Examining the Effect of Predictability on Eye Movement During Story Reading in Neurotypical Adults and People with Aphasia

by

Jacquelyn F. Stochel

Bachelor of Science, Penn State University, 2022

Submitted to the Graduate Faculty of the

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2024

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This thesis was presented

by

Jacquelyn F. Stochel

It was defended on

March 8, 2024

and approved by

Michael Walsh Dickey, Ph.D., Professor, Associate Dean for Graduate Studies in the School of Health and Rehabilitation Sciences, Departments of Communication Science and Disorders and Psychology

Tessa Warren, Ph.D., Professor, Senior Scientist in the Learning Research and Development Center, Departments of Psychology, Linguistics, and Communication Science and Disorders

Thesis Advisor: Sarah E. Wallace, Ph.D. CCC-SLP, ASHA-Fellow, Professor, Director of Master's Degree Program in Speech-Language Pathology, Department of Communication Science and Disorders

Copyright © by Jacquelyn F. Stochel

2024

Examining the Effect of Predictability on Eye Movement During Story Reading in Neurotypical Adults and People with Aphasia

Jacquelyn F. Stochel, BS

University of Pittsburgh, 2024

There is a current need to improve and expand options for reading intervention that support people with post-stroke aphasia. The present study used pre-existing eye tracking data, which tracked the reading behaviors of people with aphasia and neurotypical adults during story reading. The objectives of this analysis were to compare the eye movement patterns of each participant group, and to retrospectively describe an effect of a word predictability measure (surprisal) on eye movement across groups during story reading. Analysis revealed that people with aphasia had longer gaze durations and total fixation durations compared to neurotypical adults, which is consistent with previous findings of group effects in sentence-level stimuli. The positive effect of surprisal on probability of first pass fixation also aligns with previous studies of predictability and eye movement, but the negative effect of surprisal on probability of first pass regression, gaze duration, and total fixation duration contradicts much of the previous evidence on predictability effects. Additionally, no interactions of group and surprisal were found in this dataset, suggesting that neurotypical adults and people with aphasia respond similarly to the predictability of words during reading. While these findings could be attributable to the design of the original experiment or the current analysis approach, it is also possible that these findings are unlocking reading behaviors not previously shown in predictability literature. The current findings have the potential to inform future interventions for acquired reading impairment in post-stroke aphasia.

Table of Contents

List of Tables

List of Figures

Preface

I would like to express my gratitude to my thesis committee for their guidance, feedback, and their individual contributions to this work. Their unique expertise and ongoing encouragement was instrumental to the success of this project. I must also extend my gratitude to Dr. Amanda Hampton Wray for agreeing to serve as the moderator for my thesis defense.

I would also like to thank Alex Swiderski (PhD Student) and Candance van der Stelt (PhD Student) for supporting the analysis presented in this paper, and for offering feedback and mentorship on the project. I extend my appreciation to a dedicated team of undergraduate and graduate research assistants, who aided me in the manual organization, recording, and transfer of raw research data. These students are Sophia Coolen, Rachel Donahue, Shannon Dunphy, Emily Fowler, Sammi Kimmbis, Gracie Long, Meredith Marmolejos, Manasvita Rajesh, and Mikayla Hetrick.

I am grateful to Dr. Kelly Knollman-Porter (Principal Investigator of Neurogenic Language and Cognition Lab at Miami University of Ohio) and Dr. Karen Hux (Principal Investigator at Quality Living, Inc) for their collecting the research data presented in this study, allowing me to conduct a retrospective analysis of their dataset, and being available to answer questions throughout the project's development.

I want to thank Dr. Lin Chen and Dr. Yashan Wang at the University of Pittsburgh for meeting with me in the early stages of this project, and providing me with consultation that guided the analysis presented in this paper. I would also like to thank the clinical education team for MA/MS program in speech-language pathology, particularly Dr. Erin Lundblom and Ashley

Kreiger, for coordinating a clinical placement schedule that was compatible with my thesis timeline.

I must thank my extensive support network, which includes friends and family, roommates, professors, clinical instructors, and former and current research teams. I thank them for their ongoing encouragement and confidence in my abilities as I worked on this project. Finally, I must acknowledge the clients and speech-language pathologists that inspire me to engage in research that improves our field's impact on their world.

1.0 Introduction

Aphasia is a neurogenic language disorder that affects various language domains, including reading (Purdy et al., 2019). In a recent large and representative study of people with post-stroke aphasia (n=99), an estimated 68% of participants met the criteria for alexia, an acquired reading impairment (Brookshire et al., 2014). While alexia is a common feature of aphasia, the nature and severity of impairment can be highly variable (Knollman-Porter et al., 2015). People with aphasia may demonstrate impairment in different reading tasks (i.e., silent reading versus oral reading; see Webb & Love, 1983 for review) and different reading subprocesses (e.g., decoding, comprehension, fluency; see Knollman-Porter et al., 2015 for review). Furthermore, people with aphasia may have difficulty processing written language at any or all linguistic levels (i.e., subword, word, phrase, sentence, narrative). People with aphasia report that alexia can pose significant barriers to life participation (Parr, 1995), and have a profound impact on well-being (Rose et al., 2011), and quality of life (Parr, 2007).

Reading comprehension deficits in aphasia may be explained by different breakdowns in the language system. For example, some people with aphasia will have difficulty with sound-to-letter correspondence, which could impair or slow the decoding of written words (Brookshire et al., 2014; Purdy et al., 2019). Others may have difficulty accessing semantic knowledge during reading, resulting in consequences such as impaired reading comprehension (Purdy et al., 2019). Understanding the precise challenges and contexts of impairments for people with aphasia is important for designing assessment and treatment approaches which serve the variety of clients a speech-language pathologist may encounter.

