for exploring the effects of the semantic context on lexical access (Sailor, Brooks, Bruening, Seiger-Gardner, & Guterman, 2009). Since PWI tasks provide information about the time course of production, it may help to understand the locus of the word frequency effect while naming by showing the separate contributions of the various stages of word production (Dell & O'Seaghdha, 1992; Hashimoto & Thompson, 2010; Levelt et al., 1991; Wilshire, Keall, Stuart, & O'Donnell, 2007).

Pictures are presented to a participant along with written or spoken word distractors in the PWI task. By manipulating the SOA between picture and distractor, the nature of the two stages of lexical access in speaking can be experimentally investigated. This is based on a general finding that semantic interference (i.e., semantic inhibition) occurs at a negative SOA (e.g., SOA of -150ms) whereas phonological effects (i.e., phonological facilitation) occurs at a positive SOA (e.g., SOA of 0ms and +150ms) (e.g., Damian & Martin, 1999; Schriefers et al., 1990). In this task, participants are asked to ignore the distractor words and to name the pictures as quickly and accurately as possible (Damian & Martin, 1999; Rosinski, 1977; Schriefers et al., 1990; Starreveld & La Heij, 1996).

According to the literature, PWA show similar semantic and phonological results as healthy adults on the PWI task. This suggests that the PWI task allows for estimating the time course of activation across the lexical process steps for control groups as well as for PWA (Hashimoto & Thompson, 2010; Rogers et al., 1999; Wilshire et al., 2007). In particular, Hashimoto and Thompson (2009) showed that there were similar time courses of activation—earlier semantic inhibition and later phonological facilitation—across lexical processes for PWA and healthy adults. Not only was there an increased number of errors for PWA compared to the control group, but increased RTs in naming were also found when semantic and phonological written distractors were presented to PWA. Further, similar characteristics of errors (e.g., high rates of NRs, semantic paraphasias, and semantic-competitor errors) were observed for both groups during the PWI task. In a situation where a consensus on the locus of the frequency effect was not made for the healthy adults, the findings of Hashimoto and Thompson (2009) suggests that the locus of the frequency effect in the PWI task informs understanding of the nature of the word frequency effect in PWA as well as healthy adults.

In spite of the PWI tasks benefits, the distinct time course is less transparent when using written word distractors. For example, (Glaser & Düngelhoff, 1984) showed a semantic interference effect with the SOA around 0ms during the PWI task with written word distractors. When the same type of distractors (i.e., written words) was used in Starreveld and La Heij (1996), the phonological facilitation effects were found with SOA ranging from -200ms to +100ms which covers the SOA = 0 for the semantic interference effect of Glaser and Düngelhoff

(1984), when the same type of distractors (i.e., written words) was used. Damian & Martin (1999) supported this distractor type sensitivity of RTs in the PWI task. With visual distractors, semantic interference appeared with the SOA ranging from 0ms to +200ms and the phonological facilitation effect appeared with the SOA ranging from -200ms to + 100ms. However, when spoken distractors were presented, a semantic interference effect appeared at earlier SOAs and a phonological facilitation effect at later SOAs (see Figure 16).

Given that "in the picture-word interference procedure, a longer processing duration of the distractor word translates into a larger negative SOA at which an effect can be obtained" (Damian & Martin, 1999, p. 5), visual distractors may have more rapid access to their semantic nodes than auditory distractors and result in a smaller negative SOA for semantic distractors. Damian & Martin attributed this discrepancy between two distractor types to the parallel processing of visually presented distractors, which is not the case for auditorily presented distractors. Therefore, semantic effects appear at an early or negative SOA due to the longer processing duration of auditory distractors. Considering that the purpose of this study is to identify the frequency effect on lexical selection, using spoken words as distractors will allow the demonstration of an inhibition effect at an early time point with a phonological effect appearing at a later time point by avoiding the parallel processing inherent with visual distractors.

One might question whether RTs for the condition of mixed distractors can be used to examine the interactivity. The theoretical rationale for this originates from Damian & Martin (1999) where interactivity was identified by using mixed distractors. They hypothesized that with the DTS model; RTs for the mixed distractors should yield the same semantic interference effect as semantically related but phonologically unrelated distractors at a negative SOA. Further, they should yield facilitation effects comparable to those obtained from phonologically related but semantically unrelated words at a positive SOA. Conversely, these patterns are not predicted from the IA model.

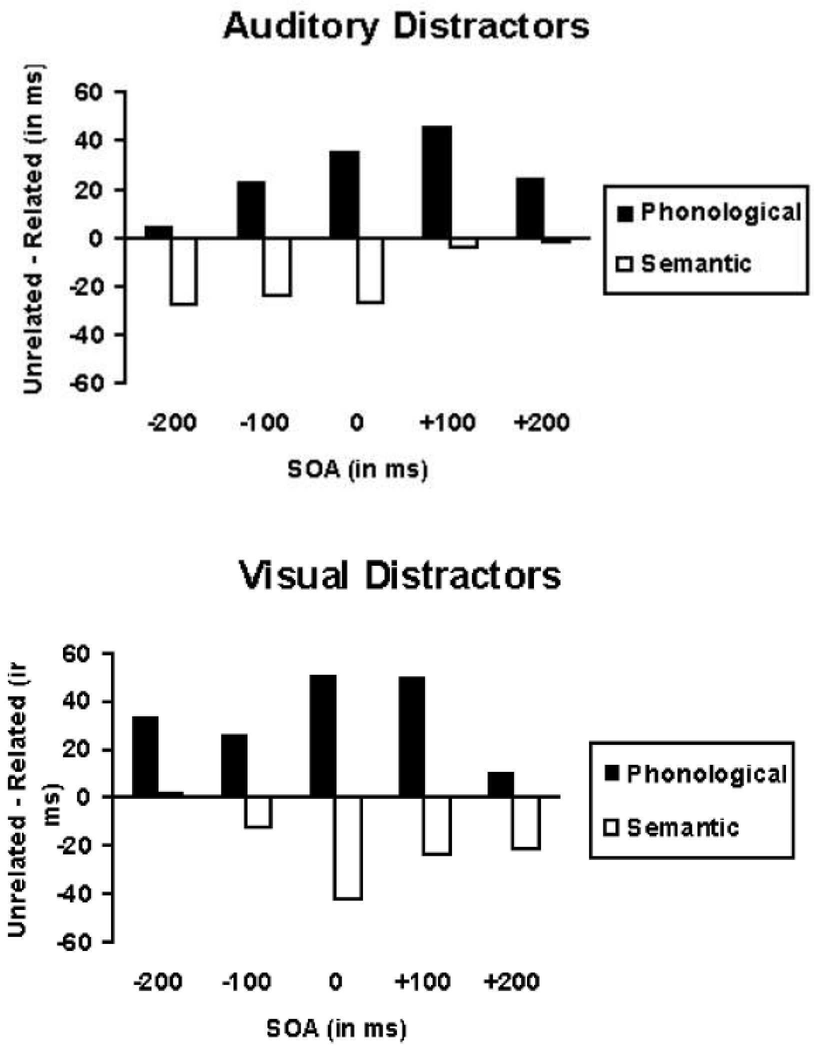

Figure 16. "Experiment 1: Effects of semantically and phonologically related distractors (unrelated minus related condition) varied by distractor modality (visual vs. auditory) and stimulus-onset asynchrony" (from Damian & Martin, 1999, p.5).

Results from Damian and Martin (1999) showed a significant semantic interference for semantic distractors and attenuated semantic interference effects in semantically and phonologically related (i.e., mixed) distractors at the earliest SOA (-150ms). However, facilitation effect size for phonological distractors was small. At the 0ms SOA, the facilitation effect size was increased although the semantic interference for semantic distractors was still greater than that of phonological facilitation. The semantic interference effect at the -150ms SOA for semantically and phonologically related distractors was entirely eliminated. At the longest SOA (+150ms), a greater phonological facilitation effect size was shown compared to the reduced semantic interference. Unlike the preceding two SOAs (i.e., -150ms and 0ms), semantic and phonological effects were added to the semantically and phonologically related distractor words (see Figure 17). Damian and Martin interpreted this additive relationship as a complete independence of lemma and lexeme stages, supporting the discrete two-step model. They also identified an interactive mechanism between semantic and phonological relatedness from the findings of the non-additive relationship which was shown at SOA = -150ms and 0ms. These results showed that there was interactivity between the two steps. Thus, Damian & Martin's (1999) finding supports the idea that presenting mixed distractors at a -150ms or 0ms SOA for different frequency target items can be used to examine whether word frequency impacts interactivity using RTs.

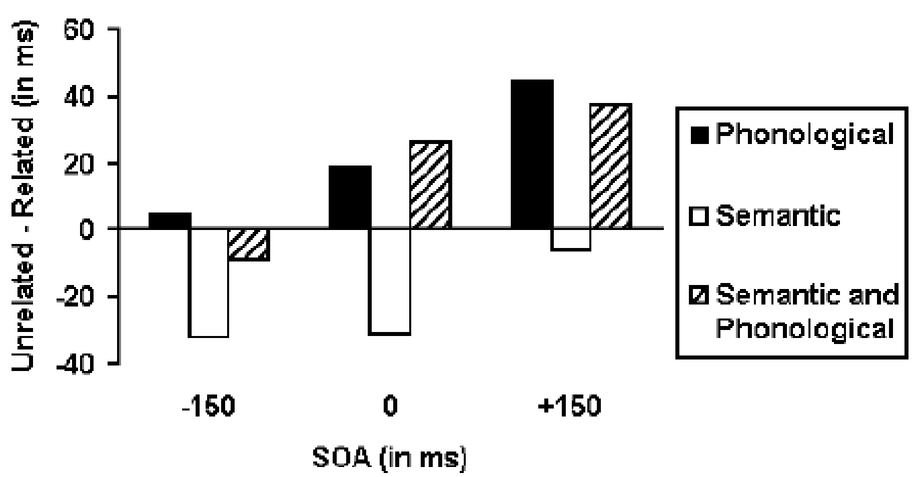

Figure 17. "Effects of semantically related, phonologically related, and semantically and phonologically related distractors varied by stimulus-onset asynchrony" (from Damian & Martin, 1999, p.11).

2.1.5.2 Distractor frequency effect in PWI

Response latencies within the PWI task can vary as a function of the relationship between target and distractor words (Finocchiaro & Navarrete, 2013). As previously reviewed, HF distractors that interfere more with a target word's retrieval can be interpreted under the activation level hypothesis (Miozzo & Caramazza, 2003). A finding of no difference between HF and LF distractors can be accounted for by the selection threshold hypothesis. Two contradicting viewpoints (competitive lexical selection versus non-competitive lexical selection) relevant to distractor frequency effects have been proposed. The predominance of evidence suggests that LF distractors produce more interference than HF distractors (Dhooge & Hartsuiker, 2010, 2011; Miozzo & Caramazza, 2003). According to the competitive model, the speed of selecting target lexical nodes is dependent on the activation levels of non-target nodes because distractor words share a semantic relationship with the target words (e.g., dog-cat). As a result, the distractor words increase the activation levels of non-target lexical nodes during the selection process and therefore interfere with the selection of the target nodes, compared to the case where they do not share any relationship (e.g., dog-phone). Accordingly, this phenomenon can be understood under the activation level hypothesis (Miozzo & Caramazza, 2003). HF distractor words have stronger interference effects on the retrieval of target words than LF distractor words. Several studies support this view under PWI by measuring picture-naming latencies to provide evidence for semantic interference or the semantic inhibition effect in order to explain competitive selection processes operating at the lexical level (Damian & Martin, 1999; Hutson, Damian, & Spalek, 2013; La Heij, 1988; Levelt et al., 1999; Roelofs, 2003).

Conversely, the non-competitive model holds that a lexical node is selected without influencing the activation level of non-target nodes (e.g., Caramazza, 1997; Dell, 1986; Dhooge

& Hartsuiker, 2010, 2011; Miozzo & Caramazza, 2003; Rapp & Goldrick, 2000; Stemberger, 1985). Miozzo and Caramazza (2003) found that LF distractor words interfere more with retrieving target words in the PWI task than HF distractor words. The traditional competitive lexical selection model cannot explain this phenomenon, where HF should interfere more. To account for this finding, the non-competitive account or "response exclusion hypothesis" (REH) was suggested (Mahon, Caramazza, Peterson, & Vargas, 2007). This hypothesis assumes that the target response can be produced only if the single-channel output buffer, which is located at the post lexical level, is not occupied by non-target words. That is, non-target words must first be cleared in the output buffer before articulating the target picture name. With respect to the frequency of non-target words, the REH proposes that HF distractors are rejected earlier in the buffer, and therefore interfere with picture naming less than the LF distractors.

Given that distractor frequency has not been shown to have consistent effects on lexical selection, controlling the distractor frequencies to examine the variable of primary interest, examining the effects of target word frequency on lexical selection, may be risky. For example, if HF distractors are chosen as stimuli and there happens to be a change on RTs depending on the target items' frequency, one could claim that the RT change resulted from the influence of HF distractors, not from the influence of target frequencies. Following the competitive hypothesis, the HF distractors can be interpreted as the main cause for inhibiting the retrieval of target items, which delays the lexical retrieval time. Thus, it is challenging to interpret the presence of the word frequency effect. In order to avoid this, a wide range of distractor frequencies should be assessed rather than choosing either (or both) HF and LF distractors. These effects should be tested using a regression model, as mentioned above, where a continuous variable can be entered as a predictor to examine its effect along with other variables' effect on the dependent variable.

Without examining it directly, it is not clear whether there is an interaction between target word frequency and distractor frequency at different processing steps. So far, no studies have directly addressed this issue. Thus, analysis of interaction effect on RT under PWI task needs to be assessed using semantic, phonological and mixed distractor conditions.

2.2 SUMMARY AND STATEMENT OF PURPOSE

In reviewing AAC technologies currently available for PWA, limitations were noticed; primarily a lack of consideration of PWA's superior HF word retrieval being brought into the process of developing AAC systems. Given that a graphic symbol-based AAC interface provides semantic representation, a question was raised whether this HF word advantage can be expected in PWA's AAC performance. Due to the lack of consensus on the effect of frequency on lexical selection, areas relevant to this debate were discussed. These areas include: characteristics of word retrieval deficits in PWA, two steps of lexical access, word retrieval models, evidence for a robust frequency effect, the mechanism of the word frequency effect, and the PWI paradigm as an alternative way to investigate the locus of the word frequency effect.

The evidence of error types leads the proposal of a mechanism involved in the retrieval process that is closely related to the two-step assumption: lexical selection and phonological encoding. Despite the robust effect of word frequency, the locus of frequency has remained subject to controversy. Jescheniak & Levelt (1994) argued that the locus of the frequency effect is at the phonological encoding step using evidence from the gender-decision task and the homophone-naming task. However, contradictory results have also emerged (e.g., Caramazza et al., 2001; Cuetos et al., 2010; Kittredge et al., 2008).

The controversy about the locus of the frequency effect was discussed under the two currently influential models—the DTS and the IA models. Although both models are based on the two-step assumption, several different perspectives on the word retrieval mechanism were discussed. The primary discrepancy was the interactivity between the two steps. Drawing on the DTS model where interactivity is not accepted, Jescheniak & Levelt, (1994) claimed that there was no frequency effect on lexical selection. Conversely, using the IA model, Kittredge et al. (2008) argued that word frequency affects both lexical selection and phonological encoding because strongly activated phonological-representation-nodes transmitted feedback to the prior step. This controversy is illustrated in the theoretical schematic of the current study in Figure 18.

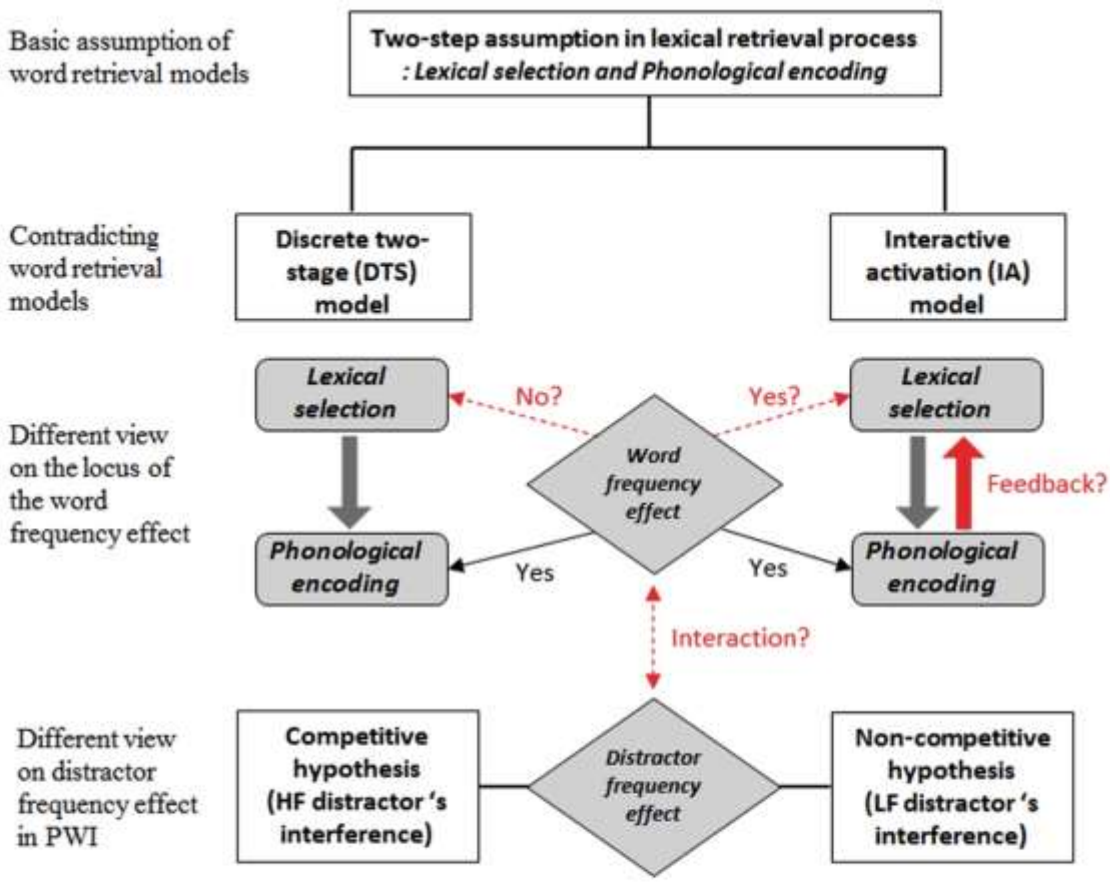

Figure 18. Theoretical Schematic of the current study

Limitations in previous studies were discussed with regard to word frequency effects on lexical selection. Although Kittredge et al. (2008) argued that lexical selection engaged interactivity in the frequency effect , mixed errors that have played a critical role to support the interactivity were missing in their data analysis. In addition, only errors were analyzed, which failed to identify the influence of word frequency on the retrieval *process*. The need to use RTs is highlighted by working on identifying the locus of the frequency effects in the PWI paradigm.

The current study sought to investigate the word frequency effect in lexical selection by measuring RTs using the PWI paradigm with healthy adult participants. The rationale for using the PWI paradigm in a normal healthy population was to gain a firmer understanding of the word frequency effects in healthy individuals that could then be extended to PWA. This extension may be justified under a unified model of lexical access that can account for both non-aphasic and aphasic speech errors (Dell et al., 1997; Rapp & Goldrick, 2000) under the assumption that speech and language performance exits on a continuum from normal to aphasic (McNeil, 1982) (also known as the continuity hypothesis; Dell et al., 1997; Freud, 1953; Schwartz et al., 2006). Specific aims and hypotheses are as follows.

<u>Specific Aim 1</u> is to investigate whether word frequency affects lexical selection during a PWI task. According to the DTS model, only phonological encoding is influenced by word frequency due to the absence of feedback activation at the semantic representation level (e.g., Jescheniak & Levelt, 1994). In contrast, in the IA model, word frequency is considered to affect both steps and interactivity is assumed to play a role in spreading the frequency effect to the higher step (e.g., Dell, 1990; Kittredge et al., 2008).

Hypothesis 1a proposes that for both the semantic and mixed distractor conditions, RTs will decrease as the target item’s frequency increases. Findings in previous studies that showed a word frequency effect on the lexical selection (Caramazza et al., 2001; Cuetos et al., 2010; Gahl, 2008; Kittredge et al., 2008; Navarrete et al., 2006) would be consistent with this pattern, and an interactive network model would be most consistent with these findings. Generalized linear mixed models (GLMMs) were used for analysis to accommodate the nesting of responses within the individual. Section 3.8 addresses a rationale and provides a detailed description of GLMMs.

An alternative hypothesis 1b proposes that there will be no significant change in RTs for both semantic and mixed distractor conditions as a function of the target's frequency. This alternative hypothesis is consistent with Jescheniak & Levelt (1994), consistent with the DTS model.

An additional alternative hypothesis 1c proposes that for the semantic distractor condition, there will be a significant effect of word frequency on RT. However, for the mixed distractor condition, no significant effect on the dependent variable is hypothesized. The rationale for this alternative hypothesis is that word frequency's effect on lexical selection will oppose the traditional view on the locus of frequency effect. The absence of a frequency effect on the mixed distractor condition will account for the non-involvement of indirect route for the influence of word frequency on the lexical selection via the interactive network (Dell & O'Seaghdha, 1991; Dell, 1986; Foygel & Dell, 2000). In this case, a direct route for the frequency effect on the lexical selection will be proposed.

Specific Aim 2 investigates whether the target item's frequency interacts with the distractors frequency during lexical selection using the PWI task. Given the influence of the distractors' frequency on the target item's retrieval in PWI, the second aim of this study sought to identify the existence of an interaction between the target items' frequency and the frequency of the distractors for the different distractor conditions. As in Specific Aim 1, GLMM was used to accommodate the nesting of responses within an individual.

As outlined above, either of two contradicting viewpoints make predictions regarding the distractor's frequency effect in the PWI task and the interaction between distractor frequency and target word frequency. One is that lexical selection is competitive, and thus the speed with which a target lexical node is selected depends on the activation levels of non-target (i.e., distractor) nodes (e.g., La Heij, 1988; Levelt et al., 1999; Roelofs, 2003). According to this argument, HF distractors should interfere more with target retrieval than LF distractors. The opposing view is that lexical selection is not competitive and the speed with which a target lexical node is selected is independent of the activation levels of non-target (i.e., distractor) nodes (e.g., Caramazza, 1997; Dell, 1986; Rapp & Goldrick, 2000; Stemberger, 1985). Recently, a series of studies have shown that LF distractors interfere more with the picture naming than HF distractors, which is at odds with the studies based on the competitive hypothesis (Dhooge & Hartsuiker, 2010, 2011; Hutson et al., 2013; Miozzo & Caramazza, 2003).

Hypothesis 2a proposes that there will be a significant RT interaction between the target items' frequency and the distractor frequency. The Competitive hypothesis will be supported if HF distractors interfere more with retrieval of target items than LF distractors (Damian & Martin, 1999; Hutson et al., 2013; La Heij, 1988; Levelt et al., 1999; Roelofs, 2003). If LF distractors interfere more with retrieval of target items than HF distractors, the non-competitive hypothesis will be supported (Caramazza, 1997; Dell, 1986; Dhooge & Hartsuiker, 2010, 2011; Miozzo & Caramazza, 2003; Rapp & Goldrick, 2000; Stemberger, 1985).

An alternative hypothesis 2b proposes that no significant interaction effect will be shown for RTs. This finding will be interpreted to support the position that the target word frequency independently modulates the lexical retrieval speed without being influenced by distractor frequency.

Specific Aim 3 proposed to investigate whether there is a difference between RT and response type when examining the frequency effect for healthy adults during the PWI task. Due to the issue of sensitivity of measurement, the frequency effect may be observed differently depending on the dependent variable. A discrete variable was applied into the GLMM to answer this question and the findings were compared with a continuous variable, RT.

