Bowling Green State University

ScholarWorks@BGSU

Word Error and Word Frequency Correlation in Broca's Aphasia

Breawna Heal
bheal@bgsu.edu

Follow this and additional works at: https://scholarworks.bgsu.edu/honorsprojects

Part of the Speech Pathology and Audiology Commons

Word Error and Word Frequency Correlation in Broca's Aphasia

Breawna Heal

Bowling Green State University

Introduction

Aphasia is a language disorder that is caused to the brain (Shames, Wiig, Secord, 1998) This damage may be caused by a stroke, traumatic brain injury, lesions to the brain, or degenerative neurological diseases(Shames, Wiig, Secord, 1998). Depending on the location of the brain damage, a person with aphasia will present with different communication impairments. People with fluent aphasia produce speech that may feature many grammatical errors or incorrectly produced words but their speech retains a typical speaking rhythm. People with fluent aphasia's speech is much slower and more halting, and there tends to be frequent and lengthy pauses between words. People with mixed aphasia have moderate production and comprehension difficulties. Global aphasia is characterized by severe production and comprehension difficulties.

The participants in this research project are people who have Broca's aphasia, a type of non-fluent aphasia. Broca's aphasia is caused by lesions to Broca's area that is in the left frontal lobe. Broca's aphasia is associated with several symptoms; word retrieval difficulties, frequent pauses in speech, and errors in fluidity (Shames, Wiig, Secord, 1998). People with Broca's aphasia frequently exhibit paraphasias in their speech. Paraphasias are speech errors that are can be characterized by phonemic substitutions for instance saying "kroom" instead of "boom". Another type of paraphasia is a semantic substitution, this is when the patient replaces a full word including the meaning, for instance saying "flower" for "grass" (Brookshire, 2015).

The goal of treatment is to lessen the patient's day to day communication struggles, these goals will depend on the patient and the severity of the aphasia (Brookshire, 2015).We utilized speech samples from a Speech-Language Pathology directed narrative therapy session in this research project. We acquired these narrative speech samples from an online database called

AphasiaBank. We analyzed these narrative speech samples and discovered the frequency in language each word has by using the SUBTlexus Corpus.

Through the SUBTlexus Corpus we were able to look up each word's word frequency. Word frequency is something that people might not consciously thinking about but word frequency has effected all of us. The difference between reading a doctoral thesis or a tabloid magazine is word frequency. The higher word frequency the more frequent the word is in language and subsequently it is easier for you to read. A low frequency word comes up less frequently in language and is proportionately harder to read (Oxford University Press, 2007). We will be using the FREQcount feature of the SUBTlexus corpus. The SUBTlexus is an advanced version of Kucera and Francis. Kucera and Francis was a corpus made up of Anglo-Saxon writings that they based word frequency on. SUBTlexus is far more technologically advanced and computerized than the previous Kucera and Francis (Hudson, 2003). The SUBTlexus uses manuscripts from all crevices of the English language.

Word errors and word frequency in people with Broca's aphasia in a narrative is important to investigate. This is a critical gap in research because there is no research out there that addresses this possible correlation. Also there is little research on people with Broca's Aphasia speaking for longer periods of time, like they do with the narratives. Knowing the extent of this correlation will not only fill this research gap but will also aid Speech-Language Pathologists with therapy.

Aim

The aim of this study was to determine if there is a relationship between a word's frequency and the probability that people with Broca's aphasia will produce the word erroneously during a narrative production task.

Methodology

Data

For this study we collected our narrative samples from an online database designed to support teaching and research related to aphasia called "AphasiaBank". AphasiaBank stores videos of a large number of people with different kinds of aphasia engaged in a variety of language/communication activities such as standardized language testing and conversation (MacWhinney, Fromm, Holland, Forbes, & Wright, 2014). Every person with aphasia who contributed to Aphasiabank was assessed by trained clinicians and information about each person's aphasia symptoms and type is available on the website.

Most of the video data on Aphasiabank has been transcribed; transcripts are also available on the website. The transcripts are verbatim in that they detail every verbal item produced by the participants including speech errors and vocalizations. Information about the targets and errors are also provided (for example, if the person with aphasia attempted to produce the word 'many' but produced an error such as 'dany', both forms are listed in the transcript).

One common task that people with aphasia are asked to complete as part of assessment and treatment activities is a narrative task, in which the person is asked to tell a story. Many of the people with aphasia featured on Aphasiabank were asked to tell the story of Cinderella. We focused on this data.