While there are several treatment programs have been created to restore reading abilities in people living with aphasia (i.e., oral reading, strategy-based, cognitive, and hierarchical interventions) there remains high variability in reading quality outcomes following treatment for reading comprehension (Purdy et al., 2019). Purdy and colleagues also stressed that existing reading interventions must be carefully selected to match the profile of the individual receiving treatment. For instance, an individual with impaired letter-sound correspondence, resulting in sublexical breakdowns in reading, may benefit from segmenting sounds in an oral reading intervention. In contrast, a person with spared decoding skills but impaired comprehension in higher level text may not benefit from the same oral reading intervention.

Given the variable outcomes of existing treatment and the limited number of interventions tailored to specific individual profiles, there is rationale to continue the research of reading processes in neurotypical adults and people with aphasia. Further exploration of impaired versus successful reading can be used to inform designs of patient-centered treatment approaches. A number of research techniques have helped to compare reading abilities and behaviors in people with aphasia relative to neurotypical adults. These methods include comprehension and reading time measures (e.g., Knollman-Porter et al., 2022a), self-paced reading protocols (e.g., DeDe, 2013), and self-report metrics (e.g., Hux et al., 2021). However, one of the most descriptive and ecologically valid approaches to studying reading behavior is the use of eye tracking methodologies (Sharma et al., 2021).

1.1 Eye Tracking to Examine Reading

Eye tracking is a method in which eye movements across a visual field can be recorded, measured, and interpreted (Rayner, 2009). Measurements often characterize two components of eye movement: saccades (i.e., eye movement events) and fixations (i.e., intervals of focused, stable gaze). Within these measures, one may capture more specific aspects of the saccades (i.e., direction, number) and/or fixations (i.e., probability, count, duration) to operationally describe reading behaviors given the manipulation of variables, such as etiology (e.g., neurotypical adults versus people with aphasia), reading material (e.g., sentences versus paragraphs), and word and text properties (e.g., word length). The experimental manipulation of these variables may lead to significant changes in the rate, count, and duration of fixations during reading.

1.1.1 Overview of Relevant Factors Affecting Eye Movement

Several factors may affect eye movement during reading, including length of written material. Reading materials vary in length from single words to multiple page books, and the materials have implications on the reading behavior observed through eye tracking. The subsequent paragraphs will focus on the different sentence- and paragraph-level stimuli.

Compared to sentence stimuli, paragraphs include a greater number of words, lines, or anaphoric references (i.e., a pronoun reference to previously introduced subject), which may lead to more regressive saccades if information is no longer in the reader's working memory and perceptual span (Cook & Wei, 2010). These authors also highlight that paragraph level text requires integration and validation of information across sentences, a behavior unlike what is observed in sentence processing studies.

Visually, sentences and passage will vary in terms of their alignment of line and sentences (Kuperman et al., 2010). In other words, a longer text may start and end a sentence on different lines, and this is less likely to be true of single-sentence stimuli used in reading experiments. This becomes important in the context of a word position in text effect, or a noted increase in fixation times that is observed at the end of a line or in the final line of passage. While there are also distinctive wrap up effects, which involve longer fixations for words at the end of linguistic sentence structures (see Warren, 2009 for review), Kuperman et al. (2010) argue that "word position in text" effects are more robust, and can be found in passage-level like the ones used in this study. However, these effects are dissociable in passage-level texts like the current study's, where both the effects of visual line endings and sentence endings are likely to impact fixation durations and reading times.

In both sentences and paragraphs, the lexical and syntactic properties of text will also influence eye movement. Examples of variables that have well-evidenced effects on reading include word length (i.e., the number of characters per word) and word frequency (i.e., the number of times a word occurs in a corpus of text). Word length is frequently associated with a probability of fixating words in text, where longer words have a greater probability of being fixated and short words have a greater probability of being skipped (Brysbaert et al, 2005; Rayner, 1998). Word frequency is typically found to have robust effects on early measures of fixation duration, which represents the amount of time a reader spends on a word during an initial first pass through text (Rayner, 1998). Word frequency and word length effects on eye movements are not only present but exaggerated in people with aphasia compared to neurotypical adults (DeDe, 2017; Huck et al., 2017; Knilans & DeDe, 2015).

In addition to word length and word frequency effects, there is also strong and consistent evidence of a predictability effect in reading (Rayner, 2009; Staub, 2015). Predictability is the expectedness of a word given its context, where "context" is inclusive of preceding syntax and semantic relations in the text, and the knowledge a reader brings to the text (Balota et al., 1985). Possible explanations of predictability effects include "anticipatory" processes (i.e., context exerted to hypothesize upcoming text) and "bottom-up" views of reading (i.e., exertion of context once a word is recognized) (see Federmeier, 2007 for review). A number of eye tracking behaviors show sensitivity to predictability, as demonstrated in both studies of neurotypical adults and people with aphasia.

1.2 Manipulation of Word Predictability in Written Text

To successfully study predictability, a researcher must take an objective approach to defining and measuring this variable. Predictability effects have not only been operationalized for the studies utilizing eye tracking (Balota et al. 1985; Boston et al., 2008; Binder et al., 1999; Calvo & Meseguer, 2002; Demberg & Keller, 2008; Ehrlich & Rayner, 1981; Huck et al., 2017; Kliegl et al., 2004; Rayner et al., 2004; Rayner et al., 2009; Rayner & Well, 1996; Schustack et al., 1987; Vitu, 1991), but also self-paced reading (e.g., Dickey et al., 2014), event-related potentials (e.g., Federmeier, 2007), and functional magnetic resonance imaging (e.g., Brennan et al., 2016).