Section 2.1.4.1 addressed the psycholinguistic mechanism of word frequency in lexical access (Miozzo & Caramazza, 2003). It was proposed that either resting activation level (Coltheart et al., 2001; Goodglass & Wingfield, 1997; McClelland & Rumelhart, 1981) or the selection threshold level (e.g., Jescheniak & Levelt, 1994; Morton, 1969) plays a major role in determining the retrieval speed depending on the word frequency. In brief, according to the activation level hypothesis, HF words are processed faster because they have a higher level of resting activation, and can reach a selection threshold faster than LF words (Coltheart et al., 2001; Goodglass & Wingfield, 1997; McClelland & Rumelhart, 1981). The selection threshold hypothesis proposes that HF and LF words start from the same resting activation level, but HF words are retrieved faster due to the lower selection threshold than LF words (e.g., Jescheniak & Levelt, 1994; Morton, 1969). Both hypotheses emphasized the temporal measures of target word selection. As HF words are selected faster than LF words through the time-dependent mechanism, RT observed in the PWI paradigm is expected to provide a more sensitive measurement than error analysis.

Hypothesis 3a proposes that RT will yield a higher sensitivity to word frequency effects than response type analysis. Alternative hypothesis 3b is that there is no difference in detecting the frequency effect between two dependent variables. Applying the psycholinguistic mechanism of word frequency to distractor frequency, hypothesis 3b will support the assumption that PWI task modulates not only the response time for each distractor condition but also the number of errors. This finding will extend the time-dominant feature of the PWI paradigm to the potential of creating the erroneous naming behavior of healthy adults.

2.3 SIGNIFICANCE

2.3.1 Theoretical Significances

The debate about the locus of the frequency effect in lexical processing is ongoing. The heart of this debate is whether word frequency influences lexical selection as well as phonological encoding. This study will provide evidence for one of the two or both positions by analyzing the difference in semantic and phonological distractors in the PWI paradigm.

In addition, it is uncertain whether the feedback activation from phonological encoding, which is strongly affected by the frequency effect, is transmitted to lexical selection, as Kittredge et al. (2008) argued. This study investigates the effects of word frequency on the interactivity between the two steps by analyzing the RTs and the number of mixed errors during a PWI task for mixed distractors. The findings about whether word frequency influences the interaction mechanism will help determine whether word frequency influences lexical selection through feedback from the second step, apart from the direct influence to lexical selection.

2.3.2 Clinical Significances

A consistent report on the advantage of HF words' influence on RT and accuracy can be used as a rationale for prioritizing HF words in an AAC system as well as in the selection of treatment stimuli and tasks. However, it is questionable whether the word frequency effect occurs in graphic symbol-based AAC systems. Graphic symbols used in AAC provide semantic prompts for each word (e.g., 'water') that contains its relevant semantic features (e.g., a symbol presenting a glass and liquid in it). PWA with impaired reading ability have to rely heavily on semantic cues in graphic symbols. This study will provide evidence of whether word frequency influences lexical selection and it will provide information relevant to determining whether word frequency effects can be expected in the graphic symbol-based AAC system for PWA in their daily communication.

Since this study presents distractors with a variety of frequencies, it will provide evidence as to whether an interaction between the target item's frequency and the distractor item's frequency is present. In the case of an interaction effect between the target and distraction items' frequency, the HF word advantage will be diminished depending on the neighboring symbols' frequency. Such a finding would indicate that careful vocabulary organization is required when creating an AAC application.

3.0 RESEARCH DESIGN AND METHODS

3.1 PARTICIPANTS

Fifty healthy adults participated in this experiment. A-priori estimates for an appropriate sample size was based on the statistical literature (Kreft & de Leeuw, 1998; Maas & Hox, 2005; Schwab, 2002), a power analysis (see 3.1.1 'sample size' for details), and several word frequency studies that ranged from sixteen to fifty participants (Cuetos et al., 2010; Jescheniak & Levelt, 1994; Kittredge et al., 2008; Navarrete et al., 2006).

All participants met the following inclusion criteria: (1) Monolingual native American English speakers, determined by asking potential participants whether they had used or understood any languages other than English when they were learning to speak as a child; (2) between 40 and 64 years old; (3) score above 13.86 on the listening Computerized Revised Token Test (CRTT-L) (McNeil, Pratt, Szuminsky, Sung, Fossett, Fassbinder, & Lim, 2015); (4) A score above 25 on the CFL form of the Word Fluency Measure (WFM; Borkowski, Benton, & Spreen, 1967) which is known to involve the processes of initiation, verbal fluency, memory and ability to organize thinking (Lezak, Howieson, & Loring, 2004); (5) a score above 18 on the Raven's Coloured Progressive Matrices (CPM A, B, AB forms; Raven, 1962) to test individual's nonverbal reasoning ability; (6) performance that yielded a ratio (the delayed recall/immediate recall×100) greater than .70 on the immediate and delayed story retell subtest from the Arizona

Battery for Communication Disorders of Dementia (ABCD) (Bayles & Tomoeda, 1993), a measure used to screen out individuals with early dementing disease or primary intermediate-term memory disorders. Additionally, considering the task requirements of experiment requiring participant's intact vision and hearing, all participants passed; (7) a vision screening involving the reduced Snellen chart at 20/40 or better with a viewing distance that was equal to that of the computer screen in the experiment; (8) a pure-tone air-conduction hearing screening via conventional headphones at 25 dB HL at 0.5, 1, 2, and 4 kHz in at least one ear and (9) a score of 22 or above on the Northwestern University Auditory Test No. 6 (NU-6) (Tillman & Carhart, 1966; Wilson, Zizz, Shanks, & Causey, 1990), a test of spoken word recognition. In addition to screening tests, an object naming task from the Porch Index of Communicative Ability (PICA) (Porch, 1981) was conducted to evaluate the individual's object naming behavior as a potential covariate.

Exclusion criteria included a self-reported history of neurological damage or language, speech, and hearing impairment. Participants using hearing aid were excluded.

Potential participants were recruited from the Clinical Translational Science Institute (CTSI) Research Participant Registry and from recruitment flyers that were posted on buildings of the University of Pittsburgh.

In total, nine individuals were excluded from participation due to the following reasons: five failed to meet one of the screening test criterion including pure tone air conduction and CRTT-L; three were non-native American English speakers; one fell asleep in the middle of the experimental task.

3.1.1 Sample Size

Under the statistical framework of the multi-level effect models, where level-1 (i.e., items) is nested within level-2 (i.e., participants), sample size was calculated considering level-2. The rationale for this is based on the results of a simulation study by Maas and Hox (2005), which showed that level-1 sample size does not significantly affect the estimates in multi-level effect model analysis. Findings revealed that level-2 sample size influenced the accuracy of the estimates (regression coefficients and variances) and their standard errors.

The nature of the data derived for the two dependent variables, response types (i.e., discrete) and RTs (i.e., continuous), was considered when determining level-2 sample size. Regarding response types, Schwab's (2002) guideline was followed to determine the sample size for multinomial logistic regression, which is an extension of binary logistic regression. This equation allows for more than two categories for the dependent variable (Starkweather & Moske, 2011). According to Schwab (2002), sample size guidelines for multinomial logistic regression require a minimum of 10 participants per independent variable. In the current study, there are two primary independent variables (i.e., target item's frequency and the interaction of target × distractor's frequency). Therefore, at least 20 participants are needed to obtain enough statistical power.

Optimal Design (OD) Software (Version 3.01) (Raudenbush, Spybrook, Congdon, Liu, & Martinez, 2011; available online at http://sitemaker.umich.edu/group-based/optimal_design_software), was used to determine the level of statistical power with regard to RTs, which is based on a ratio scale. The OD program allows for appropriate power analysis for multi-level and longitudinal research (Raudenbush et al., 2011; Spybrook, Raudenbush, Liu, & Congdon, 2011). Two approaches can be used in the OD program for conducting a power analysis: a "power determination approach," or an "effect size approach" (Spybrook et al., 2011, pp. 6-7). The former is used when the effect size is already determined and the calculation of sample size is the main interest of the researchers to achieve a specific power (e.g. 0.80). This option represents the power on the y-axis. On the other hand, the latter is used when the researchers seek to compute the minimum detectable effect sizes (MDES) (Bloom, 1995) that can be observed at a specified level of power and significance level for any given sample size. By using either approach, a power analysis can be conducted to reach the same conclusions.

In the current study, the effect size approach was adopted, which shows MDES on the y-axis. This was because in previous studies (e.g., Kittredge et al., 2008) based on multi-level effect models, authors did not provide effect sizes due to the technical limitation of computing R^2 from the models. It is regrettable that other word-frequency-effect-locus studies that used different statistical approaches did not also provide information about effect size as a reference for the current study. When the effect size is unknown, selecting an option with the MDES on the y-axis from the OD program is more logical (Spybrook et al., 2011). The MDES range considered in the current study was between 0.2 and 0.3, which represents a small to medium effect size according to Cohen (1988). Following Spybrook et al., (2011), an alpha level was set to 0.05, level-1 *n* to 43 (number of items per each distractor condition), and power to 0.80. From the MDES curve, the required sample size to achieve a small to medium effect size was observed when J = 28 for effect size = 0.3 and J = 57 for effect size = 0.2 (see Figure 19). Here J represents number of subjects. Within a range of two J measures, Kittredge et al.'s (2008) sample size (J = 50) was used as one of the empirical rationales. In their study, a statistical power for determining the word frequency effect on lexical selection was obtained with 50 participants.

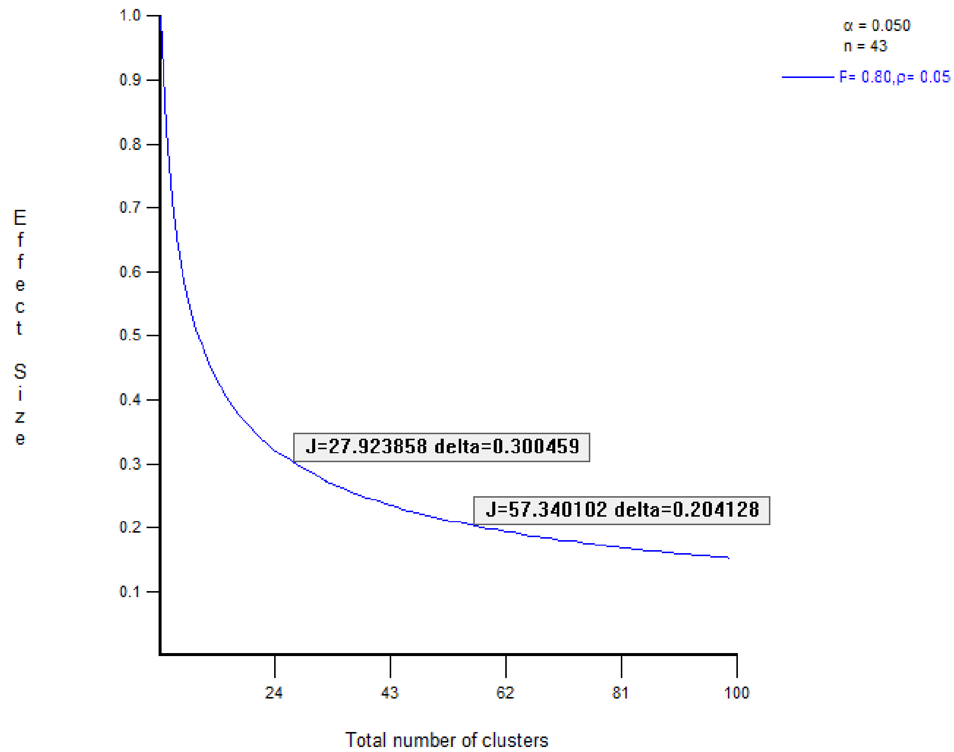

Figure 19. The power curve obtained by the Optimal Design (OD) program
Note. J indicates the total number of clusters used to achieve statistical power (F = 0.80). The alpha level and the power were set to 0.05 and 0.80 respectively. Two MDES points for 0.2 and 0.3 (effect size, delta; δ) are labeled with J (J = number of subjects).

This sample size of 50 is supported by literature. According to Kreft and de Leeuw (1998), level 2 sample size of 50 is known to be a frequently occurring number, and 30 is the smallest acceptable number in practice. From their perspective, the decision in the current study was made to take a more conservative approach and recruit additional participants than that required by the estimated power analysis for an effect size of 0.3 because power may be reduced in case of a lower delta level. Combining the two sample sizes obtained respectively for each of the dependent variables—20 for discrete variable, 50 for continuous variable, the larger minimum sample size (i.e., 50) was selected for this study.

3.1.2 Screening and Descriptive Measures

Demographic characteristics of the remaining participants who passed screening tests and completed the full experiment are described in Table 2. Fifty participants (37 females and 13 males) consisted of 33 white and 17 black individuals with ages ranging from 40 to 64 years old (mean = 56.36; SD = 6.17) and education years ranging from 10 to 25 years (mean = 16.37; SD = 2.95). The mean score of the PICA naming test was 14.90 (SD = 0.22; max = 15.0; min = 14.0). In screening tests, the mean score of the NU-6 was 24.96 (SD = 0.20; max = 25; min = 24); the mean score of the CRTT-L was 14.48 (SD = 0.30; max = 14.89; min = 13.86); the mean score of the ABCD was 0.96 (SD = 0.08; max = 1.14; min = 0.70); the mean score of the WFM was 46.04 (SD = 10.05; max = 64; min = 26); and the mean score of the RCPM was 31.88 (SD = 4.32; max = 36; min = 19).

Table 2. Participant biographical data and descriptive performance measures

Participant	Sex	Race	Age	Education level (years)	PICA naming	Vision by Reduced Snellen chart at 20/40 or better	Pure-tone air-conduction (at 25 dB HL at 0.5, 1, 2, and 4 kHz)	[a]NU-6 (cut off score = 22)	[b]CRTT-L (cut off score =13.86)	[c]ABCD ratio (cut off score = .70)	[d]WFM (cut off score = 25)	[e]RCPM (cut off score = 18)
1	Female	Black	40	12	15	Pass	Pass	25	14.32	0.93	57	35
2	Female	White	53	16	14.8	Pass	Pass	25	14.51	0.87	31	29
3	Female	White	59	16	15	Pass	Pass	25	14.77	1.00	41	34
4	Female	White	57	18	15	Pass	Pass	25	14.62	0.92	43	34
5	Male	White	49	16	15	Pass	Pass	25	14.44	1.00	51	33
6	Female	White	44	15	15	Pass	Pass	25	14.89	1.00	36	35
7	Male	White	64	17	15	Pass	Pass	25	14.10	1.00	32	36
8	Male	White	53	19	14.2	Pass	Pass	25	14.27	0.87	33	31
9	Male	White	64	16	15	Pass	Pass	25	13.97	0.80	26	35
10	Male	Black	47	12	14.3	Pass	Pass	25	13.86	0.87	36	29
11	Male	White	59	16	15	Pass	Pass	24	14.47	0.88	47	35
12	Female	Black	64	15	15	Pass	Pass	25	14.26	1.00	32	25
13	Female	White	54	18	15	Pass	Pass	25	14.60	1.00	35	35
14	Female	Black	44	20	14.5	Pass	Pass	25	14.45	1.08	62	31
15	Female	White	57	22.5	15	Pass	Pass	24	14.81	0.94	52	36
16	Female	Black	43	14	14	Pass	Pass	25	14.23	0.70	53	27
17	Female	White	44	14	15	Pass	Pass	25	14.83	1.00	46	34
18	Female	White	59	19	15	Pass	Pass	25	14.46	1.00	43	32
19	Female	Black	48	14	14.5	Pass	Pass	25	14.02	1.08	47	31
20	Female	Black	50	10	15	Pass	Pass	25	14.13	1.07	49	22
21	Female	White	50	25	15	Pass	Pass	25	14.75	1.00	46	35
22	Female	White	57	16	15	Pass	Pass	25	14.64	0.79	47	35
23	Male	Black	61	16	15	Pass	Pass	25	14.68	1.00	63	32
24	Male	White	59	22	14.7	Pass	Pass	25	14.70	0.88	54	35
25	Female	White	56	16	15	Pass	Pass	25	14.63	0.93	51	33
26	Female	White	59	14	15	Pass	Pass	25	14.64	1.14	52	30
27	Female	Black	57	14	14.8	Pass	Pass	25	14.38	1.00	58	22

Table 2 *(Continued)*

Participant	Sex	Race	Age	Education level (years)	PICA naming	Reduced Snellen	Pure-tone air-conduction	[a]NU-6	[b]CRTT-L	[c]ABCD ratio	[d]WFM	[e]RCPM
28	Female	White	56	17	15	Pass	Pass	25	14.62	0.94	36	33
29	Female	White	56	16	15	Pass	Pass	25	14.67	0.93	41	36
30	Male	White	64	18	15	Pass	Pass	25	14.64	1.00	32	34
31	Male	Black	56	16	14.8	Pass	Pass	25	14.43	1.09	41	31
32	Male	Black	55	14	14.8	Pass	Pass	25	14.15	1.00	36	27
33	Female	White	62	15	15	Pass	Pass	25	14.76	0.80	53	36
34	Female	Black	60	18	15	Pass	Pass	25	14.83	1.00	64	34
35	Female	White	63	16	15	Pass	Pass	25	14.79	1.00	62	36
36	Female	Black	57	13	15	Pass	Pass	25	14.78	1.08	49	31
37	Male	White	61	16	15	Pass	Pass	25	14.04	0.93	42	33
38	Male	White	60	23	15	Pass	Pass	25	14.71	1.00	52	35
39	Female	White	61	18	15	Pass	Pass	25	14.57	0.88	55	36
40	Female	Black	58	20	15	Pass	Pass	25	13.91	1.00	51	31
41	Female	White	62	12	15	Pass	Pass	25	14.41	1.00	49	34
42	Female	Black	57	14	15	Pass	Pass	25	13.94	0.93	40	19
43	Female	Black	61	12	15	Pass	Pass	25	13.96	0.93	27	20
44	Female	White	56	17	14.8	Pass	Pass	25	14.66	1.06	51	35
45	Female	White	61	16	15	Pass	Pass	25	14.60	0.94	36	29
46	Female	White	64	18	15	Pass	Pass	25	14.70	1.00	53	29
47	Female	Black	59	15	15	Pass	Pass	25	13.90	1.00	45	29
48	Female	White	62	18	15	Pass	Pass	25	14.78	1.00	59	36
49	Female	White	61	17	15	Pass	Pass	25	14.79	1.00	64	36
50	Female	White	55	17	15	Pass	Pass	25	14.68	0.94	41	33
Mean	(13M;37F)	(33W;17B)	56.36	16.37	14.90	(50 Pass)	(50 Pass)	24.96	14.48	0.96	46.04	31.88
SD			6.17	2.95	0.22			0.20	0.30	0.08	10.05	4.32

Note: [a]NU-6 = *Northwestern University Auditory Test No. 6* (Tillman & Carhart, 1966; Wilson et al., 1990); [b]CRTT-L = the listening *Computerized Revised Token Test* (McNeil et al., 2015); [c]ABCD ratio = *Arizona Battery for Communication Disorders of Dementia* (Bayles & Tomoeda, 1993), determined by number of delayed recall items/number of immediate recall items × 100; [d]WFM = CFL form of the *Word Fluency Measure* (Borkowski et al., 1967); [e]RCPM = *Raven's Coloured Progressive Matrices* (Raven, 1962).

3.2 STIMULI

3.2.1 Target Items

The Philadelphia Naming Test (PNT) (Roach, Schwartz, Martin, & Grewal, 1996), consisting of a set of 175 pictures, was used as the target stimuli. The PNT is a picture-naming test developed to collect a large corpus of naming responses from a standardized set of items. It has favorable psychometric properties and has been used in many studies investigating the underlying nature of aphasic naming deficits (e.g., Dell et al., 1997; Schwartz et al., 2006). The pictured items were selected on the basis of their high familiarity, name agreement, and good image quality (black-and-white line drawings of minimal complexity and confusability). Target names are all basic level concepts (i.e., not subordinate or superordinate; no targeting of famous faces or landmarks) and cover a relatively wide range of word lengths (1 to 4 syllables), and semantic categories (animals, body parts, clothing, food, furniture, tools, vehicles etc.). A high percentage of concept familiarity and name agreement are also reported (97% correct) from the naming performance of unimpaired controls (Dell et al., 1997; Roach et al., 1996).

In the current study, several lexical variables including target items' frequency, age of acquisition (AoA), neighborhood density, word length, and imageability were included in order to examine the effect of the primary predictors in the GLMMs while controlling for the confounding effects of the lexical variables. In addition, when allocating 175 PNT target items into four distractor conditions, these lexical variables were controlled across the conditions. The description of the databases used to match lexical variables follows.

Frequency was matched though the Celex corpus (Baayen, Piepenbrock, & Van Rijn, 1993; available at http://Celex.mpi.nl). This corpus is based on 16.6 million written and 1.3

million spoken words and is commonly used for word frequency studies in psycholinguistic research.

Unlike Kittredge et al. (2008) who extracted AoA information from the published norms for the American version of the MacArthur Communicative Development Inventories (MCDI), or CDI (Fenson, Tomasello, Mervis, & Stiles, 1994), the current study used a different AoA database. Only three categorical scales are used in the MCDI database (i.e., infants from 8 to 16 months; toddlers from 16 to 30 months; ages from 30 months). Thus, there is a chance that statistical power may be low, resulting in a failure to identify the co-varying effect of AoA on items. Also, the full corpus consists of 795 items (520 objects and 275 actions), which limits the full coverage of the PNT object items as well as their distractors. Thus, the current study utilized Kuperman, Stadthagen-Gonzalez, & Brysbaert's (2012) corpus, which presents AoA ratings for 30,121 English content words (nouns, verbs, and adjectives), using the web-based crowdsourcing technology offered by the Amazon Mechanical Turk. It reports as high validity and reliability as those collected in laboratory conditions, and reaches 93% for a subsample of 2,500 monosyllabic words. The database is available at http://crr.ugent.be/archives/806/.

Next, phonological neighborhood density, which refers to the number of each word’s phonological neighbors, was obtained by deleting one phoneme from the given word. The current study used the Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities online database (CLEARPOND; Marian, Bartolotti, Chabal, & Shook, 2012); available at http://clearpond.northwestern.edu/englishpond.html). Since auditory stimuli were used in the current study, phonological neighbor type was selected with the neighbor metric of ‘deletion.’

In addition, imageability was derived from the MRC Psycholinguistic Database (Coltheart, 1981), which contains 150,837 words and information on many lexical variables. A high rate indicates high imageability. The database is available at http://www.psych.rl.ac.uk. Finally, word length was determined as the number of phonemes in the word. This information was extracted from CLEARPOND (Marian et al., 2012).

3.2.2 Outliers and Distribution of Lexical Properties

After obtaining lexical properties for the stimuli, the descriptive statistics and a histogram of word frequency of target items, a variable of most interest in the current study, were obtained to determine outlier that may confound the research outcome. The mean of frequency for the 175 target items was 1194.40 (SD = 2815.90; min = 2, max = 29231). As shown in Figure 20, some extreme values were identified that may be considered outliers. In order to determine the outliers, frequency values were transformed to z-scores for all 175 items with cut-off score of +3.29 or -3.29 (equivalent to an alpha level of 0.001). Three items exceeded +3.29, as shown in Table 3. Since extreme values included in the analysis can result in a possible confounding effect on the results, those three target items (i.e., *hand, man, house*) were excluded from the analysis. Each outlier was removed from semantic, phonological, and mixed distractor conditions, thus all four conditions had an identical number of stimuli with an *n* of 43.