 We reviewed the videos and transcripts associated with the Cinderella narratives as produced by people with Broca's aphasia. Subsequently, we picked five people with Broca's aphasia from Aphasiabank. To be included in our project, the participants had to produce at least twenty paraphasias during the course of the narrative production. We need this minimum amount

of paraphasias because we need to be able to have enough data in each sample to strengthen our statistical analysis.

We retrieved every word produced by the people with Broca's aphasia during the narrative task from the transcripts. We then classified every word as "incorrect" (the word is a paraphasia or neologism) or "correct". We then looked up the frequency of each word, correct and incorrect, using the SUBTlexus Corpus.

The SUBTlexus is an online resource that will be used to generate word frequencies for the words we are analyzing.

Statistical Analysis

The independent variable in our analysis is the word frequency of every word in the narratives. Word frequency is a ratio value because it will be represented on the graph as a variety of scale values. The dependent variable is the correctness of each word in the narratives. This was represented as a nominal value because it can only be correct or incorrect, there was no in between represented in this analysis.

Given the nature of the independent and dependent variables, we conducted a simple logistical regression.

Results

An independent samples t test was conducted to compare frequencies for words that we produced correctly and words that were errors. There was a significant difference in the frequencies for the correct words (M= 20699.691, SD=345072.51); t (263)= -3.285, p=0.001. This result indicates that there is a statistically significant difference between the correct and incorrect groups.
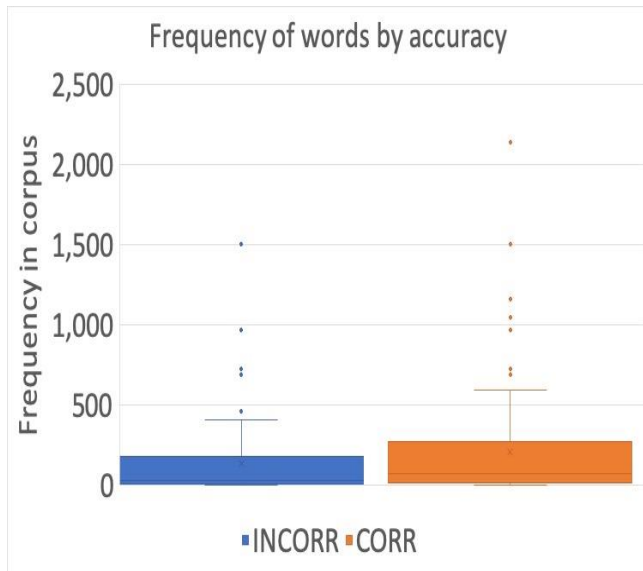
Figure 1. Box and whisker plot of frequencies associated with each condition.

Note: digits on x-axis should be multiplied by 1000

Conclusion

Given the small number of participants that we included in this study, the results we furnish should be viewed as preliminary. Despite this shortcoming, our results suggest that in narratives produced by people with Broca's aphasia, significant frequency differences exists between words produced correctly and those produced erroneously.

This finding comports with previous studies of word frequency in neurotypical and neurologically injured language users, which have tended to conclude that word frequency affects language processing.

Our finding concerning frequency differences can inform interventions for aphasia. During treatment a therapist looks back at the errors made during the session and adjusts the future therapy sessions and short-term goals accordingly. If "shoe" and "slipper" were seen as equal errors instead of an error in "shoe" being more severe than "slipper", it could slow down the

rehabilitation of the patient. Speech-Language Pathologists could also keep this in perspective during the initial screening and diagnosis for aphasia occurs. This would not only alter the initially diagnosed severity of the aphasia, but could also change the course of treatment and long-term goals of the patient.

References

Baayen, R. H. (2002). Word frequency distributions Springer Science & Business Media.

Damasio, A. R. (1992). Aphasia. New England Journal of Medicine, 326(8), 531-539.

Darley, F. L. (1982). Aphasia. Philadelphia: WB Saunders Company.

Dronkers, N., & Baldo, J. V. (2010). Language: Aphasia. Encyclopedia of neuroscience (pp. 343-348) Elsevier Ltd.

Kean, M. (1977). The linguistic interpretation of aphasic syndromes: Agrammatism in broca's aphasia, an example. Cognition, 5(1), 9-46.

Sarno, M. T. (1998). Acquired aphasia Elsevier.