1.2.1 Measurement of Predictability

Predictability is typically operationalized using a cloze procedure, where a cohort of participants must complete a phrase, sentence, or passage with a missing word (e.g., completing the phrase, "HE SEARCHED THE_______", with "NEIGHBOORHOOD") (Taylor, 1953). The proportion of accurate responses will inform the measure of predictability. While cloze predictability has been used to develop and validate stimuli in many eye tracking studies, this method is ultimately limited because a large sample of responses is required to ensure estimates of predictability are generalizable and reproducible. This is ultimately expensive, time-consuming, and difficult to achieve (Bianchi et al., 2020).

An alternative to cloze predictability is computer estimation of predictability. Artificial intelligence tools, such as large language models (LLMs), provide researchers with an opportunity to capture properties of text, including the probability of a word, over other training sample words, given context. This statistical probability of a word can help characterize predictability for experimental purposes. One of the most common estimates of predictability is surprisal, a measure which objectively represents the unexpectedness of a word in its written context (see Lowder et al., 2018 for review). Surprisal is a negative log-function of word probability, which can be derived from LLMs. Surprisal has a negative relationship with predictability. If a word is 100% predictable in in its context, it conveys no new information and has a surprisal rating of 0. Like cloze predictability, surprisal has been used to represent word predictability in eye tracking experiments (Boston et al., 2008; Demberg & Keller, 2008), but this work is fairly new to our field. The impact of both cloze predictability and surprisal on eye movement during reading will be expanded upon in the next section of this paper.

1.2.2 Predictability and Reading in Neurotypical Adults

Decades of research, featuring various reading materials (sentence, paragraph) and the two major approaches to measuring predictability (cloze predictability, surprisal), identify that predictability has robust effects on several measurements of eye movement in neurotypical adults. For instance, neurotypical adults exhibit changes in probability of first pass fixation (PF) in response to the experimental manipulation of predictability. PF captures the likelihood of fixating a word during the first pass through a text (see Rayner, 2009 for review). Neurotypical adults have reduced PF for predictable words compared to unpredictable words when reading sentences (Balota et al. 1985; Binder et al., 1999; Rayner et al., 2011, Rayner & Well, 1996) and paragraphs (Ehrlich & Rayner, 1981; Schustack et al., 1987). This means that neurotypical adults will frequently skip predictable words in the first opportunity they have to read them.

Predictability is also linked to the probability of first pass regression (PR), also called regression out (e.g., Kliegl et al., 2004) or regression probability (e.g., Boston et al., 2008). This measure represents the likelihood of shifting backward in text after fixating a word on first pass. In general, regression is interpreted as a process for resolving ambiguity during reading (Clifton Jr. et al., 2003; Frazier & Rayner, 1982). PR is higher for unpredictable words than predictable words when fixated in sentence-level contexts (Kliegl et al., 2004). Complimentary findings in the surprisal literature suggest that PR is also correlated to increases in word surprisal (or decreases in word predictability) (Boston et al., 2008).

Evidence from neurotypical adults has also identified predictability effects on reading time measures. For instance, gaze duration (GD), which measures the length of time a word is gazed upon before a reader shifts to a different word, is sensitive to word predictability. GD is an early measure of reading behavior and is seen as marker of text integration during reading (Inhoff, 1984).

GD is highly sensitive to cloze predictability, where readers spend less initial time on predictable words compared to unpredictable words in sentence-level text (Balota et al. 1985; Binder et al., 1999; Rayner, 2004; Rayner et al., 2011; Rayner & Well, 1996; Vitu, 1991) and paragraph-level text (Ehrlich & Rayner, 1981; Schustack et al., 1987). These patterns have also been replicated in studies using surprisal estimates, where increases in word surprisal (decreases in word predictability) are related to significant increases in GD (Boston et al., 2008; Demberg & Keller, 2008).

One other measure related to predictability is total fixation duration (TFD). TFD represents the total amount of time a reader fixates a given word in the text, including both time from an initial encounter with the word and any added time from re-reading the same word. TFD is considered a global measure of reading time, and is interpreted by researchers as an overall marker of reading difficulty (Clifton Jr. et al., 2007). Experiments using cloze predictability note that TFD is higher for unpredictable words than predictable words in sentence-level reading (Balota et al. 1985; Binder et al., 1999; Calvo & Meseguer, 2002; Kliegl et al., 2004; Rayner, 2004; Rayner et al., 2011, Rayner & Well, 1996) and paragraph-level reading (Ehrlich & Rayner, 1981; Schustack et al., 1987).

1.2.3 Predictability and Aphasia

Predictability effects were studied for many years before questions were posed about these same effects in special populations, such as people with aphasia. The motivation to initiate this research came from the question of whether prediction was an impaired process or a possible strategy for language processing in aphasia. Recent work has inquired about how people with aphasia respond to proceeding context, and this work has expanded knowledge on the integrity of predictive processes in people with aphasia.

Prediction has been tested in a number of studies using a visual-world eye tracking paradigms, where participants would be primed to look at pictures following presentation of auditory or written sentences. Across these studies, people with aphasia demonstrated increased anticipatory fixations to target pictures in response to increased syntactic constraint, such as use of inflectional morphemes (Hanne et al., 2015), verb argument status (Hayes et al., 2016), and restrictive verbs (Mack et al., 2017) in the preceding sentence context. Mack and Thompson (2013) described a different but related phenomenon in which people with aphasia showed no predictive behavior in response to a preceding agent, but significant predictive responses to preceding agents following a syntactic structure training intervention.

There is additional supporting evidence of spared predictive processing when people with aphasia engaged in self-paced reading (Warren et al., 2016). This research investigated if healthy adults and people with aphasia engaged in structural prediction given a preceding "either" in written sentences (e.g., "Emily painted (either) a lovely still life or a beautiful portrait of her mother"). The authors found that the presence "either" resulted in faster reading times for predictable structures in the upcoming sentence (i.e., "or" and second disjuncts). Taken together, there is overlapping evidence of residual predictive processing in language comprehension for people with aphasia. These findings justify studies of predictability effects in text-specific eye tracking paradigms, as prediction may be a strength of readers with post-stroke aphasia.