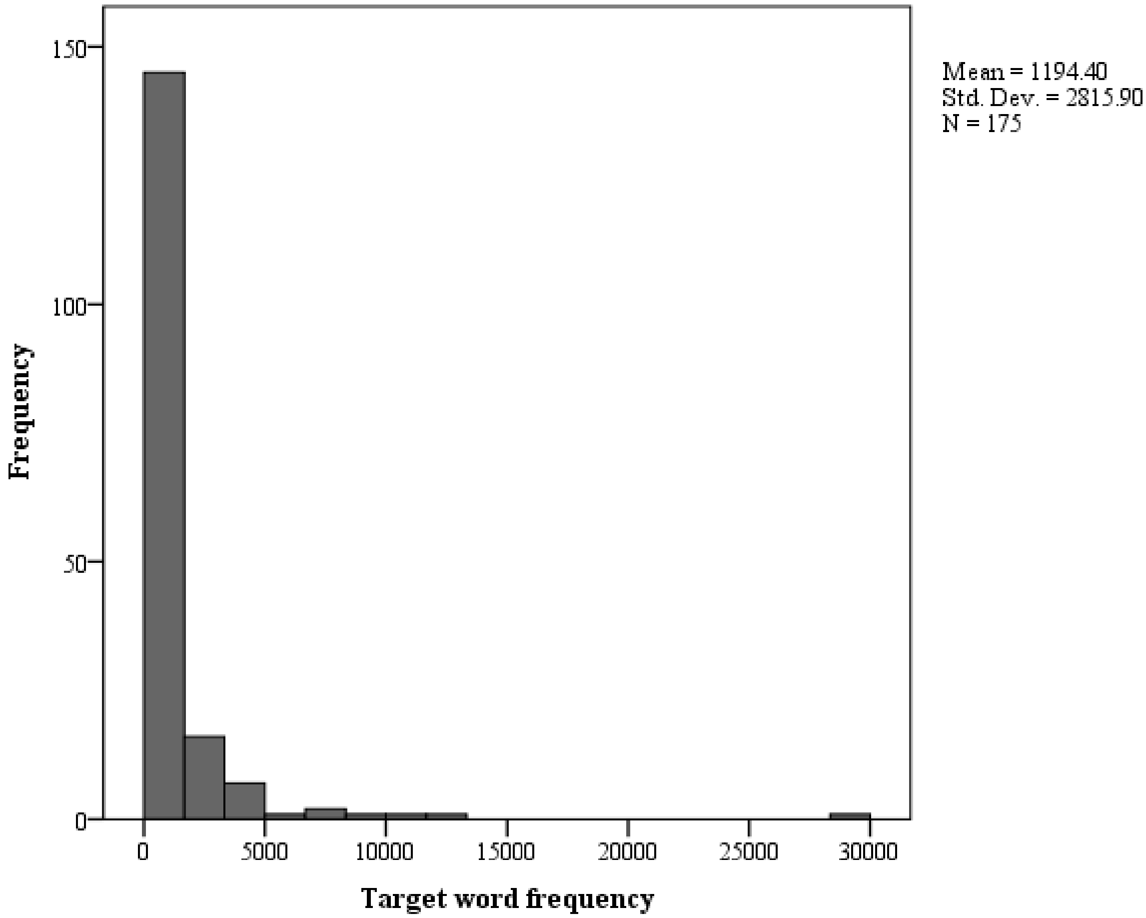

Figure 20. Histogram of 175 target word frequency

Table 3. Z-scores of three items that exceed z = 3.29

Distractor condition	Item	Frequency	Z score
Semantic	hand	12983	4.18644
Phonological	man	29231	9.95653
Mixed	house	10864	3.43393

For the remaining 172 items, the skewness of variables was examined along with means, SDs and ranges to determine appropriate statistical approaches (parametric vs. non-parametric) for further analysis. According (Bulmer, 1967), a distribution less than −1 or greater than +1 was interpreted as highly skewed, and a distribution between −1 and −½ or between +½ and +1 was interpreted as moderately skewed. As shown in Table 4, distributions of target words' frequency (3.01) and length (5.47), and distractor frequency (5.47) were highly and positively skewed. Target words' AoA (0.63), density (0.88) and distractor words' AoA (0.89), density (0.99) were moderately and positively skewed. A moderate-negative skewness was shown in target words' imageability (-0.79). Accordingly, non-parametric approaches were utilized in the current statistical analyses. Details are addressed in the section of 3.8. 'Statistical Analysis'.

Table 4. Descriptive statistics of 172 PNT target word and paired distractor word properties

	PNT target words				Distractors			
	Mean	*SD*	*Range*	*Skewness*	*Mean*	*SD*	*Range*	*Skewness*
Frequency	906.64	1424.37	2-9384	3.01	398.71	785.62	0-7528	5.47
AoA	4.97	1.32	2.50-9.16	0.63	6.22	1.93	2.78-14.88	0.89
Density	1.15	1.25	0-5	0.88	0.90	1.07	0-4	0.99
Length	4.28	1.69	1-10	1.05	4.79	1.49	2-9	0.49
Imageability	596.24	28.49	486-644	-0.79	585.87	31.07	507-652	-0.34

Note: PNT = Philadelphia Naming Test; AoA = age of acquisition; Density = Phonological neighborhood density; Length = number of phonemes.

3.2.3 Distractors

Among the error categories used in Schwartz et al. (2006) and Dell et al. (1997), semantic, phonological, mixed, and unrelated word distractors that indicate errors in the process of lexical access were presented aurally. Since PWA generated those error types due to damage in either one or both of the two lexical steps, the current study matched the distractor types with the four error types. Each distractor condition consisted of 43 items making the sum of items 172. Each distractor item was paired with the target item.

Each distractor was presented at a different SOA concurrently with the corresponding picture stimulus. The details on how distractors were developed for the experiment of the current study are illustrated as follows. The current study selected semantic and mixed distractors within the same category to increase the inhibition effect. This was based on the evidence that semantic distractors can show an inhibition effect when the distractors are from the same semantic category (e.g., CAR to the target item TRUCK), or a facilitation effect when the distractors are semantically associated (e.g., BUMPER to the target item TRUCK) (Caramazza, Alario, & Costa, 2005). Semantic categories of the distractors included body, animals, furniture, vegetables, fruits, cooking tools, containers, accessories, buildings, rooms, tools, household devices, persons, occupations, sports, plants, geography, rooms, clothes, shoes, weapons, musical instrument, and vehicles, which are identical to those of the target items. The unrelated words belonged to one of the same semantic categories of target items but did not share the same category of the matched target items. None of the distractors overlapped with the PNT word list.

The same database sources for the target's lexical variables (i.e., the frequency, AoA, neighborhood density, Imageability, and word length) were used when developing distractor items for the experiment.

3.2.4 Lexical Properties of Target Words and Distractors

Lexical properties of the 172 PNT target words and paired distractors for the four distractor conditions where semantic, phonological, mixed, and unrelated distractors, were presented in the PWI task are summarized in Table 5 and Table 6.

Due to the non-normal distribution of variables, the Kruskal-Wallis test was performed for target words and distractors separately as a function of distractor conditions. This was done in order to determine whether lexical properties of target words and distractors would be controlled across all four distractor conditions. None of lexical variables were significantly ($p < .05$)

different between distractor conditions. There was no significant difference among distractor conditions in the frequency, χ^2 = 1.421, p = .701; AOA, χ^2 = 3.274, p = .351, neighborhood density, χ^2 = 1.569, p = .666, word length, χ^2 = 3.659, p = .301, or imageability, χ^2 = 068, p = .995.

Table 5. 172 Target items' mean and standard deviation for five lexical variables

Lexical variable	Distractor condition	Actual *n* for SPSS	Mean	SD	Range
Word frequency	Semantic	43	966.88	1684.57	27-9384
	Phonological	43	885.42	1210.34	2-4620
	Mixed	43	548.49	662.65	33-2937
	Unrelated	43	1225.77	1810.72	16-7780
AoA	Semantic	43	5.10	1.33	2.89-8.17
	Phonological	42	5.12	1.10	2.95-8.50
	Mixed	43	4.70	1.14	2.50-7.55
	Unrelated	43	4.95	1.63	2.60-9.16
Density	Semantic	43	0.95	1.13	0-4
	Phonological	43	1.09	1.17	0-4
	Mixed	43	1.35	1.46	0-5
	Unrelated	43	1.19	1.22	0-4
Length	Semantic	43	4.49	1.82	1-9
	Phonological	43	4.44	1.68	2-10
	Mixed	43	4.23	1.48	2-7
	Unrelated	43	3.98	1.79	2-9
Imageability	Semantic	33	596.70	29.95	522-644
	Phonological	29	597.52	22.84	543-642
	Mixed	37	597.35	25.65	550-462
	Unrelated	36	593.64	34.39	486-369

Note: 'Actual n for SPSS' indicates the number of items which were derived from lexical database and used for statistical analysis.

Table 6 shows the lexical properties of the distractors. None of lexical variables showed a significant (p < .05) difference between distractor conditions and there was no significant (p < .05) difference in the frequency, χ^2 = 7.619, p = .055, AOA, χ^2 = 3.402, p = .334, neighborhood density, χ^2 = 2.723, p = .436, word length, χ^2 = 0.941, p = .816, and imageability, χ^2 = 4.855, p = .183.

Table 6. 172 Distractors' mean and standard deviation for five lexical variables

Lexical variable	Distractor condition	Actual *n* for SPSS	Mean	SD	Range
Word frequency	Semantic	43	338.51	497.69	6-2644
	Phonological	43	689.00	1346.99	4-7528
	Mixed	43	187.70	258.13	1-1441
	Unrelated	43	379.63	498.20	0-2517
AoA	Semantic	43	6.07	1.97	3-10.22
	Phonological	42	6.23	1.68	3.38-9.82
	Mixed	43	6.65	2.21	3.58-14.88
	Unrelated	43	5.90	1.80	2.78-11.22
Density	Semantic	42	1.02	0.98	0-4
	Phonological	42	1.00	1.19	0-4
	Mixed	40	0.85	1.14	0-4
	Unrelated	43	0.74	0.98	0-3
Length	Semantic	43	4.79	1.71	2-9
	Phonological	42	4.90	1.43	3-8
	Mixed	40	4.63	1.41	2-8
	Unrelated	43	4.84	1.41	2-8
Imageability	Semantic	25	586.20	29.67	508-645
	Phonological	29	578.86	33.83	507-647
	Mixed	27	586.19	33.60	515-652
	Unrelated	27	595.78	25.21	527-635

3.3 APPARATUS

This study collected all experimental data with an ASUS X550CA Notebook PC. E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002) was used to control the presentation of stimuli, timing operations, and data acquisition. Visual stimuli were the 172 PNT pictures. All pictures were extracted from the Moss Rehabilitation Research Institute (http://mrri.org), which are digitized at a size of approximately 18×20cm. Each picture was presented as a black line drawing on a white background. The auditory distractors were recorded by a female American English native speaker and presented to participants via headphones. In order to analyze response types and RTs, participants' verbal responses were all recorded using an Olympus digital voice recorder VN-702PC.

3.4 DESIGN

In order to establish a baseline, picture-naming without distractors was conducted before the experimental data collection. Pictures were presented in random order across participants. In the experimental condition, all participants named 172 PNT items while hearing auditory distractors consisting of 43 items for semantic, phonological, mixed distractor, and unrelated distractor conditions. Paired pictures and distractors were presented in random order across participants.

This study manipulated the SOA between picture and distractor referencing the procedures established at +150ms for semantic, mixed, and unrelated distractor conditions, and -150ms for the phonological distractor condition. These values are based on the literature demonstrating an inhibition effect of semantically related words on picture naming latencies, which appears at an early SOA (- 150ms) (distractor precedes picture) and the facilitation effect of phonologically related words that occurs at (0ms [distractor and picture presented concurrently], + 150ms [picture precedes distractor]) when distractors are presented auditorily (Damian & Martin, 1999; Sailor et al., 2009; Schriefers et al., 1990).

3.5 PROCEDURE

The overall procedure of the current study is depicted in Figure 21. All experimental procedures including informed consent was obtained by the author from each potential participant before he/she participated in the research study. The screening tests and the experimental tasks each took about 40 minutes. Two 10 minutes-breaks were given, one after the screening tests and the other one after the Baseline task. An approximate total duration was between 1.5 to 2 hours including the time for obtaining informed consent.

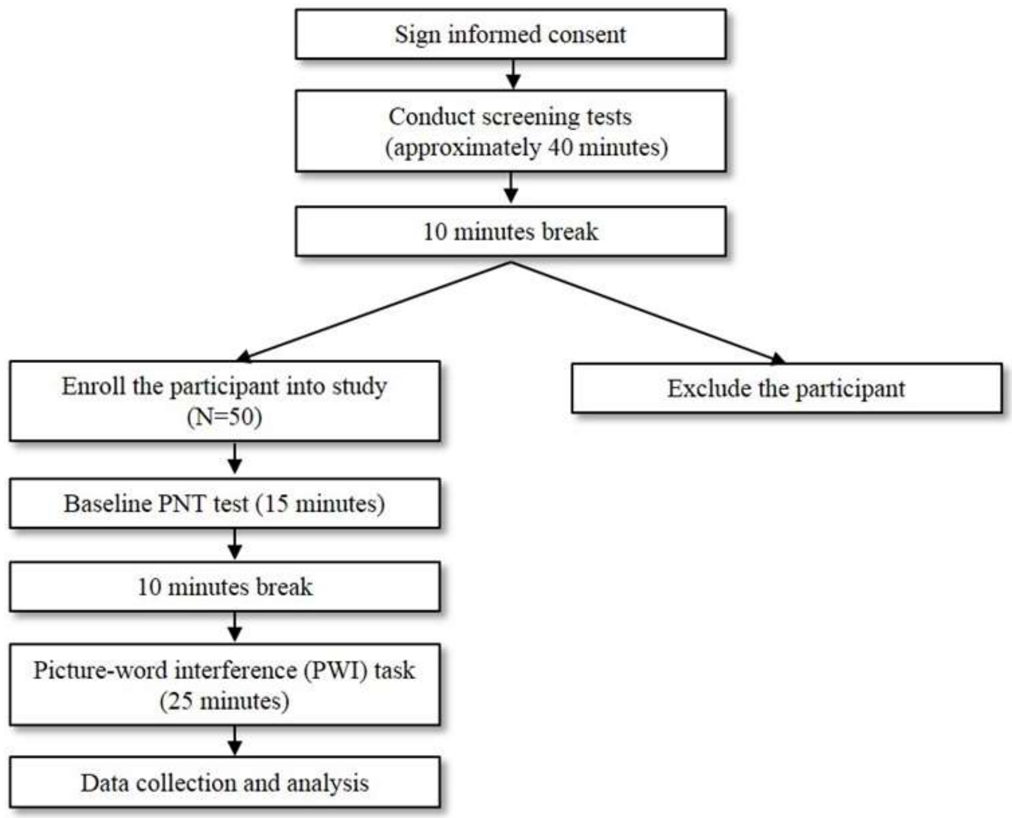

Figure 21. Study flowchart

All data were collected within a one-day visit. Hearing screening tests were administered in the Auditory Processing Laboratory of the Department of Communication Sciences and Disorders at the University of Pittsburgh, and remaining screening tests and experimental tasks were implemented in a lab located in the same facility on the day of the appointment. In order to avoid distracting the participant's attention during the experiment, the investigator sat to the side and slightly behind the participants' field of vision and did not interact with the participant until each task was completed.

Enrolled participants were tested individually in a quiet room and at a comfortable viewing-distance in front of the computer display. First, a baseline task instruction was provided on the screen as shown in Appendix B. Then, a plus sign (+) was presented for 500 ms to serve as a participant's visual fixation point on the screen. After a blank interval of 500ms, ten practice picture items were presented sequentially. Whenever the participant completed naming a single picture, the principal investigator clicked a mouse button to display the next picture. Between pictures, a 500ms blank interval, a 500ms fixation with a beep sound and another 500ms blank interval appeared (see Figure 22). After confirming participant's understanding of the picture naming task, individual PNT picture was presented on the monitor with the same procedures described above. To avoid order effects of items across participants and between the Baseline and PWI tasks, pictures were presented in a random order.

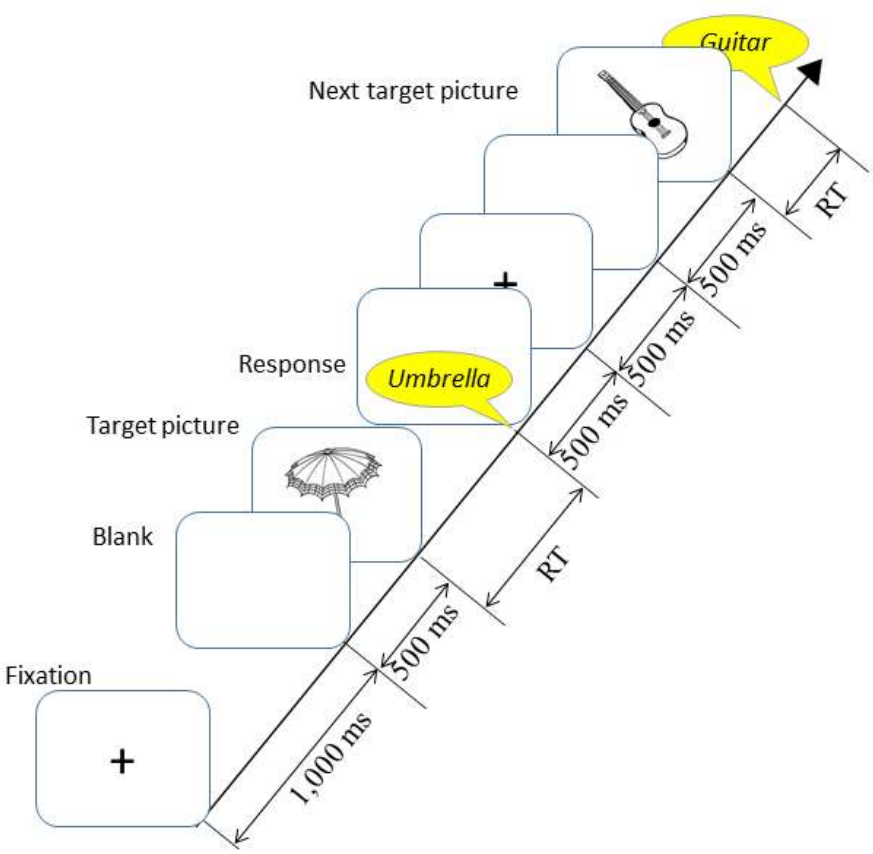

Figure 22. Baseline Task procedure

After finishing the Baseline task, participants had a 10 minute-break. An instruction for the PWI task was given to the participants (see Appendix B), which asked them to ignore the distractor words and name the pictures as quickly as possible. As shown in Figure 23, this was followed by a fixation cross at the center of the screen for 500ms with a beep sound in each individual trial. After a blank interval of 500ms, 10 practice items along with paired distractors were presented. Depending on the distractor type, either at the SOA= -150ms (for semantic, mixed, and unrelated distractors), or the SOA= +150 (for phonological distractors) was selected based on Damian and Martin (1999). Thus, distractors were presented 150 ms before or 150 ms after the picture onset. All distractors were presented auditorily via headphones. The experimenter clicked a mouse button as soon as a participant responded. This was followed by a

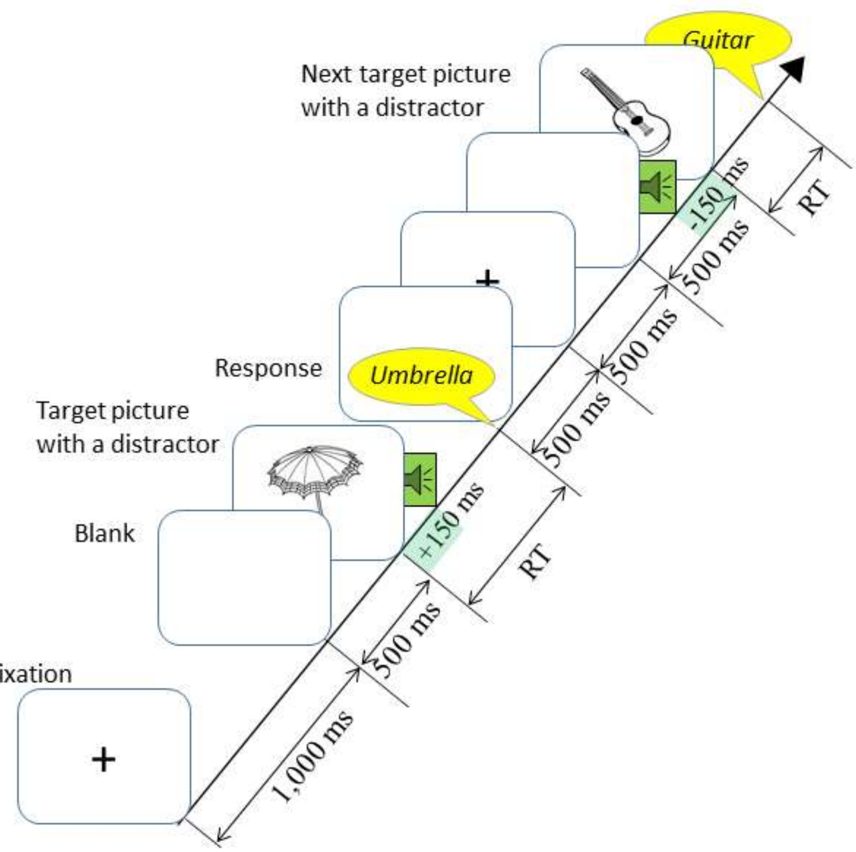

Figure 23. Diagram of the PWI Task Procedure

500ms blank interval and a 500ms fixation slide with a beep sound and then another 500ms blank interval, designed to appear before the onset of next target picture. After confirming that participants understood the instruction and named practical picture items as fast as they could, they were asked to name the randomly ordered 172 picture items while hearing paired distractors. The procedures with the experimental stimuli were identical to the practice procedures for the PWI task, as described above.

Each participant's verbal responses were recorded for analyses of RTs and response types. RTs were operationally defined as latencies from the onset of picture stimuli to the onset of the speech sound for the correct response. Due to cases of participants' false starts (e.g., clearing throat or use of interjection), poor voice key activation in the E-Prime was expected when it came to detecting the onset of the speech sound. Thus, manual examination of the speech wave forms using an audio file editing program, Audacity® (version 2.1.1; Retrieved from http://audacityteam.org/download on April 1, 2015) was used. In order to reliably determine the onset of speech sound, acoustic characteristics of each vowel and consonant for initial sounds of 172 target words were considered based on the literature (Kent & Read, 1992; *Speech Waveforms* by Mannell, R. Retrieved from http://clas.mq.edu.au/speech/acoustics/ waveforms/speech_waveforms.html on April 1, 2015). The key acoustic features of vowels and consonants that were taken into account in the current study are described in Appendix E. After obtaining a list of speech sounds onsets for Baseline and the PWI task separately, the time information extracted from the speech waveforms was synchronized with the one from the E-Prime report by matching the onset time of the acoustic signal (i.e. the beep sound).

For response type analysis, each individual's verbal response was dictated on the response sheet (see Appendix C) for 172 items of Baseline and PWI tasks. Those were evaluated based on the description of response types illustrated in Appendix D and then the initial letter of each type was inserted in the response sheet as follows: C = correct response, S = semantic error, P = phonological error, M = mixed error, and O = others. For statistical data analysis, these initial letters were then coded as 1 = C, 2 = S, 3 = P, 4 = M, and 5 = O. If the responses were real words, orthographical transcription was used (e.g., “candle”). Whereas. if the response was not a real word, broad phonetic transcription was used by two raters.

3.6 INTER-RATER RELIABILITY

Ten percent of the total responses from 50 participants (i.e., 5 participants’ samples) were randomly selected for the inter-rater reliability test between the author as a first rater and a research assistant as a second rater. The second rater was a senior undergraduate student in the Department of Communication Science and Disorders at University of Pittsburgh, who met eligibility criteria including: 1) completion of 'Speech Science' course, which was a prerequisite for this position due to the nature of task that dealt with a wide range of acoustic characteristics; 2) having a research experience in analyzing speech sound waves using Audacity software; 3) understanding of error patterns for people with neurogenic language disorders; and 4) completion of all required IRB training courses for accessing human subject data.

Before the actual analysis by the second rater, a one-day training was completed focusing on 1) reviewing the acoustic characteristics of vowels and consonants; 2) measuring and labeling the onset of each speech sound from speech waveforms in the audio file editing program; 3) analyzing response types of participants. First and second raters did the analyses independently.

The two raters' coded response types as well as RT data were computed for inter-rater reliability. The reliability of response type was calculated by Cohen's chance-corrected kappa statistic (Cohen, 1960) to account for chance agreement. The intra-class correlation coefficient (ICC) (Shrout & Fleiss, 1979) was used for the continuous RT measures.

The results of the response type analysis were Kappa = 0.819 with $p < .001$ for the 1,720 data (172 items × 10 samples), which is interpreted as an *almost perfect agreement*: As a rule of thumb, values of Kappa from 0.21 to 0.40 are considered *fair agreement*, from 0.41 to 0.60 *moderate agreement*, from 0.61 to 0.80 *substantial agreement*, and from 0.81 to 1.00 *almost perfect agreement* (Landis & Koch, 1977). Discrepancies between the two raters were resolved by consensus before the actual data analysis step for the full data set.