Only one known study to date has assessed the effect of predictability on eye movement during reading in people with aphasia. Huck et al. (2017) conducted an eye tracking experiment where neurotypical adults and people with aphasia were tasked with reading sentences designed to contain either: 1) a high predictability and high frequency target word, 2) a high predictability and low frequency target word, 3) a low predictability and high frequency target word, or 4) a low predictability and low frequency target word. Their goal was to identify the effects of group (people with aphasia, neurotypical adults), predictability (high, low), and word frequency (high, low) on relevant eye tracking measures (PF, PR, GD, TFD).

Huck et al. (2017) demonstrated main effects of predictability across participants, where readers had lower PF, PR, GD and TFD for predictable words in the experimental sentences, and higher PF, PR, GD, and TFD for unpredictable words in the experimental sentences. Another unique finding of this research was an interaction effect between group and predictability, where people with aphasia had significantly longer TFD for unpredictable words in the experimental sentences compared to neurotypical adults. These results were consistent with previous findings on neurotypical sentence processing, which notes decreases in PF, PR, GD, and TFD when reading predictable words in experimental sentences (Balota et al. 1985; Binder et al., 1999; Rayner & Well, 1996; Vitu, 1991). One possible explanation proposed for this outcome was that people with aphasia benefit from context clues to scaffold the reading process, at a greater or exaggerated level compared to neurotypical adults.

1.3 Present Inquiry

Several gaps exist in understanding the role of predictability in the reading patterns of neurotypical adults and people with aphasia. Review of the literature on reading in neurotypical adults reveals only a small number of studies that explore predictability effects in paragraph-level materials (Ehrlich & Rayner, 1981; Schustack et al., 1987). As for studies of people with aphasia, there are no known studies to date which assess predictability effects in paragraph-level text. Given evidence of predictability effects in both sentence- and paragraph-level reading for neurotypical adults, there is rationale for studying predictability effects in people with aphasia when reading paragraph-level texts. Furthermore, research has evidenced an attenuation of word frequency effects in paragraph reading paradigms (Radach et al, 2008) suggesting that salient linguistic variables may not be processed the same for sentence versus paragraph contexts. This justifies a concrete investigation of predictability effects on eye movement during paragraph reading in neurotypical adults and people with aphasia.

Another critical gap in the literature is an investigation of surprisal effects on eye movement involving people with aphasia. The one known investigation of predictability effects on eye movement in people with aphasia utilized a cloze predictability measure. While there is evidence to suggest that cloze predictability and surprisal yield parallel effects on eye movement during reading in neurotypical adults, there must be a concrete investigation utilizing surprisal estimates to confirm that people with aphasia show similar sensitivity to both measures. If people with aphasia demonstrated responses to surprisal in salient eye movement, then there would be compelling evidence to utilize surprisal measures in future research.

Taken together, there remains a need to expand the literature with a comparative study of surprisal effects in narrative-level reading for people with aphasia and neurotypical adults to address gaps in the theoretical understanding of silent reading. This data would also contribute knowledge about a clinically modifiable factor, which a speech-language pathologist could leverage in the design of assessment and treatment materials to support a subset of people with aphasia experiencing reading impairments.

1.3.1 Research Questions

Given the current gaps in the surprisal and reading literature, there were three main questions the study aimed to address. These questions were:

1. What differences will be observed in eye movements (i.e., probability of first pass fixation, probability of first pass regression, gaze duration and total fixation duration) during story reading in neurotypical adults versus people with aphasia?
2. What effect will word surprisal have on eye movements (i.e., probability of first pass fixation, probability of first pass regression, gaze duration and total fixation duration) during story reading in neurotypical adults and adults with aphasia?
3. What differences will exist in surprisal effects on sensitive eye movements (i.e., probability of first pass fixation, probability of first pass regression, gaze duration and total fixation duration) during story reading in neurotypical adults versus people with aphasia?

1.3.2 Specific Aims and Hypotheses

To address the first research question, I measured and compared group (people with aphasia, neurotypical adult) effects on relevant eye movement measures (i.e., probability of first pass fixation, probability of first regression, gaze duration and total fixation duration) during story reading. I expected people with aphasia to yield increases in all eye movement measures due to recent insight on the eye movement in people with aphasia compared to neurotypical adults during sentence-level reading (Huck et al., 2017).

To address the second research problem, I measured the effects of surprisal on relevant eye movement measures (i.e., probability of first fixation, probability of first regression, gaze duration and total fixation duration) during story reading. I hypothesized that an increase in target word surprisal will lead to increases in all eye movement measures given converging evidence of predictability effects in the eye tracking literature for neurotypical adults (Balota et al. 1985; Binder et al., 1999; Boston et al., 2008; Demberg & Keller, 2008; Calvo & Meseguer, 2002; Ehrlich & Rayner, 1981; Huck et al., 2017; Kliegl et al., 2004; Rayner et al., 2004; Rayner et al., 2009; Rayner & Well, 1996; Schustack et al., 1987; Vitu, 1991) and people with aphasia (Huck et al., 2017).

To address the final research question, I examined surprisal effects on relevant eye movement measures (i.e., probability of first fixation, probability of first regression, gaze duration and total fixation duration) between groups of participants during story reading. I hypothesized that people with aphasia would demonstrate consistent or exaggerated eye tracking measures compared to neurotypical adults. This expectation is justified given evidence of these effects in sentence-level reading (Huck et al., 2017).

2.0 Methodology

The present study investigated narrative-level predictability effects by using previously collected data from a study completed at Miami University of Ohio (Knollman-Porter et al., 2022). The research aim for the previous study was to assess the effects of text-to-speech tools in people with aphasia and neurotypical adults given stories to read. To address present questions specific to predictability as factor influencing reading, I extracted data from trials where both groups were not given text-to-speech to read.

2.1 Participants

Data from nine people with aphasia and nine neurotypical adults was analyzed in the present study. All neurotypical participants were matched to a participant with aphasia, with each match being within 6 years of age and 2 years of education of each other. Participants spoke American English and had earned at least a high school diploma.

All people with aphasia and neurotypical adults passed screenings for normal or corrected-to-normal vision and hearing to ensure each participant was equipped to complete experimental tasks. For the hearing screening, participants needed to detect 1000, 2000, and 4000 Hz pure tones in at least one ear, when presented at 40 dB via headphones. For vision screening, participants needed to identify their written name among four other written names with 100% accuracy. Names were presented in ten rows of 28-point Times New Roman font, and were situated on a white background. Finally, participants needed to meet a threshold validation accuracy during eye gaze calibration procedures. Three prospective participants with aphasia did not meet this validation accuracy, resulting in exclusion from the study.

2.1.1 People with Aphasia

Nine people with aphasia (five men and three women) were included in the analysis for the present study. The ages of people with aphasia ranged from 37 to 76 years (M = 61.5, SD = 14.50), and the highest level of education for people with aphasia ranged from 12 to 19 years (M = 16.3, SD = 2.65). Aphasia onset ranged from 5 to 269 months prior to enrollment in the study (M = 134.5, SD = 100.93). All but one participant reported aphasia onset following left hemisphere stroke, and all participants reported right-hand dominance prior to onset of aphasia. Table 1 provides a summary of all demographic information for people with aphasia.

Table 1 Demographic Information for People with Aphasia

Participant	**Demographic Information**					
Code	**Age** (Years)	**Education** (Years)	**Race**	**Gender**	**Aphasia Onset** (Months Post)	**Aphasia Classification**
A	76	16	White	Female	269	Broca's
B	37	18	White	Female	79	Anomic
C	52	18	White	Male	26	Transcortical Motor
D	52	16	White	Male	162	Broca's
E	72	18	White	Male	63	Transcortical Motor
F	75	12	White	Female	117	Anomic
G	54	12	White	Male	186	Broca's
H	74	19	White	Male	294	Transcortical Motor
I	47	18	White	Female	14	Anomic

All people with aphasia completed language and reading comprehension testing prior to the experiment. Testing did not serve an inclusionary or exclusionary purpose during recruitment or analysis. The Western Aphasia Battery—Revised (WAB-R; Kertesz, 2007) was selected to capture classification of aphasia through weighted subtest scoring, and the severity of aphasia using an aphasia quotient (AQ) score. Additional testing included the Comprehension of Spoken Paragraphs subtest of the Comprehensive Aphasia Test (CAT; Swinburn et al., 2004) and the Paragraph-Factual subtest of the Reading Comprehension Battery for Aphasia–Second Edition (RCBA-2; LaPointe & Horner, 1998) to provide additional information on spoken discourse and written paragraph comprehension. Appendix C provides a summary of all standardized testing data from people with aphasia.

2.1.2 Neurotypical Adults

Nine neurotypical adults (four men and five women) with no prior history or neurological disease or injury were enrolled in the study to serve as age- and education-matched controls to the people with aphasia. The ages of neurotypical participants ranged from 37 to 81 years (M = 61.22, SD = 15.94), and highest level of education ranged from 12 to 18 years (M = 15.56, SD = 2.55) See Table 2 for a summary of neurotypical adult demographic information. All neurotypical adults were given the Montreal Cognitive Assessment (MoCA; Nasserdine et al., 2005) prior to completion of the experiment. Scores ranged from 26 to 30 out of 30 possible points, which suggest that of cognitive performance of the neurotypical adults was within normal limits.

Table 2 Demographic Information for Neurotypical Participants

Participant	Demographic Information				
Code	**Age** (*Years*)	**Education** (*Years*)	**Race**	**Gender**	**Participant Match** (*Code*)
J	77	16	White	Female	A
K	37	16	White	Female	B
L	50	16	White	Male	C
M	54	14	White	Male	D
N	71	18	White	Female	E
O	81	12	White	Female	F
P	53	12	White	Male	G
Q	79	18	White	Male	H
R	49	18	White	Female	I

2.1.3 Eye Tracker

The original investigators presented stories on a 55cm × 18cm ×_6cm screen using Tobii Pro Lab software©. They used a Tobii Dynavox Pro Spectrum© to collect eye movement data while participants read stories. They set the eye tracker resolution to 1920 × 1080 progressively displayed pixels, and set the sampling rate to 1200 Hz.

2.1.4 Experimental Materials

All materials were created in Microsoft PowerPoint© with black, 28-point Times New Roman font on a white background, and in alignment aphasia-friendly formatting characteristics (i.e., enlarged text and increased white space; Brennan et al., 2005; Rose et al., 2003). Materials included instructions, twenty stories serving as experimental stimuli, and one additional practice story. All materials were exported and projected onto the screen used during the experiment.

All narratives included a main character, problem, and resolution, and had a Flesch-Kincaid Grade Level of 6.0 for readability (Flesch, 1979; Kincaid et al, 1975). Each story contained five or six sentences (M = 5.4, SD = 0.49) and 65–70 words (M = 67.76, SD = 1.3). Words typically measured >1cm in length, with the exception of one- to two-letter function words (e.g., “it” or “a”). Spaces between words were approximately 0.75 cm, and lines were double-spaced. Finally, story titles were provided at the top of the page and centered. Each story included two gist comprehension questions. They were presented as multiple-choice questions with four options. See Appendix A for a sample narrative.