ICCs were computed for a one-way model for the RT dependent measure. In the one-way random-effects model, the n targets are rated by a different set of k raters randomly drawn from the population of potential raters. The ICC ranges between 0.00 and 1.00. Values that are closer to 1.00 represent stronger reliability. As a rule of thumb, values above 0.75 are interpreted as good reliability. However, in clinical studies, values above 0.90 tend to be required to ensure reasonable levels of association (Portney & Watkins, 2009). Results showed that the estimated correlation between individual ratings was 0.996 with 95% CI (0.9960, 0.9967), indicating extremely high correspondence between ratings within a target.

3.7 STATISTICAL ANALYSIS

3.7.1 Prescreening Procedures for Potential Covariates

Before preforming primary data analysis, potential covariates were prescreened to increase the statistical power and alleviate the adverse effects on estimated coefficients. Predictors that might show significant correlation with dependent variables but low correlations with the other predictors were determined as potential covariates that would be entered into the models for analysis. The details of the procedures are as follows:

Step 1. The Spearman’s rank correlation test was conducted between predictors and continuous dependent variable RTs with a cut off value of $p < .10$ using StataSE 14 (StataCorp) to identify the potential predictors of the dependent variables. Also, considering the nature of non-normality, rank biserial correlation coefficients (Cureton, 1968; also known as Somers' D) were obtained with the same cut off. Somers’ D is known as a nonparametric statistic for the point biserial correlation.

Step 2. For those predictors that were screened at step 1, the presence of the multicollinearity among predictors was evaluated by creating a correlation coefficient matrix (Mansfield & Helms, 1982; Neter, Kutner, Nachtsheim, & Wasserman, 1996). Using two highly correlated independent variables in the same model results in less predictive power and adversely affects reliable estimates of the individual coefficients by increasing the estimated standard deviations of the regression coefficients (Mansfield & Helms, 1982; Neter, Kutner, Nachtsheim, & Wasserman, 1996). Based on the rule of thumb, predictor correlation coefficients that were equal and above the cut off value .70, were interpreted as high correlation (Hinkle, Wiersma, & Jurs, 2003; Mukaka, 2012). One of the two variables that might show high correlation was removed from the model to reduce the collinearity. The decision of which of the two predictors to remove depended on the predictors' scientific or practical implication for current and future studies. Predictors with high implication remained in the models.

Step 3. The predictors that met the requirements of step 1 and 2 were chosen as potential covariates for primary data analysis. Only those covariates that contribute significantly to the dependent variables were retained in the final model.

3.7.2 Generalized Linear Mixed Models (GLMMs)

Multilevel-effect models take into account person-specific variability, to recognize the partial interdependence between responses and to account for the clustered nature of the data (Hofmann, 1997; Raudenbush & Bryk, 2002; Terhorst, 2008). Thus, variation in RTs for items within a participant (Baayen, Davidson, & Bates, 2008; Baayen & Milin, 2010), individual differences in the resting activation level and selection threshold per each item, along with the effects of covariance of other lexical variables (e.g., AoA, density) in the nested data structure (Kittredge et al., 2008; Raudenbush & Bryk, 2002) were expected to be accounted for by using this approach.

Non-normality of fixed variables was reported in section 3.2.2 'Outliers and distribution of lexical properties'. The dependent variable, RT also showed high positive skewness (19.99) (interpreted according to Bulmer [1967]). In this situation, GLMMs are known to be a flexible approach for analyzing dependent variable that are not normally distributed in multilevel effect models. In addition, GLMM allows for researchers to conduct analysis for continuous as well as discrete data (Bolker et al., 2009; Garson, 2013; McCulloch, Searle, & Neuhaus, 2008).

All analyses were computed within StataSE 14 (StataCorp). The package *gllamm* was used to run GLMM. Tests are two-tailed and the initial alpha level was set at 0.05. Multiple comparisons were corrected using Bonferroni correction, which is a commonly used method for several planned comparisons as a protection against Type I error (Portney & Watkins, 2009).

3.7.2.1 Fixed effect, random effect, covariates, and dependent variable

With the two-level structure of the data, level-1 examined the predictors and covariate effects in dependent variables in terms of the item, and level-2 examined the predictors and covariate effects in the same dependent variables in terms of participants.

To achieve specific aim 1, target word frequency was designated as a fixed predictor of level-1 (items) and RT was set as the dependent variable. For specific aim 2, distractor frequency was examined as a covariate in the models, and then the fixed effect was replaced with the interaction between the target word frequency and distractor frequency. For specific aim 3, either the fixed effect of target word frequency or the fixed effect of interaction between targets and distractors was examined as binary data for response type (0 for a non-occurrence of a particular response type, 1 for an occurrence of a particular response type).

For each model, the level-2 (individuals') intercept was considered a random effect, which showed the individual difference in the dependent variables. Potential covariates such as AoA, density, length, imageability at item-level, age, education years, and PICA naming scores at an individual-level were entered into the models in order to increase statistical power (Scherbaum & Ferreter, 2009). Covariates that were significant on the dependent variable remained in the final models.

Note that providing a well-established set of distractor conditions is important to evaluate the specific aims. In other words, unlike the Baseline, where distractors were not presented, participants should show different patterns of behavior depending on the distractor type provided during the PWI task. For example, increased RT with implementation of the semantic distractor condition compared to RT in unrelated distractor conditions indicates that the semantic distractors played a role in lexical selection at the semantic representation level (i.e., inhibition effect). Conversely, decreased RTs in the phonological distractor condition compared to RT in unrelated conditions reflects the phonological encoding stage (i.e., facilitation effect). Increased RT in the mixed distractor condition compared to RT in the unrelated condition, reflects the interaction between two representation levels (i.e., inhibition effect). Once this assumption is verified, the results of the main analysis can be interpreted under the two-step word retrieval model. Thus, additional GLMM analyses were added with different fixed predictor and dependent variables. To support specific aims 1 and 2, the RTs interaction of task type and distractor type were entered as fixed predictors. Meanwhile, to support specific aim 3, where response types were analyzed, the fixed predictor was replaced with the interaction of task type and response type, and the number of responses (i.e., data count for each individual's response types) was set as the dependent variable.

3.7.2.2 Estimation

Numerous methods of estimation are available and in general, researchers most frequently choose Maximum Likelihood (ML). ML estimation allows for choosing estimates of parameters for which the likelihood of observing the outcome Y is at a maximum (Snijders & Bosker, 1999). Other estimation method, such as restricted maximum likelihood (ReML), was not considered because ReML estimates are biased when using GLMM (Noh & Lee, 2007). Therefore, calculations cannot be used with GLMM (Grilli & Rampichini, 2006).

3.7.2.3 The models

In this section, three multilevel models were fitted into GLMM using *gllamm* syntax in StataSE 14 were explored depending on the characteristics of dependent variables: identity-gamma models, binary-logit models, and Poisson models. For specific aim 1 and 2 where continuous data, RTs, were used, the simple form of the two-level hierarchical linear models (Hofmann, 1997; Raudenbush & Bryk, 2002) was extended and fitted to the GLMM.

$$\text{Level-1: } Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij},$$

$$\text{Level-2: } \beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}, \beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j},$$

where there are $i = 1, \ldots, n_j$ level-1 units nested with $j = 1, \ldots, J$ level-2 units. Y_{ij} is the outcome measure for item i in participant j. X_{ij} is a level-1 predictor for item i in participant j (i.e., target word frequency or the interaction between targets' and distractors' frequencies were tested as predictors at level-1 in the current study). β_{0j} and β_{1j} are intercept and slope estimated separately for each participant and r_{ij} is the level-1 residuals. At level-2, γ_{00} and γ_{01} are level-2 coefficients. W_j is a level-2 predictor for participant j. u_{0j} and u_{1j} are level-2 residuals.

According to the description of Raudenbush & Bryk (2002) (pp. 293-294), the sampling model for a two-level model with continuous outcomes might be expressed as

$$Y_{ij} \mid u_{ij} \sim \text{NID}\ (u_{ij}, \sigma^2),$$

meaning that the level-1 outcome Y_{ij}, given the predicted value, u_{ij}, is normally and independently distributed with an expected value of u_{ij} and a constant variable, σ^2.

However, as addressed earlier, the data are positively skewed for RTs and some predictor variables. Considering the non-normally distributed u_{ij}, an adjustment of the distribution type was made. Instead of a Gaussian model, which has a normal and symmetric distribution, an identity-gamma model was selected in the current study for analyzing RTs. This model is frequently selected for positively skewed, untransformed, continuous data (Anderson, Verkuilen, & Johnson, 2010; Hardin & Hilbe, 2007).

Binary-logit models were fitted using *gllamm* for the hierarchical binary data. From the basic models above, a transformed predicted value η_{ij} is replaced with u_{ij}. The logit link for binary data is characterized as:

$$\eta_{ij} = \log\{\varphi_{ij} / (1 - \varphi_{ij})\},$$

where η_{ij} is the log of the odds of occurrence of a certain response type and φ_{ij} is the probability of that occurrence.

For count outcomes, a Poisson model (Poisson, 1837) was used with a log link function. Counts refer to a simple counting of events and count data may be the form of a rate of occurrence (Hardin & Hilbe, 2007). The log link is,

$$\eta_{ij} = \log(\lambda_{ij}),$$

where λ_{ij} is the even rate.

4.0 RESULTS

The summaries and results of the statistical analyses related to the research questions are provided in this section. The first subsection describes the response time (RT) results, with topics including: 1) rescreening test results potentially used as covariates for the continuous variable, RT; 2) results of interaction effects between task type and distractor type on RT used to determine whether the PWI task reflects well the lexical retrieval mechanism; 3) results of target word frequency effects on RT in three different distractor conditions (semantic, phonological and mixed conditions) in the PWI task used to address specific aim 1; and 4) interaction effects of target word frequency and distractor frequency on RT in the same distractor conditions used to address specific aim 2. A similar content structure was applied to the second subsection, but with response type as the dependent variable.

When deciding the optimal GLMM, several characteristics of the dependent variables were taken into account. That is, special cases of family of distributions and link functions were applied into *gllamm* syntax to estimate models. In addition, different statistical methods for examining the correlation coefficient were used analyze the prescreening tests for potential covariates. Details are illustrated at each subsection.

4.1 RESPONSE TIME (RT)

4.1.1 Analysis of Potential Covariates

Prior to GLMM analysis, associations between a dependent variable, RT and several variables for both level-1 (item) and level-2 (individual) were evaluated to identify the potential predictors of the dependent variables with the cut off value of $p < .10$. The nonparametric Spearman's rank order correlation coefficient (ρ) was used to estimate relationships among variables. Each correlation was based on all pairs with no missing values were. For those statistically significant ($p < .10$) predictors, correlations with other predictors were examined if the correlation coefficient was equal or above the cut off value ($rs = .70$) in order to reduce collinearity.

Table 7 summarizes the correlation coefficients for various variables that were considered in the baseline. Different variables at level-1 were included depending on the task type (i.e., variables about lexical properties of distractor were only considered in the PWI task), and separate matrixes of Spearman's rank correlation coefficients were computed for Baseline and the PWI task. There were significiant correlations between RT and target word proeprties including frequency ($rs = -0.13$, $p < .001$), AoA ($rs = 0.23$, $p < .001$), density ($rs = -0.12$, $p < .001$), and length ($rs = 0.16$, $p < .001$) at the alpha level =.10 (2-tailed). In addition, significiant correlations were found between RT and demographic properties including age ($rs = -0.02$, $p < .10$) and education years ($rs = -0.05$, $p < .001$), and a performance variable PICA ($rs = 0.02$, $p < .001$). As shown in the matrix, none of these variables were equal or greater than rs = .70 for detecting collinearity. Thus, all eight variables were entered to the models initially, then after an exploration for a parsimonious model, non-significant covariates were eliminated from the model.

Table 7. Spearman's rank correlation coefficients (ρ) matrix for Baseline

	RT(ms)	Item properties					Individual properties		
		Freq.	AoA	Density	Length	Image	Age	Edu.	PICA
RT(ms)	1								
Target words									
Freq.	**-0.13*****	1							
AoA	**0.23*****	-0.47***	1						
Density	**-0.12*****	0.18***	-0.20***	1					
Length	**0.16*****	-0.48***	0.37***	-0.39***	1				
Image	**-0.02***	0.16***	-0.09***	-0.24***	0.13***	1			
Demographics									
Age	**-0.06*****	0.00	0.01	0.00	0.00	0.00	1		
Edu.	**-0.05*****	0.00	0.00	0.00	0.00	0.00	0.18***	1	
PICA	**0.02****	0.00	0.01	0.00	0.00	0.00	0.46***	0.07***	1

$^{*}p < .10$, $^{**}p < .01$, $^{***}p < .001$.

Note: RT = response time (ms); PNT = Philadelphia Naming Test; AoA = age of acquisition; Density = phonological neighborhood density; Length = word length; Image = imageability; Edu = education years; PICA = Porch Index of Communicative Ability (Porch, 1981); Correlation coefficients in bold font meet or exceed the predetermined $p < .10$ for identifying covariates.

Table 8 shows the correlation coefficients derived from the PWI task. Significiant ($p < .10$; 2-tailed) but low (ranging from .10 to .14) correlations were found between RT and target word proeprties including frequency, AoA, density, and length. In addition, there were significiant but low (ranging from .02 to .08) correlations between RT and distracor word proeprties including word frequency, density, and imageability. Signficiant and low correlation coefficients between the demographic properties of age, education years, and PICA scores and the dependent variable were found. None of these variables met or exceeded the predetermined *rs* = .70 criteria for detecting collinearity. Thus, eleven variables were entered to the models initially, then non-significant covariates were eliminated for a parsimonious model.

Table 8. Spearman's rank correlation coefficients (ρ) matrix for PWI task

		Item properties										Individual properties		
		Target words					Distractor words							
	RT	Freq.	AoA	Density	Length	Image	Freq.	AoA	Density	Length	Image	Age	Edu.	PICA
RT(ms)	1													
Target words														
Freq.	**-0.12*****	1.00***												
AoA	**0.14*****	-0.47***	1											
Density	**-0.10*****	0.18***	-0.20***	1										
Length	**0.12*****	-0.48***	0.37***	-0.38***	1									
Image	**-0.02***	0.16***	-0.09***	-0.24***	0.12***	1								
Distractors														
Freq.	**-0.03*****	0.18***	0.02**	0.03**	-0.08***	-0.10***	1							
AoA	-0.01	-0.06***	0.06***	0	-0.07***	-0.14***	-0.38***	1						
Density	**0.02***	0.01	0.05***	0.02*	-0.03***	0.14***	0.19***	-0.11***	1					
Length	0.01	-0.14***	-0.01	-0.13***	0.21***	0.02*	-0.41***	0.26***	-0.49***	1				
Image	**0.03***	-0.19***	0.02	-0.27***	0.07***	0.09***	0.19***	-0.16***	-0.08***	0.12***	1			
Demographics														
Age	**-0.08*****	-0.01	0.01	0	0	0	0	0	0	0	0	1		
Edu.	**-0.05*****	0	0	0	0	0	0	0	0	0	0	0.18***	1	
PICA	**0.06*****	0	0	0	0	0	0	0	0	0	0	0.46***	0.06***	1

$^{*}p < .10$, $^{**}p < .01$. $^{***}p < .001$.

Note: RT = response time (ms); PNT = Philadelphia Naming Test; AoA = age of acquisition; Density = phonological neighborhood density; Length = word length; Image = imageability; Edu = education years; PICA = Porch Index of Communicative Ability (Porch, 1981); Correlation coefficients in bold font meet or exceed the predetermined $p < .10$ for identifying covariates.

In addition to the information about the potential covariates, other information can also be obtained from Tables 7 and 8. The coefficient of determination is calculated as the proportion of variablity in one variable that can be determined from the relationship with the other variable. The coefficient of determination for target word frequency on RT was 1.69% (r^2 = 0.0169) in Baseline and 1.44% (r^2= 0.0144) in the PWI task. Relative to Cohen's (Cohen, 1988) recommended interpretaton of effect size, these data revealed a small effect size (where r^2 = 0.01= small effect; r^2= 0.09 = medium effect; r^2= 0.25 = large effect). However, it may be unwise to guage the magnitude of the relationship between RT and word frequency solely by this effect size calculation because it does not consider data dependency for each participant. This limitation points to the importance of examining the relationships between variables under multi-level effect models.

4.1.2 Interaction Effect Between Task Type and Distractor Type

In this section, the interaction effect between task type and distractor type was reported. Task type consists of Baseline and the PWI task. Distractor type includes semantic, phonological, mixed, and unrelated. The primary purpose of this analysis is to determine if the semantic, phonological, and mixed distractor conditions represented well the semantic representation level, phonological representation level, and their intermediate stage of interaction by showing inhibition, facilitation, and combined effects respectively. Interpretations for specific aims 1 and 2 can be derived from the two-step theory if the interaction is found to be significant.

Mean RTs are illustrated in Table 9 for Baseline and the PWI tasks respectively for all four distractor types. A decreased RT in the PWI task was observed compared to Baseline. This decrease seems to be due to a learning effect from the repeated use of the 172 items, which was a predicted phenomenon from the literature. Identifying the task effect is not the research question of interest; an interaction effect of task type and distractor type was the focus of the current study. When an interaction effect was found, a significant difference in RT between distractor types was tested with reference to the unrelated distractor condition, which was used in the PWI task.

Table 9. Mean RT (ms) for Baseline and PWI Task

Distractor Type	Baseline			PWI task		
	Mean	*SD*	*n*	*Mean*	*SD*	*n*
Semantic	833.72	406.11	2091	767.54	376.9	2070
Phonological	853.39	462.84	2053	595.6	386.93	2079
Mixed	790.71	410.87	2091	719.27	301.17	2090
Unrelated	829.32	760.42	2068	696.15	322.51	2047
Total (N=172)	826.66	530.52	8303	694.59	354.3	8286

The GLMM was estimated in order to examine the interaction of task type and distractor type on RT. Considering hierarchical continuous data with a positively skewed distribution, identity-gamma models were employed by using *gllamm* syntax. Task type, distractor type, and their interactions were entered as a fixed effect along with confounding variables identified in the previous subsection. Confounding variables that showed no significant effect on RT were removed to obtain a parsimonious model. These results are reported in Table 10.

With a reference combination of Baseline × unrelated, a significant interaction effect was found for all three distractor types of interest: PWI × semantic (Coef. = 58.58, $p < .001$), PWI × phonological (Coef. = -155.51, $p < .001$), and PWI × mixed (Coef. = 48.51, $p < .001$). These interaction effects are depicted in Figure 24, Figure 25, and Figure 26 for semantic, phonological, and mixed distractors respectively.

Table 10. Effects of task type and distractor type and interaction effect (task × distractor)

	Coef.	SE	*P*	CI
Task type				
Baseline (ref.)	-	-	-	-
PWI	-104.25^{***}	7.89	0.000	[-119.71, -88.79]
Distractor type				
Semantic	-7.27	8.7	0. 403	[-24.31, 9.77]
Phonologic	0.76	9.21	0. 934	[-17.29, 18.81]
Mixed	-31.16^{***}	8.38	0.000	[-47.59, -14.72]
Unrelated (ref.)	-	-	-	-
Task × Distractor				
Baseline × Unrelated (ref.)	-	-	-	-
PWI × Semantic	58.58^{***}	11.68	0.000	[35.69, 81.46]
PWI × Phonologic	-155.51^{***}	11.46	0.000	[-177.97, -133.05]
PWI × Mixed	48.51^{***}	11.02	0.000	[26.92, 70.11]
Item properties				
Target word frequency	-0.01^{***}	0.00	0.000	[-0.01, -0.00]
″ AoA	31.34^{***}	1.81	0.000	[27.80, 34.89]
″ length	11.19^{***}	1.63	0.000	[8.00, 14.39]
″ image	-0.16^{*}	0.07	0. 014	[-0.29, -0.03]
Individual properties				
Age	2.68^{***}	0.35	0.000	[2.00, 3.36]
Education years	-4.56^{***}	0.72	0.000	[-5.98, -3.15]
PICA naming	-202.30^{***}	5.64	0.000	[-213.35, -191.26]
Random part	Variance	SE		
Level-2 variance	6183.32	254.1		

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.
Note: *n* of level 1 units = 12,907, *n* of level 2 units = 50.
Covariates that appeared for both Baseline and the PWI task were entered in the model.
ref. = reference, Coef. = coefficient, SE = standard error, CI = confidence interval, PWI = Picture word interference, AoA = age of acquisition.

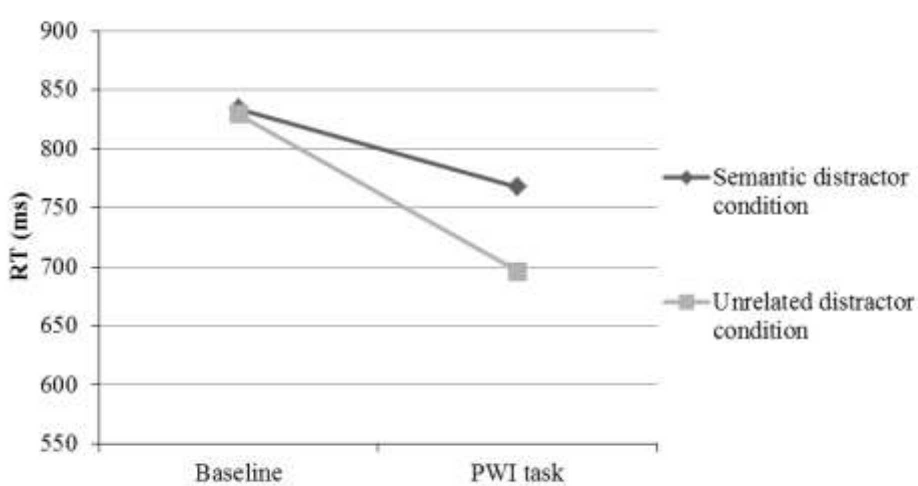

Figure 24. Mean RT of the semantic distractor condition and the unrelated distractor condition during Baseline and PWI task.

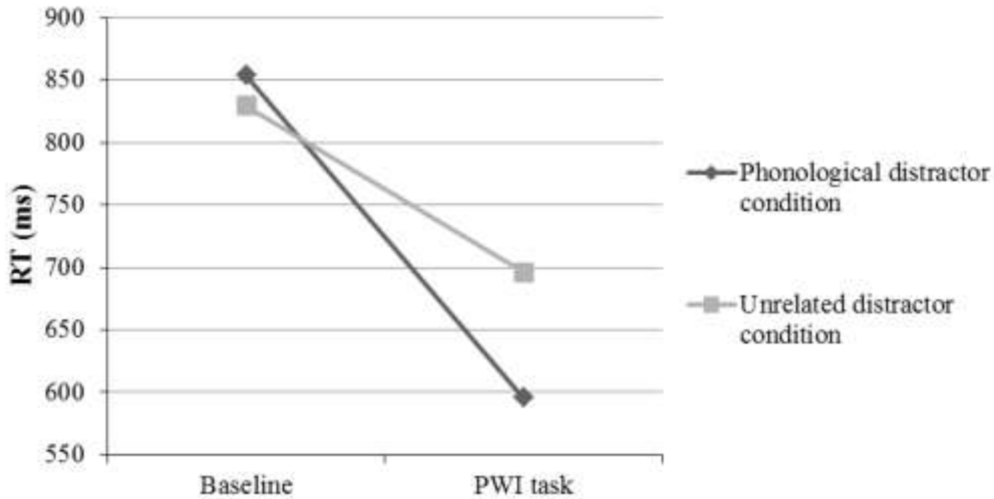

Figure 25. Mean RT of the phonological distractor condition and the unrelated distractor condition during Baseline and PWI task.