2.2 Procedures

Consent, intake, and data collection procedures were completed during one, in-person session. Consent procedures were pre-approved by the Institution Review Board (IRB) at Miami University of Ohio. After verbalizing consent, participants went through assessment and screening procedures followed by four runs of the experimental protocol, which included eye tracker calibration, review of task instructions, a practice narrative with accompanying gist questions, and five experimental stories with two follow-up gist questions per story. The original experiment utilized sets to create time separation between control and text-to-speech conditions. The assignment of stories and conditions was random to control for confounds such as practice effects. Figure 1 illustrates a summary of experimental procedure.

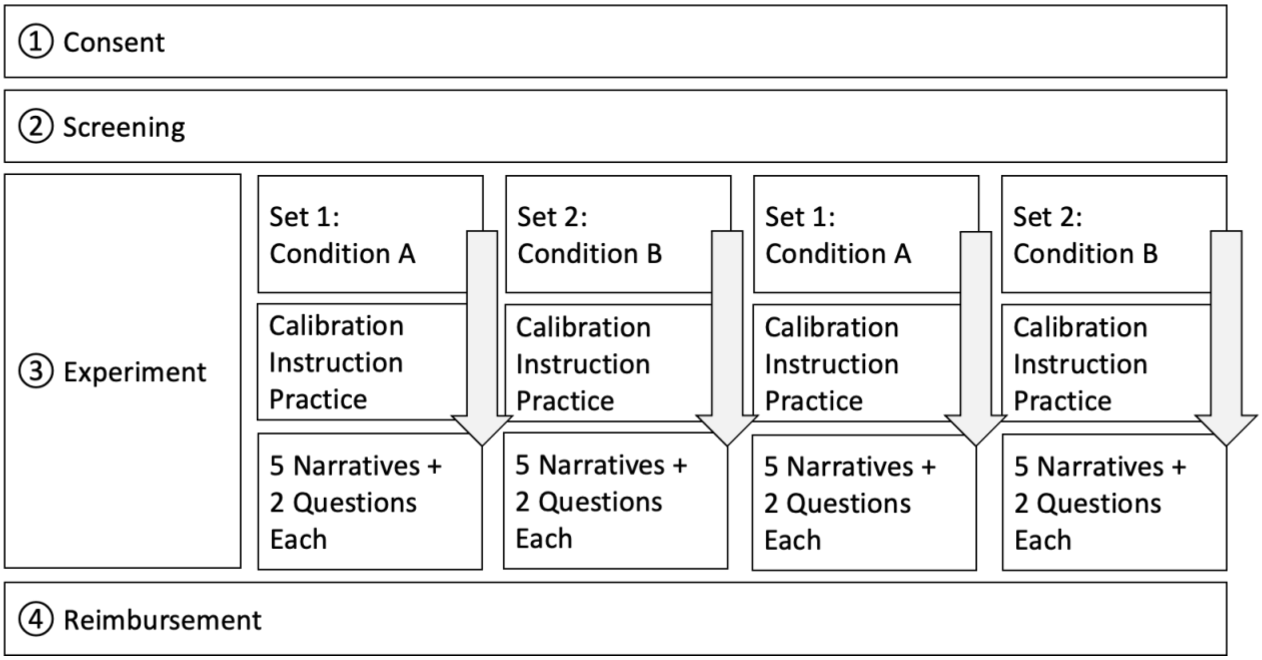

Figure 1 Summary of Procedures for Data Collection (Figure Adapted from Knollman-Porter et al., 2022)

2.2.1 Calibration

At the beginning of each story set, participants completed eye gaze calibration procedures to ensure estimated gaze position and eye movement tracking were accurate for the subsequent trials of the experiment. The procedures for calibration were adopted from Huck et al. (2017), where participants were prompted to fixate nine points on a screen in a standard sequence to calibrate. The eye tracking system then used data from this procedure to compute a validation accuracy, which represents the average difference between the targeted region of fixation versus the participant's true region of fixation, as measured though degree of difference (Raney et al., 2014). Across participants, average validation accuracy was 0.497° angle (SD = 0.24°; range: 0.19°-0.98°).

2.2.2 Experiment Presentation

Following the calibration procedures, participants were provided with task instructions. Instructions were displayed to participants on the screen and read aloud by the investigator to facilitate comprehension of the instructions. Participants had the opportunity to ask questions and were encouraged to indicate verbally that they understood the task. Following reinforcement of the instructions, participants were presented with a black transition screen with a red dot in the middle, intended to draw the gaze to a consistent region of the screen before every trial. The transition screen was repeated before every practice and experimental narrative throughout the experiment.

Participants then six stories and answered two gist questions per story. The first story and questions in each set served as an opportunity for practice. Following each practice round, participants read and answered questions for five experimental stories. The research team did not restrict the time participants had to view each story and encouraged them to read each story silently. After participants finished each story, they would need to indicate readiness for gist questions using a verbal or gestural cue. The participants viewed the questions and choices on the screen while the investigator read each question and choices aloud. The participants were permitted to respond by pointing, indicating the letter matching their choice, or reciting the full answer aloud. Optional breaks were offered at the end of each experimental set. Eye movements during the story were captured in an audio and video recording using the AceThinker Screen Grabber Pro software©. Reviewing the fixation patterns and response to comprehension questions after practice provided investigators with additional confidence in the participant's understanding of the study instructions.

2.3 Data Preparation and Analysis

A sequence of steps was completed to investigate the relationship between independent variables of interest (participant group, computer-estimated predictability) and eye movement outcome measures. These steps included the selection of stimuli, objective measurement of the independent and dependent variables of interest, and analysis of a relationship between variables. I participated in all steps of the analysis with the support from qualified mentors, collaborators, and student research assistants.