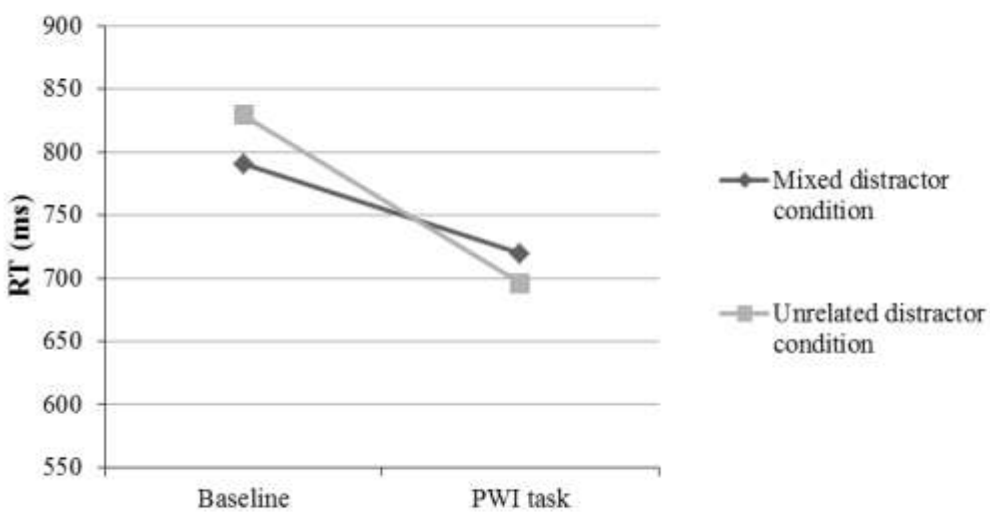

Figure 26. Mean RT of the mixed distractor condition and the unrelated distractor condition during Baseline and PWI task.

Due to the significant interactions of task type and distractor type, two additional identity-gamma models were analyzed for determining the simple main effect of distractor type for Baseline and PWI respectively, with reference to the unrelated distractor condition. An additional identity-gamma model was used for determining the simple main effect of distractor type for PWI with reference to mixed distractors. A Bonferroni adjusted alpha level was set at $p < .016$ (= 0.05/3 contrast).

A non-significant change in RT was found in the semantic distractor condition (Coef. = -3.54, p = .659) and the phonological distractor condition (Coef. = 8.43, p = .307) (see Table 11). Despite controlling the lexical variables that are frequently reported to affect RT, participants showed a significant decrease in RT in the mixed distractor condition compared to the unrelated distractor condition (Coef. = -25.93, p < .01).

Table 11. Effects of distractor type in Baseline

	Coef.	SE	*P*	CI
Distractor type				
Semantic	-3.54	8.02	0.659	[-19.27, 12.18]
Phonological	8.43	8.26	0.307	[-7.75, 24.61]
Mixed	-25.93^{**}	8.04	0.001	[-41.69, -10.17]
Unrelated (ref.)	-	-	-	-
Item properties				
Target word frequency	0.00^{*}	0.00	0.022	[-0.01, -0.00]
″ AoA	40.01^{**}	2.62	0.000	[34.87, 45.15]
″ length	16.88^{***}	2.05	0.000	[12.87, 20.89]
Individual properties				
Age	12.95^{***}	0.56	0.000	[11.85, 14.04]
Education years	-39.03^{***}	1.27	0.000	[-41.51, -36.55]
PICA naming	-155.13^{***}	4.63	0.000	[-164.20, -146.06]
Random part	Variance	SE		
Level-2 variance	22168.75	1053.27		

$^{*}p < .016$, $^{**}p < .01$, and $^{***}p < .001$.
Note: Distractor type, a predictor of interest was evaluated at $p < .016$ (=0.05/3 contrast) using Bonferroni correction.
n of level 1 units = 8,255, *n* of level 2 units = 50.
Coef = coefficient, SE = standard error, CI = confidence interval, ref. = reference, AoA = age of acquisition.

In the PWI task with the influence of a distractor (see Table 12), a significant increase in RT in the mixed distractor condition was shown compared to the unrelated distractor condition (Coef. = 18.17, $p < .025$). In addition, unlike Baseline, a significantly increased RT was shown with the semantic distractor where they were presented in the PWI task (Coef. = 52.55, $p < .001$). A significantly decreased RT was shown with the phonological distractor condition (Coef. = -138.03, $p < .001$) compared to unrelated distractor condition. All significant effects are depicted in Figure 27.

Table 12. Effects of Distractor Type in PWI task

	Coef.	SE	*P*	CI
Distractor type				
Semantic	52.55^{***}	7.39	0.000	[38.07, 67.02]
Phonological	-138.03^{***}	6.48	0.000	[-150.73, -125.33]
Mixed	18.17^{*}	7.28	0.013	[3.89, 32.44]
Unrelated (ref.)	-	-	-	
Item properties				
Target word frequency	-0.01^{***}	0.00	0.000	[-0.01, 0.00]
〃 AoA	25.36^{***}	2.3	0.000	[20.85, 29.86]
〃 length	12.02^{***}	1.71	0.000	[8.66, 15.37]
Individual properties				
Education years	-15.90^{***}	15.77	0.000	[-17.88, -13.92]
PICA naming	63.92^{***}	6.82	0.000	[45.55, 82.30]
Random part	Variance	SE		
Level-2 variance	6028.1	386.58		

$^{*}p < .016$, $^{**}p < .01$, and $^{***}p < .001$.
Note: Distractor type, a predictor of interest was evaluated at $p < .016$ (=0.05/3 contrast) using Bonferroni correction.
n of level 1 units = 8,237, *n* of level 2 units = 50.
ref. = reference, Coef = coefficient, SE = standard error, CI = confidence interval, AoA = age of acquisition.

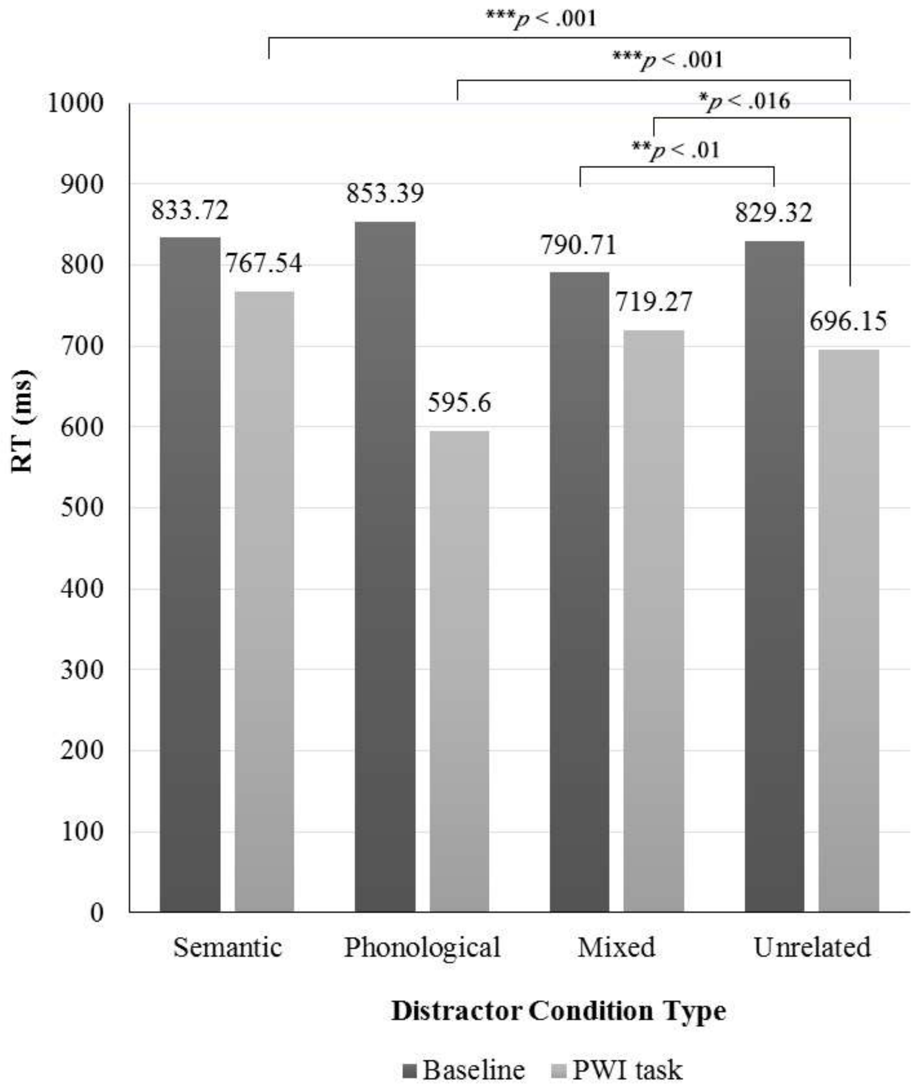

Figure 27. RT (ms) in four distractor types for Baseline and the PWI task.

Table 13 shows the results when the mixed distractor condition was set as a reference. Compared to the mixed condition, participants showed approximately 34ms delayed RT in the semantic condition (Coef. = 33.69, p < .001), but 153ms RT reduction in the phonological condition (Coef. = -156.46, p < .001), which means that the RT in the mixed condition fell between the RTs of the semantic and phonological conditions.

Table 13. Effects of distractor type in PWI task: When the mixed condition is set as a reference

Fixed part	Coef.	SE	*P*	CI
Distractor type				
Semantic	33.69***	7.62	0.000	[18.76, 48.63]
Phonological	-153.46***	6.68	0.000	[-166.55, -140.37]
Mixed (ref.)	-	-	-	
Unrelated	-17.95*	7.32	0.014	[-32.30, -3.60]
Item properties				
Target word frequency	-0.01***	0.00	0.000	[-0.01, -0.00]
″ AoA	25.45***	2.30	0.000	[20.93, 29.97]
″ length	11.80***	1.73	0.000	[8.41, 15.18]
Individual properties				
Age	5.51***	0.45	0.000	[4.64, 6.39]
Education years	-14.10***	0.87	0.000	[-15.80, -12.41]
PICA naming	-219.54***	7.66	0.000	[-234.55, -204.53]
Random part	**Variance**	**SE**		
Level-2 variance	10582.49	621.40		

*$p < .016$, **$p < .01$, and ***$p < .001$.
Note: A distractor type, a predictor of interest was evaluated at $p < .016$ (=0.05/3 contrast) using Bonferroni correction.
n of level 1 units = 8,237, n of level 2 units = 50.
ref. = reference, Coef = coefficient, SE = standard error, CI = confidence interval, AoA = age of acquisition.

4.1.3 Word Frequency Effect

Three sets of GLMM models addressed the first research question of this study; they examined the effects of word frequency on the RT in semantic, phonological, and mixed distractor conditions. For hierarchical continuous data, RTs showed a positively skewed distribution when identity-gamma models were estimated under the GLMM. Associations with target word frequency were examined along with other covariates for the semantic, phonological, mixed, and unrelated distractor conditions.

The analysis was undertaken in two steps. The first step (modeled without covariates) included one predictor of interest as a fixed effect aimed to examine whether target word frequency influenced RT. The second step (modeled with covariates) added all potential covariates to the previous model in order to investigate whether the coefficient of word frequency effect changed after considering the effects of all confounding variables at level-1 (i.e., item) and level-2 (i.e., individual) of the model. Consistent findings in terms of significance in both models were considered as a robust effect of word frequency on RT in a certain distractor condition. Separate analyses were conducted for four distractor conditions in the PWI task.

In Table 14, the first model (without covariates) summarizes the significant effect of word frequency on RT in the semantic distractor condition. As a unit of word frequency increased by100, the RT was reduced 2ms (Coef. = -0.02, $p < .001$). Even with the addition of nine covariates that had shown significant effects on RT, a significant word frequency effect was

Table 14. Word frequency effect on RT for the semantic distractor condition in PWI task: Coefficient estimates and standard errors

Fixed part	Model w/o covariates		Model w/ covariates	
	Coef.	SE	Coef.	SE
Level-1: Item				
Target word frequency	-0.02^{***}	0	-0.06^{***}	0.01
〃 AoA			-26.36^{***}	7.08
〃 density			-16.91^{*}	6.69
〃 length			-29.38^{***}	7.3
〃 image			-0.80^{*}	0.32
Distractor frequency			0.13^{***}	0.02
〃 density			23.54^{**}	7.96
〃 image			-1.08^{**}	0.31
Level-2: Individual				
Age			4.71^{**}	1.47
PICA naming			-164.32^{***}	40.35
Random part	Variance	SE	Variance	SE
Level-2 Variation	10612.53	-1360.53	19668.94	6185.97

$^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.
Note: For Model w/ covariates, *n* of level 1 units = 1,102, *n* of level 2 units = 50; Coef = coefficient, SE = standard error.

observed (Coef. = -0.06, $p < .001$). As the coefficient increased in the second model, RT was reduced by 6ms for each increase of one unit (100) of word frequency.

With regard to the covariate of distractor frequency, it was found that as the unit increases by 100, the RT increased by 13ms. Other significant covariates that occurred with changes in RT for the target words included AoA, density, length, imageability, and the distractor's density and imageability. Also, age and PICA naming score had a significant effect on RTs in the semantic distractor condition.

The data in Table 15 shows a significant word frequency effect on RT in the models without (Coef. = -0.02, $p < .001$) or with covariates (Coef. = -0.01, $p < .01$) for the phonological distractor condition. When covariates were added, the coefficient was reduced, resulting in a 1 ms decrease of RT for each increase of 100 units of word frequency.

Table 15. Word frequency effect on RT for the phonological distractor condition in PWI task: Coefficient estimates and standard errors

Fixed part	Model w/o covariates		Model w/ covariates	
	Coef.	SE	Coef.	SE
Level-1: Item				
Target word frequency	-0.02^{***}	0.00	-0.01^{**}	0.00
〃 length			11.86^{*}	5.03
〃 image			-0.63^{**}	0.22
Distractor density			9.00^{*}	4.5
Level-2: Individual				
Age			13.01^{***}	0.97
Education years			7.15^{*}	3.51
PICA naming			-356.16^{***}	34.96
Random part	Variance	SE	Variance	SE
Level-2 Variation	17036.07	1529.02	25549.39	2581.85

$^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.
Note: For Model w/ covariates, *n* of level 1 units = 1,404, *n* of level 2 units = 50; Coef = coefficient, SE = standard error.

In this phonological distractor condition, significant covariates include target word length (Coef. = 11.86, p < .05), imageability (Coef. = -0.63, p < .01), distractor's density (Coef. = 9.00, p < .05), age (Coef. = 13.01, p < .001), education years (Coef. = 7.15, p < .05), and PICA naming score (Coef. = -356.16, p < .001). Distractor frequency did not have a significant influence on RT.

Table 16 summarizes the significant effects of word frequency in the first (Coef. = -0.03, $p < .001$) and second models (Coef. = -0.05, $p < .01$) for the mixed distractor condition. When covariates were added, the absolute coefficient was increased with 100 units of word frequency yielding a 5 ms decrease in RT for the mixed distractor condition.

Table 16. Word frequency effect on RT for the mixed distractor condition in PWI task: Coefficient estimates and standard errors

Fixed part	Model w/o covariates		Model w/ covariates	
	Coef.	SE	Coef.	SE
Level-1: Item				
Target word frequency	-0.03***	-0.01	-0.05***	0.01
″ AoA			34.77***	4.27
″ length			-16.09***	3.65
″ image			1.20***	0.2
Distractor frequency			0.08***	0.02
Level-2: Individual				
Education years			-5.86***	1.83
PICA naming			-184.07*	36.66
Random part	Variance	SE	Variance	SE
Level-2 Variance	20402.36	-4516.79	12123.72	1507.77

* $p < .05$, ** $p < .01$, and *** $p < .001$.
Note: For Model w/ covariates, n of level 1 units = 1,803, n of level 2 units = 50; Coef = coefficient, SE = standard error.

With regard to the covariate of distractor frequency, as the unit of distractor frequency increases by 100, the RT increased by 8ms (Coef. = 0.08, p < .001). Other significant covariates include target word AoA (Coef. = 34.77, p < .001), length (Coef. = -16.09, p < .001), imageability (Coef. = 1.20, p < .001), education years (Coef. = -5.86, p < .001) and PICA naming score (Coef. = -184.07, p < .05).

4.1.4 Interaction Effect (Target Word Frequency × Distractor Frequency)

Interaction effects of target word frequency and distractor frequency on RT were examined focusing on semantic, phonological, and mixed distractor conditions in the PWI task. In identity-gamma models, potential covariates that showed significant effects on the dependent variable remained in the final model. The results are shown in Table 17.

A significant interaction effect was shown only with the mixed distractor condition (Coef. = 36.67, $p < .05$). As shown in Figure 28, RTs for high frequency distractors were larger than for low frequency distractors. However, RT there significantly more delayed for high frequency target words than low frequency target words. This trend was reversed in the condition of low frequency distractors.

Table 17. Effect of interaction between target word frequency and distractor frequency on RT for three different distractor conditions in the PWI task

Fixed part	Semantic distractor		Phonological distractor		Mixed distractor	
	Coef.	SE	Coef.	SE	Coef.	SE
Level-1: Item						
Target high freq.	-15.55	18.16	37.82	19.81	-15.45	12.04
Distr. high freq.	26.78	21	-52.79^{**}	16.32	6.9	14.4
Target high × Distr. high	34.45	25.62	2.94	22.49	39.55^{*}	19.35
Target word AoA			26.96^{**}	8.22	36.67^{***}	4.41
〃 density	-28.23^{***}	5.85				
〃 length			20.14^{***}	5.5	-6.98^{*}	3.45
〃 image	-1.15^{***}	0.2	-1.18^{***}	0.26	0.67^{***}	0.19
Level-2: Individual						
Age	5.19^{**}	1.69	11.58^{***}	1.18	2.668^{*}	1.13
Education years			-33.38^{***}	2.42		
PICA naming	-191.12^{***}	50.19	-252.91^{***}	29.49	-218.96^{***}	29.66
Random part	Variance	SE	Variance	SE	Variance	SE
Level-2 Variation	9967.44	1647.61	17984.64	2029.67	12164.49	2376.98

$^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.
Note: For the semantic distractor condition, *n* of level 1 units = 1,576, *n* of level 2 units = 50;
For the phonological distractor condition, *n* of level 1 units = 1,355, *n* of level 2 units = 50;
For the mixed distractor condition, *n* of level 1 units = 1,803, *n* of level 2 units = 50;
Coef = coefficient, SE = standard error.

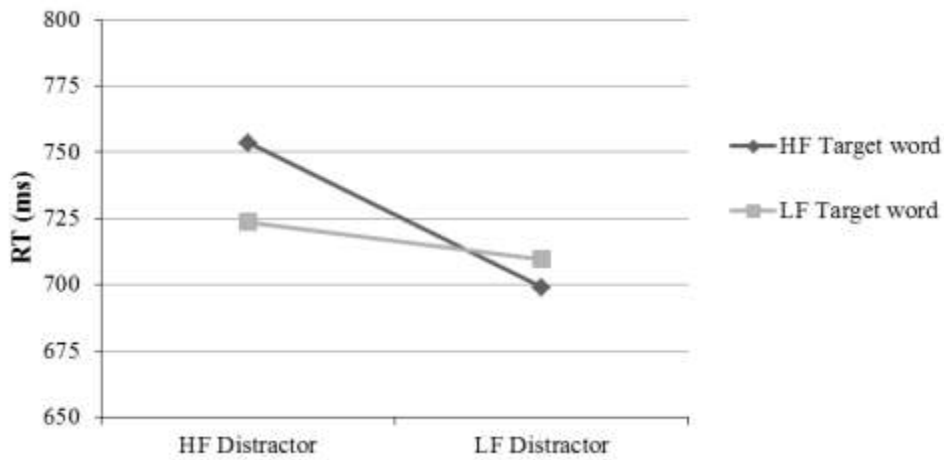

Figure 28. Mean RT of target word frequency and distractor frequency in the mixed condition in PWI task. Note: HF = high frequency, LF = low frequency.

Further analyses to identify simple effects for HF and LF distractors were implemented separately with adjusted alpha level at *p* = .016 (i.e., 0.05/3 contrasts) by Bonferroni correction. For HF distractors, target words with HF did not show significant RT change compared to the LF target words (Coef. = 15.51, *p* = .303, CI [-14.03, 45.05]). Non-significance was found for LF distractors between HF and LF target words as well (Coef. = -12.06, *p* = .325, CI [-36.06, 11.94]). For HF target words, there was no significant difference between HF and LF distractors (Coef. = 293.23, *p* = .892, CI [-3959.45, 4545.91]).

4.2 RESPONSE TYPE

4.2.1 Analysis of Potential Covariates

Prior to GLMM analysis, associations between a binary dependent variable and five different response types and several continuous variables for both level-1 (item) and level-2 (individual) were evaluated to identify the potential predictors. A cut off value of p < .10 was used. Somers' Ds were obtained based on all pairs of data values where missing values did not occur (i.e., pair-wise deletion). Likewise, collinearity of potential covariates (all are continuous data) was checked with cut off value *rs* = .70.

Table 18 shows the coefficients for the binary data for five different response types at Baseline. Correct responses were singificantly correlated with AoA (Coef. = -0.27, $p < .001$), length of target word (Coef. = -0.15, $p < .001$), age (Coef. = 0.13, $p < .01$), education years (Coef. = 0.22, $p < .001$), and PICA score (Coef. = 0.08, $p < .01$) at the alpha level =.10 (two tailed). Semantic errors were singificantly correlated with AoA (Coef. =0.15 , $p < .001$), length of target word (Coef. = 0.12, $p < .01$), age (Coef. = -0.12, $p < .01$), education years (Coef. = -0.16, $p< .01$),

Table 18. Rank-biserial correlation: Coefficient and standard errors of Somers' D for Baseline

	Correct	Semantic	Phonological	Mixed	Others
Item properties					
AoA	**-0.27** (0.03)***	**0.15** (0.04)***	0.11 (0.43)	**0.67** (0.06)***	**0.30** (0.07)***
Density	0.04 (0.03)	0.06 (0.04)	**0.55** (0.19)**	**-0.31** (0.08)***	**-0.15** (0.07)*
Length	**-0.15** (0.03)***	**0.12** (0.04)**	**-0.33** (0.16)*	**0.39** (0.09)***	0.09 (0.07)
Imageability	-0.05 (0.04)	0.03 (0.05)	0.33 (0.36)	**0.17** (0.07)*	-0.02 (0.09)
Individual properties					
Age	**0.13** (0.04)**	**-0.12** (0.05)**	**0.47** (0.23)*	**-0.16** (0.10)*	**-0.13** (0.08)*
Education yrs	**0.22** (0.04)***	**-0.16** (0.05)**	-0.40 (0.29)	**-0.32** (0.09)***	**-0.30** (0.07)***
PICA	**0.08** (0.03)**	**-0.08** (0.04)*	-0.07 (0.24)	**-0.16** (0.08)*	-0.04 (0.06)

$^{*}p < .10$, $^{**}p < .01$. $^{***}p < .001$.
Note: PNT = Philadelphia Naming Test, AoA = age of acquisition.

and PICA score (Coef. = -0.08, $p < .10$). Phonological errors were correlated singificantly with density (Coef. = 0.55, $p < .01$), length of target word (Coef. = -0.33, $p < .10$), and age (Coef. = 0.45, $p < .10$). Mixed errors were correlated singificantly with AoA (Coef. = 0.67, $p < .001$), density (Coef. = -0.31, $p < .001$), length (Coef. = 0.39, $p < .001$), imageability of target word (Coef. = 0.17, $p < .10$), age (Coef. = -0.16, $p < .10$), education years (Coef. = -0.32, $p < .001$), and PICA score (Coef. = -0.16, $p < .10$). Other error types were correlated singificantly with AoA (Coef. = 0.30, $p < .001$), density of target word (Coef. = -0.15, $p < .10$), age (Coef. = -0.13, $p < .10$) and education years (Coef. = -0.30, $p < .001$).

Correct responses were correlated singificantly with target words' AoA (Coef. = -0.12, $p < .01$), and distractors' length (Coef. = -0.07, $p < .10$), along with age (Coef. = 0.12, $p < .001$), education years (Coef. = 0.18, $p < .001$), and PICA score (Coef. = 0.08, $p < .01$). Semantic errors were correlated singificantly with taret word density (Coef. = 0.15, $p < .001$), length (Coef. = -0.07, $p < .10$), age (Coef. = -0.10, $p < .10$), education years (Coef. = -0.12, $p < 0.01$), and PICA score (Coef. = -0.07, $p < .10$). There were no variables that met the cut off p value for phonological errors (see Table 19).