2.3.1 Stimuli

To ensure a controlled but powerful analysis of word predictability effects, stimuli were restricted to story and title nouns (n= 259). Nouns were identified through part of speech (POS) tagging. Nouns were selected because they were prevalent in the stories, providing opportunity for a powerful yet constrained analysis. To identify nouns in the stories, all words from the narrative were entered into a Microsoft Excel© spreadsheet, and judged by myself and Micheal Walsh Dickey, PhD (PTARI principal investigator and thesis co-chair). We reviewed each word and labeled those which were nouns in their context. Two rounds of this process were completed to verify agreement between judges before finalizing the experimental stimuli. Proper nouns (e.g., Megan) and spaced compounds (e.g., ice cream) were not included in the final subset of stimuli.

Estimates of target word surprisal were computed by Alex Swiderski (PTARI PhD Candidate). Preparation for computer estimation, first required that I isolate preceding story context ($w_1 \ldots w_{i-1}$), which included story title, for each target word identified in the POS tagging protocol (w_i). Alex inserted the novel story text into GPT-2 (Radford et al., 2019), a large language model (LLM) accessible through the transformers library (Wolf et al., 2020) in Python 3.0. GPT-2 was used to compute probability (P) values, which could be transformed to more interpretable scales of measurement using the surprisal function: $\text{surprisal}(w_i) = -\log P(w_i | w1 \ldots w_{i-1})$ (Lowder et al., 2018). GPT-2 was determined to be the best tool for estimating word probability because of evidence that correlates word prediction by GPT-2 to brain responses (Schrimpf et al., 2022) and reading time measures captured with eye tracking (Shain et al., 2022), above and beyond more sophisticated versions of the same LLM (e.g., GPT-3) and other LLMs (e.g., BERT). Taken together, GPT-2 has successfully accomplished predictive processes and shows correlation to neurocognitive measures of language processing, justifying the use of surprisal measures in place of cloze predictability measures when estimating word predictability.

GPT-2 was also used to predict the next word given preceding story context (including story title) for each target word. The purpose of executing next word prediction was to explore a possible measure or correlate of surprisal effects. Specifically, the present study explored if surprisal was correlated with the semantic (cosine) distance between a target word (e.g., NEIGHBORHOOD) and a GPT-2 predicted word (e.g., HOUSE) for the same context, and if these variables made similar contributions to the eye tracking measures explored in this study. Word2vec (Mikilov et al., 2013) was used to estimate cosine distance between the target word and GPT-2 predicted word. I will refer to this measure as “target-prediction cosine distance” for the remainder of this paper.

2.3.2 Eye Tracking Analysis

With the support of the research assistants at Miami University of Ohio and University of Pittsburgh, eye movement parameters were manually entered into a Microsoft Excel© spreadsheet template (illustrated in Appendix B) based on careful frame-by-frame assessment of story recordings in the Tobii Pro Lab Software. Prior to data extraction, all coders used a function in the Tobi Pro Lab Software to filter fixations below 100ms. Coders worked in pairs to access and record fixated words, letters fixated in word, fixation durations, the word number and line number corresponding to fixated words, a count of total number of fixations on a word throughout the story, and a count of total consecutive (i.e., uninterrupted) fixations of words through the story. This information provided raw data to support computation of desired eye tracking measures in the final analysis.

The analysis targeted four outcome variables which show historical sensitivity to predictability. These measures were probability of first pass fixation (PF) probability of regression (PR), gaze duration (GD), and total fixation (TFD) (see section 1.22 for review of the eye tracking evidence). Micheal Walsh Dickey, PhD and Tessa Warren, PhD (thesis committee member) developed Microsoft Excel© algorithms to transcribe the dataset (outlined in Appendix D). In a separate tab, outcome measure Pivot Tables were created to summarize the outcomes per observation (i.e., measure per word given story and participant). A final tab was created to include pivot table duplicates, where tables were inserted into a separate spreadsheet tab using the ‘paste values only’ function to ensure tables were amendable for adding missing values (i.e., words skipped by individual participants). Table 3 provides a summary of definitions for each selected outcome variable.

Table 3 Definition of Eye Tracking Parameters

Measure	**Abbreviation**	**Definition**
Probability of First Pass Fixation	PF	Binary value representing whether the participant passed over (0) or fixated (1) a target word on their first pass through the region
Probability of First Pass Regression	PR	Binary value representing whether a participant advanced in the text (0) or regressed to previous text (1) following a first pass fixation of the target word. Only words fixated by a participant on first pass were included in analyses for this variable
Gaze Duration	GD	Value representing the amount time a participant fixated on a target word on first pass before shifting to a different word in the text, whether this duration was accrued from a single fixation or several consecutive fixations on the target word. Only words fixated by a participant on first pass were included in the final analysis of this variable
Total Fixation Duration	TFD	Value representing the total amount time a participant fixated on a target word over the course of reading. Words fixated on or after a first pass were included in the final analysis, and words skipped entirely were filtered out of final analysis

Note*:* the above measures do not include data from fixations of 100 or fewer milliseconds, and do not account for text (pre-)viewed in the participant's parafoveal visual field

2.3.3 Statistical Analysis

All data was analyzed using linear mixed effect models scripted and ran in R using the lme4 library (Bates et al., 2015). This statistical approach was taken to account for variance across individual stories and participants. The models were written to establish the relationship between three fixed effects (group, surprisal, group*surprisal) and four eye movements outcomes (PF, PR, GD, TFD), where separate models were scripted for each outcome variable. Binary values (PF, PR) were analyzed using separate generalized linear mixed effect models (GLMERs). Continuous variables (GD, TFD) were individually log-transformed due to an abnormal distribution of each continuous variable. GD and TFD were analyzed post-transformation using separate linear mixed effect models (LMERs).

All models featured random effects (participant, story) to account for differences due to the individual reader or story read. Additional covariates, which were expected to have complementary or confounding effects, were strategically added as fixed effects to individual regression models.