Mixed errors were correlated singificantly with taret word AoA (Coef. = 0.48, $p < .001$), density (Coef. = -0.22, $p < .10$), length (Coef. = 0.34, $p < .001$), and with distractors frequency (Coef. = -0.25, $p < .01$), and length (Coef. = 0.29, $p < .01$) age (Coef. = -0.19, $p < .10$), education years (Coef. = -0.35, $p < .001$), and PICA score (Coef. = -0.13, $p < .10$). Other types of errors were correlated singificantly with the taret words' AoA (Coef. = 0.30, $p < .01$), density (Coef. = -0.16, $p < .10$), length (Coef. = 0.23, $p < .01$), age (Coef. = -0.21, $p < .10$), and education years (Coef. = -0.38, $p <.001$). Since none of these variables yielded correlations equal to or greater than rs = .70 for detecting collinearity, all potential covariates listed above were entered into the models when investigating word frequency and interaction effects of target word frequency and distractor frequency (section 4.2.3 and 4.2.4). Note that non-significant covariates were eliminated from the model to obtain parsimony.

Table 19. Rank-biserial correlation: Coefficient and standard errors of Somers' D for PWI task

	Correct	Semantic	Phonological	Mixed	Others
Item properties (Target words)					
AoA	**-0.12** (0.03)**	-0.00 (0.04)	0.15 (0.10)	**0.48** (0.07)***	**0.30** (0.09)**
Density	-0.05 (0.03)	**0.15** (0.04)***	0.02 (0.13)	**-0.22** (0.09)*	**-0.16** (0.09)*
Length	-0.03 (0.03)	**-0.07** (0.04)*	-0.03 (0.10)	**0.34** (0.08)***	**0.23** (0.09)**
Imageability	-0.02 (0.04)	0.04 (0.05)	0.14 (0.13)	-0.01 (0.10)	-0.11 (0.10)
(Distractors)					
Frequency	-0.01 (0.04)	0.04 (0.04)	0.08 (0.14)	**-0.25** (0.09)**	0.09 (0.09)
AoA	0.03 (0.03)	-0.06 (0.04)	0.14 (0.14)	0.07 (0.10)	-0.07 (0.10)
Density	0.03 (0.03)	-0.01 (0.04)	-0.06 (0.11)	-0.13 (0.09)	-0.01 (0.08)
Length	**-0.07** (0.03)*	0.03 (0.04)	0.10 (0.11)	**0.29** (0.08)**	0.07 (0.08)
Imageability	0.04 (0.05)	-0.07 (0.06)	-0.07 (0.14)	-0.04 (0.11)	0.015 (0.14)
Individual properties					
Age	**0.12** (0.03)***	**-0.10** (0.04)*	0.07 (0.12)	**-0.19** (0.09)*	**-0.21** (0.10)*
Education yrs	**0.18** (0.04)***	**-0.12** (0.04)**	0.11 (0.13)	**-0.35** (0.09)***	**-0.38** (0.08)***
PICA	**0.08** (0.03)**	**-0.07** (0.03)*	0.12 (0.09)	**-0.13** (0.08)*	-0.10 (0.08)

$^{*}p < .10$, $^{**}p < .01$. $^{***}p < .001$.
Note: PNT = Philadelphia Naming Test, AoA = age of acquisition.
Correlation coefficients in bold font meet or exceed the predetermined $p < .10$ for identifying covariates.

4.2.2 Interaction Effect Between Task Type and Response Type

The interaction effect between task type and response type involving possible association with the count variable from each individual was examined using a two-level model, with items nested within individuals as in other GLMMs. For hierarchical count data, Poisson models were analyzed using *gllamm* syntax. Task type, response type, and their interactions were entered as fixed predictors along with confounding variables including age, education years, and PICA naming performance. Unlike RT analysis, an odds ratio (OR) was used instead of a coefficient to help interpretation of discrete (i.e., non-continuous) data. By using an OR, the odds that an outcome will occur given a particular exposure, can be compared to the odds of the outcome occurring in the absence of that exposure. An OR of 1 is interpreted as “exposure does not affect odds of outcome”; OR > 1 is referred to as “exposure associated with higher odds of outcome” and OR < 1 means “exposure associated with lower odds of outcome” (Szumilas, 2010, p. 227). Like other models used in this study, confounding variables that showed no significant effect on the dependent variable, were removed for a parsimonious model.

The data consisted of a number of responses for the five response types, including correct response, semantic error, phonological error, mixed error, and “other” error types. Participant’s responses including “I don't know”, visual misinterpretation, part of picture, description/circum-locution, no response, unrelated, and non-word were categorized as “other.” Mean number of responses, for each response type, is summarized for Baseline and the PWI task respectively in Table 20. Correct responses (96.87% in Baseline; 96.55% in PWI) were produced most frequently by participants in both tasks, followed by number of semantic errors (1.92% in Baseline; 2.27% in PWI). Phonological errors were least frequently produced in both tasks (0.03% in Baseline; 0.23% in PWI). While other errors (0.70%) and mixed errors (0.48%) were produced less than semantic errors in Baseline, mixed errors (0.49%) were observed slightly more frequently than other errors (0.47%) in the PWI task.

Table 20. Mean number of response for each response type in Baseline and PWI task

	Baseline			PWI task		
Response type	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*
Correct	166.62	4.55	50	166.06	4.44	50
Semantic	3.30	2.67	50	3.90	2.53	50
Phonological	0.06	0.24	50	0.40	0.64	50
Mixed	0.82	1.12	50	0.84	1.27	50
Others	1.20	1.71	50	0.80	1.56	50

In order to investigate whether there is a significant interaction between task type and response type, Poisson models were estimated using *gllamm* syntax. Due to the nature of the count variable that was computed for each individual (not for each item), individual characteristics including age, education years, and PICA score were entered as covariates in the model. None of individual characteristics had a significant effect on the dependent variable. To compare the number of semantic, phonological and mixed errors, which are related to the word retrieval process, mixed errors was set as the reference in the model. Mean number of mixed errors is ranked 2^{nd} in Table 21 and its mean and SD were relatively constant between Baseline and the PWI task, which helped interpret the interaction effect in graphs by providing a constant slope. As a result, except for phonological errors (OR = 2.85, p = .01), no interaction effects were found for correct (OR = 0.97, p = .901), semantic (OR = 1.15, p = .557), and other errors (OR = 0.65, p = .152). In addition, no main effect for task was found (OR = 1.02, p = .913) (see Table 21). For the error types that showed no significant interaction, an effect of response type is reported as follows. Compared to the mixed errors, significant increases in correct (OR = 203.20, p < .001), and semantic errors (OR = 4.02, p < .001) were shown but non-significant change was observed for the other type of errors (OR = 1.46, p = .060).

Further analysis of phonological errors revealed a significant interaction effect. One Poisson model for task type effect and two Poisson models for the effect of response type (one for Baseline and one for the PWI task), were implemented with adjusted alpha level at p = .016 (i.e., .05/3 contrasts) by Bonferroni correction. Compared to the number of mixed errors, participants produced significantly fewer phonological errors in both Baseline (OR = 0.07, p < .001) and the PWI task (OR = 0.48, p < .01) (see Table 22 and Table 23).

Table 21. Effects of task type and response type, and interaction effect (task × response)

Fixed part	OR	SE	*P*	CI
Task type				
Baseline (ref.)	-	-	-	-
PWI	1.02	-0.22	0.913	[0.67, 1.58]
Response type				
Mixed (ref.)	-	-	-	-
Correct	203.20***	-31.81	0.000	[149.50, 276.17]
Semantic	4.02***	-0.7	0.000	[2.86, 5.67]
Phonological	0.07***	-0.04	0.000	[0.02, 0.24]
Others	1.46	-0.3	0.06	[0.98, 2.18]
Task × Response type				
Baseline × Mixed (ref.)	-	-	-	-
PWI × Correct	0.97	-0.21	0.901	[0.63, 1.50]
PWI × Semantic	1.15	-0.28	0.557	[0.72, 1.86]
PWI × Phonological	6.51**	-4.28	0.004	[1.80, 23.58]
PWI × Others	0.65	-0.2	0.152	[0.36, 1.17]
Random part	Variance	SE		
Level-2 Variance	9.49E-20	-4.70E-12		

*$p < .05$, **$p < .01$, ***$p < .001$.
Note: *n* of level 1 units = 500, *n* of level 2 units = 50.
None of covariates was included in the model due to the failure to show a significance at $p < .05$; OR = odds ratio, SE = standard errors, CI = confidence interval, ref = reference.

Table 22. Effects of response type in Baseline

Fixed part	OR	SE	*P*	CI
Response type				
Mixed (ref.)	-	-	-	-
Correct	203.20***	-31.81	0.000	[149.50, 276.17]
Semantic	4.02***	-0.7	0.000	[2.86, 5.67]
Phonological	0.07***	-0.04	0.000	[0.02, 0.24]
Others	1.46	-0.3	0.060	[0.98, 2.18]
Random part	Variance	SE		
Level-2 Variance	3.08E-18	-3.78E-11		

*$p < .016$, **$p < .01$, ***$p < .001$.
Note: *n* of level 1 units = 250, *n* of level 2 units = 50.
OR = odds ratio, SE = standard errors, CI = confidence interval, ref = reference.

Table 23. Effects of response type in PWI task

Fixed part	OR	SE	*P*	CI
Response type				
Mixed (ref.)	-	-	-	-
Correct	197.69^{***}	-30.58	0.000	[145.99, 267.71]
Semantic	4.64^{***}	-0.79	0.000	[3.33, 6.48]
Phonological	0.48^{**}	-0.13	0.006	[0.28, 0.81]
Others	0.95	-0.21	0.825	[0.62, 1.47]
Random part	Variance	SE		
Level-2 Variance	3.37E-27	-1.25E-15		

$^{*}p < .016$, $^{**}p < .01$, $^{***}p < .001$.
Note: *n* of level 1 units = 250, *n* of level 2 units = 50.

The number of phonological errors in the PWI task was approximately seven times greater than the number in the Baseline (OR = 6.67, $p < .01$) (See Table 24). This interaction is depicted in Figure 29.

Table 24. Effects of task type for phonological errors

Fixed part	OR	SE	*P*	CI
Baseline (ref.)	-	-	-	-
PWI	6.67^{**}	-4.13	0.002	[1.98, 22.43]
Random part	Variance	SE		
Level-2 Variance	9.41E-20	-3.82E-10		

$^{*}p < .016$, $^{**}p < .01$, $^{***}p < .001$.
Note: *n* of level 1 units = 100, *n* of level 2 units = 50.

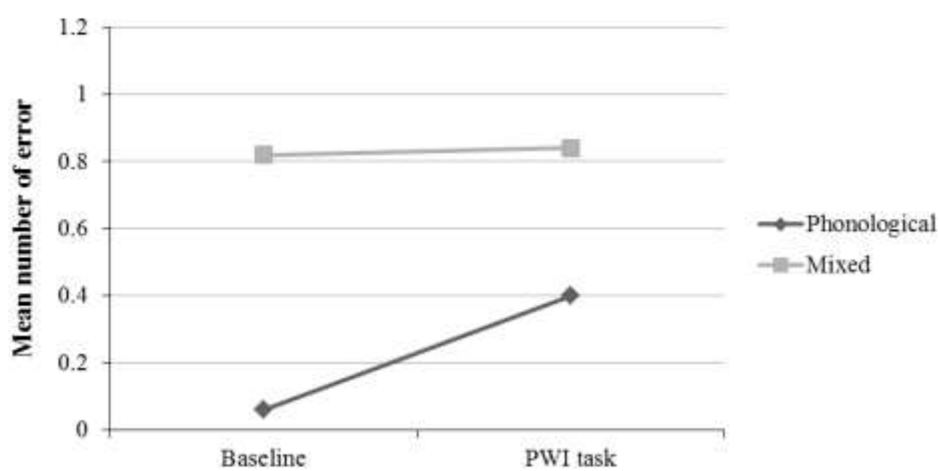

Figure 29. Mean numbers of phonological and mixed errors during Baseline and PWI task.

From among the three primary error reponse types, participants generated significantly more semantic errors followed by mixed errors and then phonological errors. The biggiest difference between two task types was that phonological errors increased significantly in the PWI task (see Figure 30).

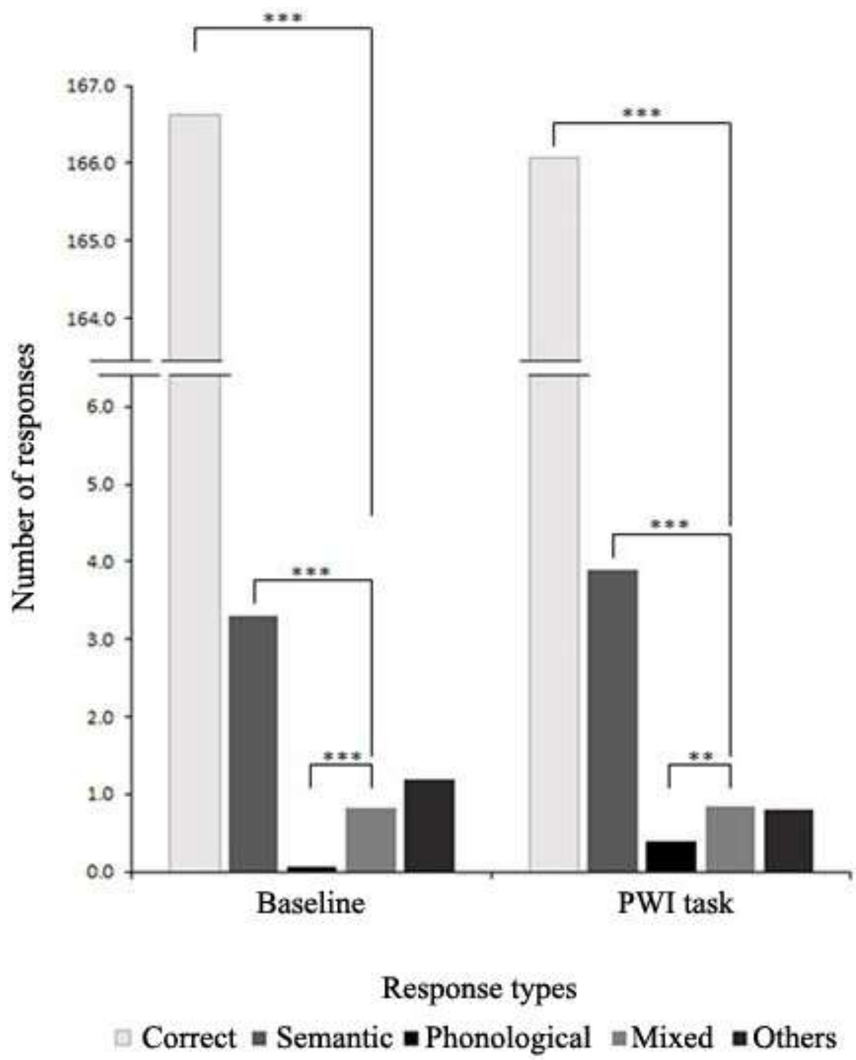

Figure 30. Number of responses for five different response types in Baseline and PWI task.
Note: Significant mean differences were indicated by $^{*}p < .016$, $^{**}p < .01$, $^{***}p < .001$ among response types and between task type.

4.2.3 Word Frequency Effect

For hierarchical binary data, taking the form of 0 for a non-occurrence of a particular response type and 1 for an occurrence of a particular response type for each item, binomial models were analyzed using *gllamm* syntax. Associations with predictors including target word frequency along with other covariates were examined in GLMMs focusing on semantic, phonological and mixed errors.

These results are reported in Table 25. None of models yielded a significant effect of word frequency on semantic, phonological, or mixed errors. All error types produced an OR = 1.00, which indicates that word frequency did not affect the odds of the outcome.

Table 25. Word frequency effect on semantic, phonological, and mixed errors in PWI task: Odds ratio and standard errors

Fixed part	Semantic errors		Phonological errors		Mixed errors	
	OR	SE	OR	SE	OR	SE
Level-1: Item						
Target word frequency	1.00	0.00	1.00	0.00	1.00	0.00
″ AoA					1.59***	-0.18
″ density	1.19**	-0.06				
″ length					1.20*	-0.11
Distractor length					1.28*	-0.14
Level-2: Individual						
Education years					0.85*	-0.06
PICA naming	0.47*	-0.17				
Random part	Variance	SE	Variance	SE	Variance	SE
Level-2 Variance	0.12	-0.08	0.01	-0.52	0.24	-0.26

* $p < .05$, ** $p < .01$, and *** $p < .001$.
Note: For the semantic errors, *n* of level 1 units = 8600, *n* of level 2 units = 50.
For the phonological errors, *n* of level 1 units = 8600, *n* of level 2 units = 50.
For the mixed errors, *n* of level 1 units = 8351, *n* of level 2 units = 50.
OR = odds ratio, SE = standard errors, AoA = age of acquisition.

4.2.4 Interaction Effect (Target Word Frequency × Distractor Frequency)

Interaction effects of target word frequency and distractor frequency on binary responses for semantic, phonological, and mixed error types were examined by estimating GLMM. Potential covariates of item properties and demographic characteristics were entered into the models, and yielded a significant effect on the dependent variable for the final model. The results are shown in Table 26. No models yielded a significant interaction effect between target word frequency and distractor frequency on semantic, phonological, or mixed errors. All yielded an OR = 1.00, which indicates that word frequency did not affect the odds of the outcome.

Table 26. Effect of interaction between target word frequency and distractor frequency on semantic, phonological, and mixed errors in PWI task: Odds ratio and standard errors

Fixed part	Semantic errors		Phonological errors		Mixed errors	
	OR	SE	OR	SE	OR	SE
Level-1: Item						
Target high freq.	2.56^{***}	-0.59	5.32^{*}	-1.21	0.81	-0.34
Distr. high freq.	1.4	-0.37	1.99	-1.82	0.47	-0.19
Target high × Distr. high	0.75	-0.24	0.33	-0.35	0.53	-0.42
Target word AoA					1.83^{***}	-0.2
″ density	1.17^{**}	-0.07				
Level-2: Individual						
Education years					0.83^{*}	-0.06
PICA naming	0.47^{*}	-0.17				
Random part	Variance	SE	Variance	SE	Variance	SE
Level-2 Variance	0.13	-0.08	0.01	-0.51	0.35	-0.25

$^{*}p < .05$, $^{**}p < .01$, and $^{***}p < .001$.
Note: For the semantic errors, *n* of level 1 units = 8600, *n* of level 2 units = 50.
For the phonological errors, *n* of level 1 units = 8600, *n* of level 2 units = 50.
For the mixed errors, *n* of level 1 units = 8551, *n* of level 2 units = 50.
OR = odds ratio, SE = standard errors, freq. = frequency, Distr. = distractor, AoA = age of acquisition.

5.0 DISCUSSION

In this study the word frequency effect was examined by adopting a PWI paradigm to provide empirical evidence to address the locus of the word frequency effect. As a starting point, some assumptions were made about the distractor conditions used in the current study to support the main findings and their application to the study's specific aims. The assumptions were that in Baseline, where distractors were not presented, participants' RTs would not be significantly different among distractor types implicating that no inhibition and facilitation effects exist due to the absence of distractors. However, in the PWI task, it was predicted that participants would show significantly different patterns of naming behavior in terms of RTs depending on the distractor conditions representing each computational step of lexical access plus their interactions. That is, it was predicted that, relative to the Baseline, there would be significantly: 1) increased RTs in the semantic distractor condition (i.e., an inhibition effect); 2) decreased RTs in the phonological distractor condition (i.e., a facilitation effect); and 3) increased RTs in the mixed distractor condition but less increased RT than in the semantic distractor condition due to the influence of a facilitation effect derived from the phonological prompts of the distractors.

The results showed no significant differences among distractor types in Baseline except for the mixed distractor condition. Participants showed significantly reduced RTs in the mixed distractor condition compared to the neutral, unrelated distractor condition. This outcome violates the assumptions above. However, a negative coefficient for the mixed distractor condition in the Baseline task switched to a positive coefficient in the PWI task. This indicates that mixed distractors affected the naming performance strongly enough to change the direction of the coefficient. Thus, this finding strongly supports the interpretation that the established mixed distractor stimuli successfully played a role in revealing the interactive connection between the two steps. Few studies have successfully demonstrated an interactive stage with a PWI paradigm. A successful manipulation of the network between the two steps by using mixed distractors provides evidence for the word frequency effect's location within the interactive mechanism.

Importantly, findings from the PWI task provide strong evidence that each of the assumptions were in fact met. There were significantly increased RTs in the semantic distractor condition compared to the neutral, unrelated distractor condition. This indicates that semantic distractors successfully manipulated lexical selection at the semantic representation level by showing an inhibition effect. Conversely, significantly decreased RTs in the phonological distractor condition indicates that those distractors affected phonological encoding. Lastly, increased RTs were observed in the mixed distractor condition. The amount of RT increased was less than the amount in the semantic distractor condition and more than the amount in the phonological distractor condition. This result suggests that both inhibition and facilitation effects occurred during picture naming when mixed distractors were presented. In conclusion, these results support the two-step interactive model.

In order to address the first research aim, the following section discusses the RT findings relative to the word frequency effect for each distractor condition.

5.1 LOCUS OF WORD FREQUENY EFFECT

The results of the PWI experiment reveal that word frequency significantly affects RTs in the semantic, phonological, and mixed distractor conditions. Negative coefficients that were obtained for all conditions indicate that participants named the HF picture items faster than LF items, showing a HF word advantage at all lexical access steps. This finding is at odds with Jescheniak & Levelt (1994) who proposed that word frequency affects only the phonological stage of processing. Simultaneously, by showing the effects on RTs in both semantic and phonological distractor conditions, the results of this study support the argument that the locus of the frequency effect is evidenced not only at phonological encoding but also at lexical selection (Caramazza et al., 2001; Cuetos et al., 2010; Gahl, 2008; Kittredge et al., 2008; Navarrete et al., 2006).

Indeed, the results are consistent with the Kittredge et al. (2008) proposed role of interactivity in word frequency effects on lexical selection. From the studies of error analysis for PWA, they had assumed that the effect on lexical selection was due to the transmission of activated phonological nodes to the semantic representation level via the spreading activation feedback. As discussed above, their data provided inconclusive evidence because the mixed errors that play a critical role to support the interactivity, was not included in their analysis. The current study successfully manipulated the interactive network between the lexical and phonological stages by using mixed distractors, demonstrating that word frequency affects both stages, presumably by shared spreading activation. This is interpreted as support for the IA model (Dell & O'Seaghdha, 1991; Dell, 1986; Foygel & Dell, 2000) where the two steps are not completely independent of each other but influence the other through an interactive network. Based on the findings of reduced RT with increased word frequency in the semantic distractor condition, as well as in the mixed distractor condition, a direct and indirect route for the word frequency effects on lexical selection seems plausible. This not only identifies the presence of the word frequency effect on lexical selection, it also provides a plausible mechanism or route for that influence.

In addition, by providing each coefficient for the semantic, phonological, and mixed distractor conditions, a relative magnitude of RT change as a function of word frequency was identified within a single study, which has not been explored in the literature. In other words, not only the existence of a word frequency effect at each lexical retrieval step, but also how much each step is affected by the frequency effect was revealed. Consistent with the literature, the robust frequency effect on RT in the phonological distractor condition was identified. However, its magnitude (Coef. = -0.01) was relatively small compared to those in the semantic (Coef. = -0.06) and the mixed distractor conditions (Coef. = -0.05). This finding led to a conclusion that the word frequency effect on the lexical selection and the interaction between steps is stronger than expected. Note that a direct comparison about the frequency effect on phonological encoding between this study and previous studies is challenging for several reasons: 1) technical issues on computing effect sizes from GLMMs; 2) different experimental tasks used in studies that may induce different RT ranges across participants; and 3) different data type (i.e., binary data in the literature as HF and LF words, but continuous data in the current study) that requires appropriate statistical approaches to obtain effect sizes.

Regarding the covariate results, the results haves revealed that not only word frequency, but also other confounding factors such as a target or distractor word's AoA, density, length, image, and such biographical and performance variables as age, education years, and PICA scores are related to the RT. Despite an apparent influence of these factors on performance, word frequency effects across the three major distractor conditions were robust.

Contrary to some previous research, the impact of each lexical variable differed depending on the location of the lexical access step. Generally, researchers have found better performance with earlier AoA, higher imageability, shorter word length (Cuetos et al., 2002; Nickels & Howard, 1995; Plaut & Shallice, 1993) without considering processing step. Regarding phonological neighborhood density, some researchers have argued that high density words impact the ease of word productions due to the more phonological information recalled and reinforced feedback to a target lemma (e.g., Dell & Gordon, 2003; MacKay & Burke, 1990; Middleton & Schwartz, 2010; Vitevitch, 2002). Others reported the opposite due to the competing phonologically related words that block retrieval of the target word (e.g., Maylor, 1990; Newman & German, 2005; Schacter, 1999). The current study showed a negative effect of imageability that appeared in the semantic and phonological distractor conditions, while a positive effect was shown in the mixed distractor condition. Imageable objects with richer semantic representations do not always result in better naming performance at the lexical selection or phonological encoding steps. Also, some variables that significantly impacted RTs did not appear significant in some distractor conditions. That is, there was no AoA or density influence on RTs in the phonological distractor condition, although such influence was present in the semantic distractor condition.

Compared to the influences of lexical variables at each processing step, individual-level variables such as age and PICA scores showed a robust impact on RTs. Interestingly, education years did not show either a consistent positive or negative influence on RTs. Older participants showed a significantly greater delayed response time than younger participants, and higher scores on the PICA test related to better performance. The different impacts between lexical variables and individual variables for the distractor conditions seem to be due to the high association of the lexical variables with lexical retrieval processing. Further studies will be needed to address the influence of lexical variables at each lexical retrieval step. This discussion will focus on the word frequency and distractor frequency effects.

Finally, there were several reasons to adopt hierarchical multilevel effect models in this study. Variation in RTs during picture naming, due to individual differences, was predicted due to differences in resting activation level and/or selection thresholds. Additionally, confounding factors such as lexical properties and demographic characteristics could affect variability. As predicted, a large variance between individuals was found in the models and these individual differences were considered when computing the estimates. The approach using GLMM enhanced the sensitivity of the study by increasing power by decreasing type II error.

In conclusion, the RT findings from this study support the first hypothesis; that there would be a decrease in RTs in both the semantic and mixed distractor conditions as the target item's frequency increases. These findings are inconsistent with the established DTS model (Jescheniak & Levelt, 1994) that word frequency has a selective influence at the phonological encoding step. By demonstrating the indirect word frequency effect on the lexical selection step via the interactive lexical network, the bi-directional spreading of activation of IA models was supported (Kittredge et al., 2008). Also, findings further extended understanding of the word frequency effect on lexical selection. That is, not only its indirect influence, but also how its direct influence operates during lemma selection.

5.2 DISTRACTOR FREQUENCY EFFECT AND ITS INTERACTION WITH TARGET WORD FREQUENCY

The second specific aim was to investigate whether the target item's frequency interacts with distractor frequency during lexical selection. Since the literature did not provide information about the relationship between target word frequency and distractor word frequency, it was expected that the current study would extend the understanding of distractor frequency effect in the PWI task.

Two hypotheses were proposed: competitive lexical selection and non-competitive lexical selection. In summary, the competitive hypothesis predicts that the speed of selecting target lexical nodes is dependent on the activation levels of non-target nodes. Therefore, HF distractor words have stronger interference effects on the retrieval of target words than LF distractor words (Damian & Martin, 1999; Hutson et al., 2013; La Heij, 1988; Levelt et al., 1999; Roelofs, 2003). The non-competitive hypothesis predicts that a lexical node is selected without influencing the activation level of non-target nodes, and thus, LF distractor words should interfere more in a PWI task (Caramazza, 1997; Dell, 1986; Dhooge & Hartsuiker, 2010, 2011; Mahon et al., 2007; Miozzo & Caramazza, 2003; Rapp & Goldrick, 2000; Stemberger, 1985). This hypothesis views the distractor frequency effects as occurring post-lexically, while the LF distractors are held longer than HF distractors, resulting in more interference.

Before examining the interaction between target word frequency and distractor frequency, it is necessary to discuss the GLMM results for each distractor condition in order to understand the fundamental impact of the distractor frequency on picture naming. Note that in the models, the distractor frequency effect was entered as a covariate predictor to examine its effect on RT holding the effect of target word frequency. Results showed that an increase in distractor frequency led to delayed RT in semantic and mixed distractor conditions, supporting the competitive hypothesis. Consistent with the competitive hypothesis, HF distractor words interfered more with the retrieval of target words than LF distractor words during the lexical selection step. This is at odds with work claiming that LF distractors interfered more (Dhooge & Hartsuiker, 2010, 2011; Mahon et al., 2007; Miozzo & Caramazza, 2003; Rapp & Goldrick, 2000). This distractor frequency effect was not found for the phonological distractor condition. Following Miozzo and Caramazza (2003), it seems that interaction with the facilitation effect of the phonological distractors led to the disappearance of the distractor frequency effect.

A significant interaction effect between target word frequency and distractor frequency was observed only in the mixed distractor condition, indicating that the interference of distractor frequency type depends on the target frequency type. The mean RT of HF target words was higher than the LF target words when HF distractors were provided and the pattern was reversed when the LF distractors were presented although the difference of interference was small. The magnitude of difference between the mean RT of HF target words and the mean RT of LF target words was higher in the HF distractor condition than in the LF distractor condition. A possible interpretation would be that HF distractors in the mixed distractor condition seemed to influence lexical retrieval of target words but their influence was mitigated in the interactive network by the partial facilitation effect that propagated back from the phonological representation level. However, findings of no simple main effect of distractor frequency type makes it hard to argue that a certain type of distractor word's frequency had a stronger interference effects on the retrieval of a certain type of target word's frequency.

In sum, based on positive (+) coefficients of distractor frequency on RTs shown in Table 14 and 16, the competitive hypothesis seems to account for the overall structure of the distractor frequency effect in the current study. In addition, findings regarding the interaction effect in the mixed distractor condition support the hypothesis 2a that the target word's frequency does not independently modulate the lexical retrieval speed without being influenced by distractor frequency. However, it is difficult to select either of these hypotheses to account for the significant interaction effect.

5.3 RESPONSE TYPE ANALYSIS

A large number of correct responses were observed at Baseline and in the experimental conditions. There was no significant difference in the errors between Baseline and the PWI task except for the number of phonological errors. However, it was not necessarily phonological distractors that led to the increase in phonological errors. It appears that the distractors in the experimental setting caused some participants to generate phonological errors, an atypical type of error, for healthy adults. There was also a small variance at level-2 in the models, which means that there was little variance across participants in terms of the number of errors.

Focusing on the errors in the PWI condition, a significant increase in semantic errors and a significant decrease in phonological errors were shown compared to the number of mixed errors. This finding is consistent with the observation that semantic and mixed errors are the most frequent error types in healthy adults (Dell & Reich, 1981; Harley, 1984; Martin et al., 1989; Schwartz et al., 2006).

Regarding specific aim 3, non-significant word frequency and non-significant interaction effects were found between target word and distractor frequencies across all error types. It turns out that error rates are too low for healthy adults to be meaningful. Thus, the first hypothesis of specific aim 3 was supported: examining RT provides a sensitive dependent measure with which to investigate the word frequency effect in a PWI paradigm by increasing the statistical power. A better dependent variable for indexing the word frequency effect (Miozzo & Caramazza, 2003) is captured by analyzing RT rather than errors when the PWI paradigm is used; at least when assessing healthy adults. Although the amount of phonological errors increased in the distractor conditions, it would be premature to conclude that the PWI paradigm can be used to create erroneous naming behavior in healthy adults or to assess the normal-to-aphasic continuum (continuity hypothesis) using the experimental design of the current study.

5.4 CLINICAL IMPLICATION

Many studies of aphasia treatment have focused on the recovery of language impairment due to acquired brain damage (Basso, Capitani, & Vignolo, 1979; Kiran & Thompson, 2003; Nickels, 2002; Rose, Douglas, & Matyas, 2002; Shewan & Kertesz, 1984). In order to achieve the ultimate goal of aphasic intervention, which is maximizing quality of life (Simmons-Mackie & Kagan, 2007; WHO, 2001), extending researchers' and clinicians' attention to the functional and/or preserved language ability is required. In this sense, the current study has clinical implications for aphasia treatment focusing on PWA's relatively well-preserved word retrieval ability for HF words. Findings showed that HF word advantage can be gained across all lexical retrieval steps. Accordingly, no matter where the damage is located in the lexical processing network, PWA with semantic network damage should still benefit from the HF word advantage with the relatively retained phonological encoding, and the PWA with phonological network damage may benefit with their relatively retained lexical selection before they articulate a word.

Via the interactive network between two steps, the damaged network expects to be strengthened by other nodes in adjacent representation levels that are connected to the weak target nodes. With frequently usable words, the treatment goal for aphasia may be more related to the improvement of daily communication. Yet to be determined is the influence of word forms. In order to expand the findings from the literature and to enhance understanding about the word frequency effect on retrieving nouns, the current study focused only on nouns. Many word forms are used in our language production. Thus, future research needs to examine the locus of word frequency effect for a wide range of word forms.

In addition, findings have implications for the current AAC treatment for PWA. Many graphic symbol-based AAC systems that were developed for PWA do not consider HF word advantage on the AAC vocabulary selection and organization. Identifying the locus of the word frequency effect, thus, was considered important in the present study to provide a rationale with which to prioritize HF words in the AAC system. From this study, word frequency was confirmed to be a relevant psycholinguistic variable for lexical selection. Thus, we can predict that a HF word advantage can be expected when using a graphic symbol-based interface where semantic information is represented. More importantly, based on the findings about the direct and indirect influence of word frequency on lexical selection, improved AAC performance is anticipated with well-organized access to the HF word symbols.

Findings about the distractor frequency effect during lexical access provided another possibility regarding AAC vocabulary organization. Non-target symbols can interfere with selecting a symbol or word when the non-targets are semantically related. Therefore, organizing words in a semantic hierarchy would not be necessary or viable for effective communication using an AAC system. Rather, allocating the highest frequency words on the main page of the interface seems practical for efficient lexical selection using symbols. Then the remaining HF words can be organized by grammatical categories. Based on the evidence about the interaction effect between the target word's frequency and the distractor frequency (non-target words), the magnitude of interference from non-target word on retrieving words in the AAC interface would differ depending on the target word's frequency type (i.e., HF or LF words). However, HF words are not necessarily affected more by HF non-target words than LF non-target words. Thus there are less constraints when allocating those words.

As the results of this study demonstrate, lexical level variables as well as individual-level variables may affect word retrieval. In addition to these variables, other interface variables may affect the performance too. Thus, additional word frequency effect studies that include the factors of AAC technology and AAC speakers will be required to explicate these interactions. Additionally, for the same reason that was mentioned above for the clinical implication for apahsia treatment, the current study focused only on nouns. AAC, and other interventions, must utilize not only nouns but also pronouns, verbs, adjectives, and preposition when considering word frequency effects (Hill 2010). Thus, when the word frequency effect studies are conducted for AAC speakers in the future, all of these form classes need to be included in order to assess the generalizability of the findings of this study.

The clinical motivation for investigating the WF effect was its application to AAC. However, findings can be extended to selecting stimuli for treatment for PWA or other populations. Additionally, PWI task may be applicable to locating the areas of impairment in PWA as well with other clinical populations.

5.5 LIMITATIONS

In spite of the advantages of using word frequency as a continuous variable and the use of the GLMM, the present study has a limitation to be addressed in terms of the statistical method. In order to examine the interaction effect between target word frequency and distractor frequency, the continuous variable had to be sorted into HF and LF words when using *gllamm* syntax in StataSE 14. Medians of target word frequencies and distractor frequencies were used as cut off points to divide the frequencies into two groups because of the non-normal distribution. For normal distributions, the mean, median, and mode can be used as a central value, but for a skewed distribution, the median is typically used (Gravetter & Wallnau, 2013). Although meaningful results were found to account for the interaction effect from the mixed distractor condition, discrepancies were shown between the GLMM results where continuous variables were entered and the GLMM results where categorical variables were entered. As expected, the latter GLMM failed to show the main effect of target word frequency and/or distractor frequency effect due to the low sensitivity of the measurement. Because the primary purpose of the latter GLMM was to examine the interaction effect, non-significant main effects were ignored in the current study.

Another limitation is related to the target item stimuli. Since this study adopted a PWI paradigm, the PNT pictures were used, which have been demonstrated to have high familiarity, high name agreement, and good image quality (Roach et al., 1996). In the current study, the lexical properties including word frequency, word length, phonological neighborhood density, and imageability were all controlled. Also, pictures were presented in a random order with and without the paired auditory distractors. Thus, possible confounding effects that pertained to the pictures were controlled. However, as the coefficients showed, the amount of RT change relative to the change of word frequency units was relatively small. For example, in the semantic condition, as a unit of word frequency increased by 100, the RT was reduced 2ms. The small amount of change seems to be due to the relatively high performance of participants with the LF words and relatively poor performance with the HF words for some participants. In order to improve the sensitivity of the stimuli, providing some cognitive stress such as a time-constrained environment during their picture naming could be considered for future studies (Silkes, McNeil, & Drton, 2004). In addition, conducting the same research procedure for PWA and comparing the results with the ones for healthy adults will be helpful to determine if the low sensitivity of the stimuli was due to the influence of the demographic characteristic.

Lastly, unrelated or involuntary reactions such as interjection (e.g., um, ah), cough, and yawn that preceded their verbal responses, were shown from some participants. Follow-up analysis showed that among 16,589 verbal responses that were included for the RT analysis, the total number of these reactions was 488 (2.94% to the total verbal responses): 425 interjections (2.56%), 19 throat clearings (0.11%), 13 laughs (0.08%), 8 coughs (0.05%), 7 yawns (0.04%), 11 carrier phrases (e.g., “that’s a”) (0.07%), 2 hiccups (0.01%), 2 sighs (0.01%), and 1 sniffle (0.01%). A close observation was made on the majority of interjections revealing that many of interjections were used by some participants as an effort to retrieve a word. Along with other involuntary reactions, the possibility exists that interjections might affect the onset of speech sounds, resulting in longer RTs. However, these interjections were dispersed across conditions and did not appear to have a differential effect on the experimental conditions. Thus, there is no likely effect on the results obtained.

6.0 CONCLUSION

The purpose of the current study was to identify the locus of the word frequency effect in order to provide theoretical and empirical justification for the use of AAC systems for PWA. The picture-word interference (PWI) paradigm was used to investigate these effects in healthy adults and the responses were analyzed using GLMMs to achieve three specific aims.

Specific aim 1 investigated whether word frequency affects lexical selection during a PWI task. Results showed that word frequency affects not only phonological encoding, but also the lexical selection and the interaction between these two stages of processing. This finding supports the rejection of the established viewpoint, derived from DTS models, that word frequency has a selective influence at the phonological encoding step.

Specific aim 2 investigated whether the target item's frequency interacts with the distractor's frequency during lexical selection, within the PWI task. No interaction effect was found at lexical selection or at phonological encoding. A significant interaction effect was found at the locus between the two steps without a simple main effect of distractor frequency type. Overall, these findings are consistent with the competitive hypothesis. However, to account for the mechanism of the interaction effect at the interactive network, further research is needed.

Specific aim 3 investigated whether there is a difference between RT and response type, when examining the frequency effect for healthy adults during the PWI task. As predicted, the frequency effect was observed differently depending on the type of dependent variable. Due to the issue of sensitivity of measurement, neither the word frequency effect nor the interaction effect between target word frequency and distractor frequency was observed when errors were investigated. RT was found to be a sensitive measurement when investigating the word frequency effect while using the PWI paradigm and in general, error type was not.

Although some limitations regarding the statistical approach and stimuli were identified, this study provides empirical data addressing the debate on the locus of the frequency effect on lexical retrieval that is prominent in aphasiology and psycholinguistics. In addition, the study extends the understanding of the characteristics of the word frequency effect and the interactive mechanism in the lexical retrieval process. These findings will help to develop theory-driven AAC treatment approaches, which are expected to enhance the autonomy and expand social networks for PWA.

APPENDIX A

PNT Target Picture Items and Paired Distractor Stimuli For Semantic, Phonological, Mixed, and Unrelated Distractor Conditions

	Semantic distractor condition (n=43)		Phonological distractor condition (n=43)		Mixed distractor condition (n=43)	
#	Target	Distractor	Target	Distractor	Target	Distractor
1	zebra	buffalo	dice	diaper	scissors	sickle
2	octopus	squid	grapes	grave	apple	apricot
3	celery	onion	Eskimo	escalator	balloon	babble
4	owl	eagle	necklace	nest	pineapple	papaya
5	strawberries	watermelon	camera	candy	lamp	lantern
6	cannon	missile	binoculars	biscuit	skull	scapula
7	vest	cardigan	ruler	ruby	dog	donkey
8	vase	pitcher	cowboy	cocktail	carrot	cabbage
9	bridge	tunnel	calendar	caramel	bus	buggy
10	fireplace	radiator	can	cactus	tent	temple
11	dragon	unicorn	cane	cave	flashlight	flame
12	pyramid	sphinx	saddle	sack	camel	calf
13	typewriter	printer	piano	picture	goat	goose
14	hammer	drill	wig	wizard	sandwich	salad
15	fork	chopstick	drum	drugstore	tractor	truck
16	helicopter	airplane	pencil	pentagon	train	trolley
17	towel	napkin	table	tail	basket	bag
18	lion	tiger	fan	family	horse	hound
19	ghost	angel	pen	pebble	pie	pizza
20	pipe	cigarette	belt	beggar	pumpkin	potato
21	football	hockey	slippers	slingshot	squirrel	skunk
22	rope	chain	map	mango	cake	custard
23	candle	torch	queen	quilt	bat	bear

(continued)

	Semantic distractor condition (n=43)		Phonological distractor condition (n=43)		Mixed distractor condition (n=43)	
#	Target	Distractor	Target	Distractor	Target	Distractor
24	closet	pantry	ring	river	nurse	nun
25	van	sedan	cat	canopy	nail	needle
26	bread	cookie	nose	notebook	elephant	elk
27	leaf	petal	stethoscope	steak	turkey	turtle
28	kitchen	bathroom	ear	easel	seal	seahorse
29	bottle	cup	king	kidney	spoon	spatula
30	glass	mug	plant	plastic	hat	helmet
31	church	library	fish	field	sailor	salesman
32	tree	moss	window	wing	butterfly	bumblebee
33	bed	couch	letter	lemon	heart	hexagon
34	eye	mouth	baby	bacon	skis	skate
35	well	reservoir	waterfall	wallet	pear	peach
36	wagon	cart	pillow	pianist	harp	harmonica
37	bowl	jar	corn	court	broom	brush
38	dinosaur	mammoth	garage	garment	monkey	mole
39	banana	orange	comb	coach	rake	razor
40	fireman	lifeguard	desk	devil	pig	panda
41	bench	sofa	anchor	antique	mountain	mound
42	cow	lamb	mustache	mud	duck	dove
43	boot	heel	chimney	chin	spider	sparrow

	Unrelated distractor condition (n=43)	
#	Target	Distractor
1	cheerleaders	eggplant
2	zipper	bicycle
3	snail	ocean
4	pirate	squash
5	sock	radio
6	volcano	snack
7	bride	rocket
8	saw	golf
9	ambulance	oven
10	whistle	garden
11	iron	ankle
12	scarf	bamboo
13	crown	pin

(continued)

Unrelated distractor condition (n=43)

#	Target	Distractor
14	cross	peanut
15	scale	kangaroo
16	glove	parrot
17	snake	broccoli
18	beard	plate
19	toilet	jelly
20	clown	pelican
21	clock	valley
22	knife	ant
23	suit	garlic
24	hair	stapler
25	shoe	tank
26	flower	elevator
27	star	toaster
28	ball	jacket
29	sun	dormitory
30	top	cucumber
31	kite	boat
32	foot	cottage
33	book	gym
34	key	basil
35	crutches	doughnut
36	hose	prince
37	thermometer	doll
38	microscope	cereal
39	frog	wrench
40	bell	attic
41	door	cherry
42	bone	guitar
43	chair	jaw

Note: Each distractor was created considering the distractors used in previous PWI studies and description of error types in literature. Semantic distractors are semantically related words within the same semantic category as the paired target items. Phonological distractors are words of which the first syllable corresponds to the first syllable of paired target items. Mixed distractors share a same semantic category and a phonological sound for at least first consonant or vowel with paired target items. Unrelated words are not semantically or phonologically related to the paired items. No item was repeated in any condition.

APPENDIX B

Instructions Provided to Participants in Baseline and PWI Tasks

1. Instruction provided through E-Prime for the Baseline Task

"I am going to ask you to name some pictures. *When you see a plus sign (+) on the computer screen with a beep sound,* *a picture will appear on it.* *Your job is to name the picture as fast as you can when you see it.* *Please use only one word."*

2. Instruction provided through E-Prime for the PWI task

"This time you will again name some pictures as fast as you can when you see a plus sign (+) on the computer screen with a beep sound. *However, this time you will hear spoken words with the picture.* *Your job is to name the picture as soon as possible and pay no attention to the spoken words.* *Sometimes the words will be spoken after you see the picture. In that case, don't wait until the spoken word is finished, but name the picture as fast as you can.* *Again, use only one word to name the picture.* *We'll practice before we begin."*

3. In the middle of both tasks, the following written instruction was provided once:

"Keep naming the picture as soon as possible."

APPENDIX C

Example of Response Sheet

Participant ID: Date: Examiner:	• Response Time (RT): Obtained based on E-Prime outcome report. • Response Type: C = correct response, S = semantic error, P = phonological error, M = mixed error, or O = others. Indicate a specific subtype of "Others" from followings: *I don't know, omission of more than 50% of number of phonemes, description/circumlocution, part of picture, no response, visual misinterpretation, non-word, and unrelated word.*

#	**Baseline**					**PWI**				
	Target word	Distractor type	Response	Response type	RT (ms)	Target word	Distractor type	Response	Response type	RT (ms)
1	sun	Unrelated				belt	Phonological			
2	ruler	Phonological				scale	Unrelated			
3	banana	Semantic				wig	Phonological			
4	waterfall	Phonological				sailor	Mixed			
5	towel	Semantic				crutches	Unrelated			
6	ring	Phonological				ear	Phonological			
7	cane	Phonological				wagon	Semantic			
8	stethoscope	Phonological				typewriter	Semantic			
9	dragon	Semantic				volcano	Unrelated			
10	corn	Phonological				snake	Unrelated			

APPENDIX D

Response Type Analysis

The first response of participant to each item was examined and checked for correctness. Correct response should be identical to the target word except for the following words: “snare drum” for “drum”; “ink pen” for pen”; “fall” for “waterfall”; “firefighter” for “fireman.” Therefore, if a participant said “snare drum” for a picture “drum”, it was regarded as a correct response. These words were excluded from the RT analysis because the phonemes that are unmated with the ones of a target word can result in confounding influence on examining frequency effect on lemma and lexeme of the target word.

Incorrect responses, which were not covered by the specifically identified exceptions, were considered to be errors. All errors were to be sorted out into four error categories. Three major error types including semantic, phonological, and mixed errors were defined based on the guideline of Schwartz et al. (2006):

(1) “Semantic” is a synonym of the target, or a coordinate (e.g., toad for frog), superordinate (e.g., cup for glass) or subordinate member of its category (e.g., bible for book). Noun associates are also included in the semantic error category (e.g., bride for wedding), whereas non-noun associates are not; they are considered to be “Other” type of errors and coded in the category “description/circumlocution” (e.g., bride for getting married; or marrying).

(2) “Phonological” is any word response (excluding proper nouns or non-words) that meets the Philadelphia Naming Test’s phonological similarity criterion. This criterion requires that the target and the error start or end with the same phoneme, that they have a phoneme in common at another corresponding syllable or word position, aligning words left to right, or that they have more than one phoneme in common in any position (excluding unstressed vowels) (e.g., “peer” for “pillow”).

(3) “Mixed” is a response that meets both the semantic and the phonological similarity criterion.

(4) Other types of errors will be grouped as “other” which will include no response, “don’t know” response, omission of more than 50% of number of phonemes (e.g., “vol” for “volcano”), description/circumlocution (“woman wearing a wedding gown” for “bride”), naming a picture part (e.g., “car” for a picture of “garage” where a “car” is parked inside), visual misinterpretation (e.g., “wrench” for “bone”), non-word, and unrelated words that do not meet neither of the criteria of semantic and phonological errors. Repeating a name produced earlier in the list is regarded as unrelated type of errors (i.e., perseverations). Note that in the case of “omission”, not all incomplete words are errors. If the number of phonemes is more than 50% of full phonemes, the incomplete words were treated as complete response. For example, if the response is /vɑl/ for “volcano”, the number of phonemes is less than 50% (3 out of 7) so it is not considered as “volcano”. However, if the response is /vɑlkeɪ/, more than 50% phonemes are produced, thus, the response is considered as “volcano.”

APPENDIX E

Acoustic Characteristics of Vowels and Consonants

The onset of each speech sound was examined by taking into account following acoustic characteristics of phonemes provided by Kent and Read (1992) and an online lecture, *Speech Waveforms* by Mannell (Retrieved from http://clas.mq.edu.au/speech/acoustics/waveforms /speech_waveforms.html on April 1, 2015). Note that in the case of appearing creaks resulting from a perturbation of vocal fold vibration, the break of the continuous waveform was ignored as long as the waveform showed one single word in the current study. Repeated initial phonemes which did not compose a whole single word were excluded from the data analysis.

1. Vowels

All voiced have periodic waves. The intensity of the vowels rises rapidly at the start, and then gradually drops.

2. Consonants

1) Stops

a. /p/ /t/ /k/: These voiceless oral stops have characteristic of aperiodic sound commencing abruptly with a burst. The bursts are very short and are followed by aspiration.

b. /b/ /d/ /g/: These voiced oral stops have an initiation of voicing that precedes the stop burst. In the case of unclear voicing, the bursts occur immediately before the onset of the vowel.

2) Fricatives

a. /f/ /ɵ/ /s/ /ʃ/ /h/: These voiceless fricatives have a gradually increasing aperiodic pattern but no bursts. These sounds have relatively long fricative aspiration compared to the aspiration of the voiceless stops and affricate. /s/ /ʃ/ have the stronger voiceless fricatives than /f/ /ɵ/ and /h/ do.

b. /v/ /ð/ /z/ /ʒ/: These voiced fricatives have periodic pattern. For the strong fricatives /z ʒ/, the mixture of periodicity and aperiodicity may be seen before the vowel.

3) Affricates

a. /tʃ/: This voiceless affricate has a weak burst. Then a very strong aspiration appears before the onset of voicing.

b. /dʒ/: The burst in this voiced affricate's waveform is barely discernable. It would look similar to the waveform of the voiced fricative /z/.

4) Nasals

Voiced nasals /m/ /n/ /ŋ/ have relatively periodic pattern without bursts like vowels but have weaker intensity than vowels.

5) Liquids and glides

Voiced liquids /l/, /r/ and voiced glides /w/ /j/ look like voiced nasals.

BIBLIOGRAPHY

Aftonomous, L. B., Steele, R. D., & Wertz, R. T. (1997). Promoting recovery in chronic aphasia with an interactive technology. *Archives of Physical Medicine and Rehabilitation*, *78*, 841–846.

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano* [Word frequency dictationary for Spanish Castillian]. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

Alario, F. X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: implications for models of lexical access. *Language and Cognitive Processes*, *17*(3), 299–319.

Anderson, C. J., Verkuilen, J., & Johnson, T. (2010). *Applied generalized linear mixed models: Continuous and discrete data*. New York: Springer.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). The CELEX lexical database (CD-ROM). Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Balandin, S., & Iacono, T. (1998). A few well-chosen words. *Augmentative and Alternative Communication*, *14*(3), 147–161.

Balandin, S., & Iacono, T. (1999). Crews,Wusses, and Whoppas: core and fringe vocabularies of Australian meal-break conversations in the workplace. *Augmentative and Alternative Communication*, *15*(2), 95–109.

Barry, C. M., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *The Quarterly Journal of Experimental Psychology. A. Human Experimental Psychology*, *50*(3), 560–585.

Basso, A, Capitani, E., & Vignolo, L. A (1979). Influence of rehabilitation on language skills in aphasic patients. *Archives of Neurology, 36*, 190–196.

Bayles, K., & Tomoeda, C. (1993). *The Arizona Battery for Communication Disorders in Dementia*. Tucson, AZ: Canyonland.

Beukelman, D. R., Ball, L. J., & Fager, S. (2008). An AAC personnel framework: adults with acquired complex communication needs. *Augmentative and Alternative Communication (Baltimore, Md. : 1985)*, *24*(3), 255–267.

Beukelman, D. R., Hux, K., Dietz, A., McKelvey, M., & Weissling, K. (2015). Using visual scene displays as communication support options for people with Chronic, severe aphasia: A summary of AAC research and future research directions. *Augmentative and Alternative Communication*, *31*(3), 234–245.

Beukelman, D. R., Jones, R., & Rowan, M. (1989). Frequency of word usage by nondisabled peers in integrated preschool classrooms. *Augmentative and Alternative Communication*, *5*(4), 243–248.

Beukelman, D. R., McGinnis, J., & Morrow, D. (1991). Vocabulary selection in augmentative and alternative communication. *Augmentative and Alternative Communication*, *7*(3), 171–185.

Beukelman, D. R., Yorkston, K. M., Poblete, M., & Naranjo, C. (1984). Frequency of word occurrence in communication samples produced by adult communication aid users. *The Journal of Speech and Hearing Disorders*, *49*(4), 360–367.

Blumstein, S. (1973). *A phonological investigation of aphasic speech*. The Hague, Netherlands: Mouton.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. Gernsbacher, Academic Press.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135.

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*(2), 135–140.

Bulmer, M. G. (1967). *Principles of statistics* (2nd ed.). Cambridge, MA: M.I.T. Press.

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA: MIT Press.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177–208.

Caramazza, A., Alario, F. X., & Costa, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review: A Journal of the Psychonomic Society, Inc*, *12*(1), 125–131.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430–1450.

Chertkow, H., & Bub, D. (1990). Semantic memory loss in dementia of alzheimer's type. *Brain*, *113*(2), 397–417.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology. A. Human Experimental Psychology*, *33*(A), 497–505.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Cuetos, F., Aguado, G., Izura, C., & Ellis, A. W. (2002). Aphasic naming in Spanish: predictors and errors. *Brain and Language*, *82*(3), 344–365.

Cuetos, F., Bonin, P., Alameda, J. R., & Caramazza, A. (2010). The specific-word frequency effect in speech production: evidence from Spanish and French. *Quarterly Journal of Experimental Psychology (2006)*, *63*(4), 750–771.

Cureton, E. E. (1968). Rank-biserial correlation when ties are present. *Educational and Psychological Measurement*, *28*(1), 77–79.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 345–361.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon friends or foes? In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9–38). Berlin, Germany: Mouten de Gruyter.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991). *Psychological Review*, *98*(4), 604–614.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An anlysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*, 611–629.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838.

Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture-word interference: Evidence for response exclusion. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *36*(4), 878–891.

Dhooge, E., & Hartsuiker, R. J. (2011). How do speakers resist distraction?: Evidence from a taboo picture-word interference task. *Psychological Science*, *22*(7), 855–859.

Dietz, A., McKelvey, M., & Beukelman, D. R. (2006). Visual scene display: New AAC interface for persons with aphasia. *Perspectives on Augmentative and Alternative Communication*, *15*(1), 13–17.

Dietz, A., Thiessen, A., Griffith, J., & Peterson, A. (2013). The renegotiation of social roles in chronic aphasia: Finding a voice through AAC. *Aphasiology*, *27*(3), 309–325.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 515–523.

Fenson, L., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). *Variability in early communicative development*. Chicago, Ill: Society for Research in Child Development.

Feyereisen, P., Van der Borght, F., & Seron, X. (1988). The operativity effect in naming: a re-analysis. *Neuropsychologia*, *26*(3), 401–415.

Fink, R. B., Brecher, A., Schwartz, M. F., & Robey, R. R. (2002). A computer-implemented protocol for treatment of naming disorders: Evaluation of clinician-guided and partially self-guided instruction. *Aphasiology*, *16*(10-11), 1061–1086.

Finocchiaro, C., & Navarrete, E. (2013). About the locus of the distractor frequency effect: Evidence from the production of clitic pronouns. *Journal of Cognitive Psychology*, *25*, 861–872.

Fox, L., & Fried-Oken, M. (1996). AAC aphasiology: partnership for future research. *Augmentative and Alternative Communication*, *12*(4), 257–271.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language*, *43*(2), 182–216.

Freud, S. (1953). *On aphasia, a critical study*. London: Imago.

Fried-Oken, M., & Granlund, M. (2012). AAC and ICF: A good fit to emphasize outcomes. *Augmentative and Alternative Communication (Baltimore, Md. : 1985)*, *28*(1), 1–2.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Languages*, *47*, 27–

52.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496.

Garrett, K. L., Beukelman, D. R., & Low-Morrow, D. (1989). A comprehensive augmentative communication system for an adult with Broca's aphasia. *Augmentative and Alternative Communication*, *5*(1), 55–61.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation.* (pp. 133–177). New York: Academic Prerss.

Garrett, M. F. (1976). Syntactic process in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 231–255). Amsterdam: North Holland.

Garson, D. (2013). Fundamentals of hierarchical linear and multilevel modeling. In D. Garson (Ed.), *Hierarchical linear modeling: guide and applications* (pp. 3–26). Thousand Oaks, CA: Sage.

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640–654.

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, 219–260.

Goodglass, H., & Wingfield, A. (1997). *Anomia: neuroanatomical and cognitive correlates*. San Diego: Academic Press.

Goral, M., Levy, E., Swann-Sternberg, T., & Obler, L. (2010). Frequency and word-length factors and lexical retrieval in sentence production in aphasia. *Procedia - Social and Behavioral Sciences*, *6*, 107–108.

Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language*, *82*(2), 113–145.

Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the behavioral sciences*. Belmont, CA: Wadsworth Cengage Learning.

Grilli, L., & Rampichini, C. (2006). A review of random effects modelling using gllamm in Stata. Retrieved February 1, 2015, from http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/reviewgllamm.pdf

Hardin, J. W., & Hilbe, J. M. (2007). *Generalized linear models and extensions* (2nd ed.). College Station, TX: Stata Press.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science: Incorporating the*

Journal Cognition and Brain Theory: A Multidisciplinary Journal, *8*(3), 191–219.

Hashimoto, N., & Thompson, C. K. (2010). The use of the picture-word interference paradigm to examine naming abilities in aphasic individuals. *Aphasiology*, *24*(5), 580–611.

Hill, K. (2001). The development of a model for automated performance measurement and the establishment of performance indices for augmented communicators under two sampling conditions. *Dissertation Abstracts International*, *62*(5), 2293.

Hill, K. (2010). Advances in augmentative and alternative communication as quality of life technology. *Physical Medicine and Rehabilitation Clinics of North America*, *21*, 43–58.

Hill, K., & Scherer, M. (2008). Matching persons and technology: data-driven AAC assessments. Los Angeles, CA: CSUN: The 23rd Annual International Conference on Technology and Persons with Disabilities.

Hinkle, D. E., Wiersma, W. G., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin Company.

Hofmann, D. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, *23*(6), 723–744.

Hutson, J., Damian, M. F., & Spalek, K. (2013). Distractor frequency effects in picture-word interference tasks with vocal and manual responses. *Language and Cognitive Processes, 28*, 615–632.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.

Jacobs, B., R., D., Ogletree, B. T., & Pierce, K. (2004). Augmentative and Alternative Communication (AAC) for adults with severe aphasia: where we stand and how we ca go further. *Disabilities and Rehabilitation*, *26*, 1231–1240.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition: International Journal of Cognitive Science*, *14*(2), 185–209.

Kent, R. D., & Read, C. (1992). *The acoustic analysis of speech*. San Diego, Calif.: Singular Pub. Group.

Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research*, *46*, 608–622.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*(4), 463–492.

Knobel, M., Finkbeiner, M., & Caramazza, A. (2008). The many places of frequency: evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology*, *25*(2), 256–286.

Koul, R., & Harding, R. (1998). Identification and production of graphic symbols by individuals with aphasia: efficacy of a software application. *Augmentative and Alternative Communication*, *14*(1), 11–24.

Kraat, A. W. (1990). Augmentative and alternative communication: Does it have a future in aphasia rehabilitation? *Aphasiology*, *4*(32), 321–338.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing Statistical Methods*. London: Sage Publications.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods (Online)*, *44*(4), 978–990.

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition*, *16*(5), 400–410.

Laine, M., & Martin, N. (2006). *Anomia: Theoretical and Clinical Aspects*. Hove, UK: Psychology Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lecours, A. R., & Lhermitte, F. (1969). Phonemic paraphasias: Linguistic structures and tentative hypotheses. *Cortex*, *5*, 193–228.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1-3), 1–22.

Levelt, W. J. M., Roelofs, A. P. A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–37.

Levelt, W. J. M., Schriefers, H., Vorberg, D., & Meyer, A. S. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*(1), 122–142.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.

MacKay, D. G., & Burke, D. M. (1990). Cognition and aging: A theory of new learning and the use of old connections. In T. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 281–300). Amsterdam: North-Holland.

Mahon, B. Z., Caramazza, A., Peterson, R., & Vargas, K. A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the Picture-Word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503–535.

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician*, *36*(3), 158–160.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS One*, *7*(8), e43230.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, *47*(4), 609–660.

Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables Influencing the Occurrence of Naming Errors: Implications for Models of Lexical Retrieval. *Journal of Memory and Language*, *28*(4), 462–485.

Maylor, E. A. (1990). Age, blocking and the tip of the tongue state. *British Journal of Psychology*, *81*, 123–134.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. Chronic aphasia. *Psychological Review*, *88*(5), 375–407.

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). *Generalized, linear, and mixed models* (2nd ed.). McCulloch: John Wiley & Sons, Inc.

McKelvey, M., Hux, K., Beukelman, D. R., & Dietz, A. (2008). Information representation in AAC systems for people with chronic aphasia. In *American Speech-Language-Hearing Association Convention*. Chicago, IL.

McNeil, M. R. (1982). The nature of aphasia in adults. In N. J. Lass, L. V. McReynolds, J. L. Northern, & D. E. Yoder (Eds.), *Speech, language, and hearing: Vol. III. Pathologies of speech and language* (pp. 692–740). Philadelphia, PA: W.B. Saunders.

McNeil, M. R., & Kimelman, M. D. Z. (2001). Darley and the nature of aphasia: The defining and classifying controversies. *Aphasiology*, *15*(3), 221–229.

McNeil, M. R., & Pratt, S. R. (2001). A standard definition of aphasia: toward a general theory of aphasia. *Aphasiology, 15*(10/11), 901–911.

McNeil, M. R., Pratt, S. R., Szuminsky, N., Sung, J. E., Fossett, T. R. D., Fassbinder, W., & Lim, K. Y. (2015). Reliability and validity of the Computerized Revised Token Test: Comparison of reading and listening versions in persons with and without aphasia. *Journal of Speech, Language and Hearing Research (Online)*, *58*(2), 311–324.

Mein, R., & O'Connor, N. (1960). A study of the oral vocabularies of severely subnormal patients. *Journal of Mental Deficiency Research*, *4*, 130–143.

Middleton, E. L., & Schwartz, M. F. (2010). Density pervades: An analysis of phonological neighbourhood density effects in aphasic speakers with different types of naming impairment. *Cogn Neuropsychol*, *27*(5), 401–427.

Miller, D., & Ellis, A. (1987). Speech and writing errors in "neologistic jargonaphasia": A lexical activation hypothesis. In M. Colthear, G. Sartori, & K. Job (Eds.), *The cognitive neuropsychology of language*. Hillsdale, NJ: Erlbaum.

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General*, *132*(2), 228–252.

Morton, J. (1969). The effects of context on the visual duratio threshold for words. *British Journal of Psychology*, *55*, 165–180.

Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69–71.

Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology*, *59*(10), 1681–1690.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models*. Chicago, IL: Irwin.

Newman, R. S., & German, D. J. (2005). Life span effects of lexical factors on oral naming. *Language & Speech*, *48*, 123–156.

Nicholas, M., Sinotte, M. P., & Helm-Estabrooks, N. (2011). C-Speak aphasia alternative communication program for people with severe aphasia: importance of executive functioning and semantic knowledge. *Neuropsychological Rehabilitation*, *21*(3), 322–366.

Nickels, L. (2002). Therapy for naming disorders. Revisiting, revising, and reviewing. *Aphasiology*, *16*, 935–979.

Nickels, L., & Howard, D. (1994). A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology*, *11*, 289–320.

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: comprehension, monitoring and lexicality. *Cortex, 31*(2), 209–237.

Noh, M., & Lee, Y. (2007). REML estimation for binary data in GLMMs. *Journal of Multivariate Analysis*, *98*(5), 896–915.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*, 273–281.

Plaut, D. C. (1996). Relearning after damage in connectionist networks: toward a theory of rehabilitation. *Brain and Language*, *52*(1), 25–82.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377–500.

Porch, B. E. (1981). *The Porch Index of Communicative Ability*. Palo Alto, CA: Consulting Psychologists Press.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice*. Upper Saddle River, N.J: Pearson/Prentice Hall.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460–499.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks: Sage Publications.

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X., Martinez, A., Bloom, H., & Hill, C. (2011). Optimal Design Plus Empirical Evidence.

Raven, J. C. (1962). Advanced progressive matrices, set II. London, UK: H.K. Lewis.

Raymer, A. (2005). Naming and word-retrieval problems. In L. L. LaPointe (Ed.), *Aphasia and related language disorders* (3rd ed.). New York: Thieme.

Roach, A., Schwartz, M. F., Martin, N., & Grewal, R. S. (1996). The Philadelphia Naming Test: Scoring and Rationale. *Clinical Aphasiology, 24*, 121–133.

Roelofs, A. (2003). Goal-referenced selection of verbal action : Modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–124.

Rogers, M. A., Redmond, J. J., & Alarcon, N. B. (1999). Parameters of semantic and phonologic activation in speakers with aphasia with and without apraxia of speech. *Aphasiology*, *13*(9-11), 871–886.

Rose, M., Douglas, J., & Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonological naming impairment. *Aphasiology, 16,* 1001-1030.

Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development*, *48*(2), 643–647.

Sailor, K., Brooks, P. J., Bruening, P. R., Seiger-Gardner, L., & Guterman, M. (2009). Exploring the time course of semantic interference and associative priming in the picture-word interference task. *Quarterly Journal of Experimental Psychology (2006)*, *62*(4), 789–801.

Schacter, D. L. (1999). The seven sins of memory. Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*, 182–203.

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, *12*(2), 347–367.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime user's guide. Pittsburgh: Psychology Software Tools Inc.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time-course of lexical access in production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.

Schwab, J. A. (2002). Multinomial logistic regression: Basic relationships and complete problems. Retrieved January 5, 2015, from http://www.unt.edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug2011.pdf

Schwartz, M., Dell, G., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming☆. *Journal of Memory and Language*, *54*(2), 228–264.

Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., & Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology*, *21*, 159–186.

Shatzmana, K. B., & Schiller, N. O. (2004). The word frequency effect in picture naming: Contrasting two hypotheses using homonym pictures. *Brain and Language*, *90*, 160–169.

Shelton, J. R., Weinrich, M., McCall, D., & Cox, D. M. (1996). Differentiating globally aphasic patients: Data from in-depth language assessments and production training using C-VIC. *Aphasiology*, *10*, 319–342.

Shewan, C. M., & Kertesz, A (1984). Effects of speech language treatment in recovery from aphasia. *Brain and Language*, *23*, 272–299.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Silkes, J. P., McNeil, M. R., & Drton, M. (2004). Simulation of aphasic naming performance in non-brain-damaged adults. *Journal of Speech, Language, and Hearing Research [H.W. Wilson - SSA]*, *47*(3), 610–623.

Simmons-Mackie, N., & Kagan, A. (2007). Application of the ICF in aphasia. *Seminars in Speech and Language*, *28*(4), 244–253.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage Publications.

Spybrook, J., Raudenbush, S. W., Liu, X., & Congdon, R. (2011). Optimal Design for longitudinal and multilevel research: Documentation for the Optimal Design software version 3.0. Retrieved August 5, 2014, from www.wtgrantfoundation.org

Starkweather, J., & Moske, A. K. (2011). Multinomial logistic regression. Retrieved from http://it.unt.edu/benchmarks/issues/2011/08/rss-matters

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 896–918.

Steele, R. D., Kleczewska, M. K., Carlson, G. S., & Weinrich, M. (1992). Computers in the rehabilitation of chronic, severe aphasia: C-VIC 2.0. *Aphasiology*, *6*, 185–194.

Steele, R. D., Weinrich, M., Wertz, R. T., Kleczewska, M. K., & Carlson, G. S. (1989). Computer-based visual communication in aphasia. *Neuropsychologia*, *27*(4), 409–426.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (pp. 143–186). Hillsdale, NJ: Erlbaum.

Stuart, S., Beukelman, D. R., & King, J. (1997). Vocabulary use during extended conversations by two cohorts of older adults. *Augmentative and Alternative Communication*, *13*(1), 40–47.

Sutton, M. (2012). Apps to aid aphasia. *ASHA Leader*, *17*(7), 32.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Académie Canadienne de Psychiatrie de L'enfant et de L'adolescent*, *19*(3), 227–229.

Terhorst, L. (2008). *A comparison of estimation methods when an interaction is omitted from a multilevel model*. University of Pittsburgh.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's wordbook of 30,000 words*. New York: Teacher's College, Columbia University.

Tillman, T. W., & Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6. technical report no. SAM-TR-66-55. Brooks Air Force Base, TX: USAF School of Aerospace Medicine.

Trembath, D., Balandin, S., & Togher, L. (2007). Vocabulary selection for Australian children who use augmentative and alternative communication. *Journal of Intellectual & Developmental Disability*, *32*(4), 291–301.

van de Sandt-Koenderman, W. M. E. (2004). High-tech AAC and aphasia: Widening horizons? *Aphasiology*, *18*(3), 245–263.

van de Sandt-Koenderman, W. M. E., Wiegers, J., & Hardy, P. (2005). A computerised communication aid for persons with aphasia. *Disability and Rehabilitation*, *27*(9), 529–533.

van de Sandt-Koenderman, W. M. E., Wiegers, J., Wielaert, S. M., Duivenvoorden, H. J., & Ribbers, G. M. (2007). High-tech AAC and severe aphasia: Candidacy for TouchSpeak (TS). *Aphasiology*, *21*(5), 459–474.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735–747.

Wallace, T., & Bradshaw, A. (2011). Technologies and strategies for people with communication problems following brain injury or stroke. *NeuroRehabilitation*, *28*(3), 199–209.

Weinrich, M., McCall, D., & Weber, C. (1995). Thematic role assignment in two severely aphasic patients: Associations and dissociations. *Brain and Language*, *48*(2), 221–237.

Weinrich, M., McCall, D., Weber, C., Shoosmith, L., Thomas, K., & Katzenberger, K. (1993). Training on alternative communication systems can improve aphasic patients' verbal output. *Annals of Neurology*, *34*, 280.

Weinrich, M., Shelton, J. R., McCall, D., & Cox, D. M. (1997). Generalization from single sentence to multisentence production in severely aphasic patients. *Brain and Language*, *58*(2), 327–352.

Weinrich, M., Steele, R. D., Kleczewska, M., Carlson, S. G., Baker, E., & Wertz, R. T. (1989). Representations of "verbs" in a computerized visual communication system. *Aphasiology*, *3*, 501–512.

Wilshire, C. E. (2002). Where do aphasic phonological errors come from? Evidence from phoneme movement errors in picture naming. *Aphasiology*, *16*(1-2), 169–197.

Wilshire, C. E., Keall, L. M., Stuart, E. J., & O'Donnell, D. J. (2007). Exploring the dynamics of aphasic word production using the picture-word interference task: A case study. *Neuropsychologia*, *45*(5), 939–953.

Wilson, R. H., Zizz, C. A., Shanks, J. E., & Causey, G. D. (1990). Normative data in quiet, broadband noise, and competing message for Northwestern University Auditory Test No. 6 by a female speaker. *Journal of Speech Hearing Disorders*, *55*, 771–778.

World Health Organization (WHO). *International Classification of Functioning, Disability and Health (ICF)*. Geneva, Switzerland: World Health Organization; 2001

Worrall, L., Rose, T., Howe, T., & McKenna, K. (2007). Developing an evidence-base for accessibility for people with aphasia. *Aphasiology*, *21*(1), 124–136.