The PF model featured a word length covariate due to the establishment of word length effects on skipping reported in the literature (Brysbaert et al., 2005; Rayner, 2009; Rayner et al., 1998). Word length was computed using a Microsoft Excel© algorithm, and manually entered into the R data structure. The GD and TFD models were each fitted with a word frequency covariate due to frequently reported effects of word frequency on reading time measures (Huck et al., 2017; Rayner et al., 1998; Rayner et al., 2009). For all target words, subtitle word frequency measures (Brysbaert & New, 2009) were extracted from the English Lexicon Project (Balota et al., 2007), and then manually entered into the R structure.

Finally, models for PF, GD and TFD were fitted with a covariate of word2vec (Mikilov et al., 2013) estimated cosine distance between original target words (e.g., NEIGHBORHOOD) and words predicted by GPT-2 (e.g., HOUSE) given preceding context. The addition of target-prediction cosine distance as covariate to the PR model was also attempted but not executed due to a failure of the model to converge. Appendix E lists the final models, written by Candace van der Stelt (PTARI PhD Student) and Alex Swiderski, used to generate results in R.

3.0 Results

We tagged 259 nouns (approximately 10-16 nouns per story) and all were included in the final analysis. Appendix F provides a sample list of stimuli and preceding story context. Following presentation of target words and preceding context to GPT-2, it was determined through one-way analysis of variance (ANOVA) that there was no significant difference in mean target word probability across stories. GPT-2 probability output was used to estimate surprisal, and the surprisal values of target words ranged from 6.070-21.15 (M=12.82, SD=3.02). Analysis of covariate values revealed that target-prediction cosine ranged from 0.0000-1.066 (M= 0.54, SD=0.37). Word frequency of target words ranged from 1.40-5.60 per million words (M=3.44; SD=0.92), and word length of target words generally ranged from 3-13 characters (M=5.80; SD=2.10), with exception of one target word, "TV".

3.1 Fixed Effects and Covariate Outcomes

The results are broken down by research question, with the addition of section describing covariate effects. A summary of all findings by outcome variable is illustrated in Table 4. Figures 2-5 visualize the fixed effect findings outlined in the upcoming sections, and Figures 6-10 visualize the covariate effect findings outlined in section 3.1.4.

3.1.1 Group Effects

The first research question inquired about the effect of group (neurotypical adults, people with aphasia) on probability of first pass fixation (PF), probability of first pass regression (PR), log-transformed gaze duration (logGD), and log-transformed total fixation duration (logTFD). The analysis revealed that people with aphasia and neurotypical adults demonstrated no significant difference in PF (β=0.17247, SE=0.5382, p=0.74710) or PR (β=0.34494, SE=59783, p=0.77966). Significant differences between neurotypical adults and people with aphasia were identified for both logGD (β=0.61790, SE=0.18750, p=0.00160) and logTFD (β=1.13000, SE= 0.21680, p=0.00001), where people with aphasia demonstrated higher fixation duration measures compared to neurotypical adults.

3.1.2 Surprisal Effects

The second research question inquired about the effect of surprisal on PF, PR, logGD, and logTFD. Surprisal was found to have a significant relationship with PF, where increases in surprisal were associated with increases in PF (β=0.08780, SE=0.02006, p=0.00001). Conversely, increases in surprisal were associated with significant decreases in PR (β= -0.07022, SE= 0.02312, p=0.00238). Analysis of temporal measures also revealed that increases in surprisal were associated with significant decreases in logGD (β= -0.01636, SE= 0.00540, p= 0.00251) and logTFD (β= -0.00988, SE= 0.00490, p= 0.04400).

3.1.3 Interaction Effects

The third research question inquired about the differences of surprisal effects on PF, PR, logGD, and logTFD across groups (neurotypical adults, people with aphasia). The analysis of interaction effects was found to be unremarkable, as there were no significant effects of surprisal and group across outcomes.

3.1.4 Covariate Effects

Covariate effects were included in the present study's models to account for well-evidenced effects of word length on PF and word frequency on fixation duration metrics (logGD, logTFD), and to explore the relationship between surprisal and target-prediction cosine distance. The results suggested that there were significant effects of word length on PF, where increases in word length were associated with increased PF (β=0.16674, SE=0.02506, p=0.00000). It was also found that both word frequency had significant effects on both logGD (β=0.0532, SE=0.01625, p=0.00174) and logTFD (β=-0.07284, SE=0.01486, p=0.0001). Finally, target-prediction cosine distance appeared to have significant positive effect on logGD (β=0.09890, SE=0.04049, p=0.01555) and logTFD (β=0.01647, SE=0.03810, p=0.00002) where increases in target-prediction cosine were associated with increases in both fixation duration measures. In contrast, cosine distance had a nonsignificant effect on PF.

Table 4 Summary of Fixed and Covariate Effects on Eye Movement

	Observations	Main Effect	*β*	SE	z	pr(<|z|)
Probability of First Pass Fixation (PF)	2136	Group	0.17247	0.53482	0.322	0.74710
		Surprisal	0.08780	0.02006	4.377	0.00001***
		Group*Surprisal	0.07188	0.03686	1.950	0.05120
		Cosine Distance	0.10992	0.02506	0.803	0.42220
		Word Length	0.16674	0.02506	6.652	0.00000***
Probability of First Pass Regression (PR)	1537	Group	0.34494	0.59783	0.280	0.77966
		Surprisal	-0.07022	0.02312	-3.038	0.00238**
		Group*Surprisal	0.06471	0.04486	1.442	0.14918
Gaze Duration (GD)	1525	Group	0.61790	0.18750	3.296	0.00160**
		Surprisal	-0.01636	0.00540	-0.302	0.00251**
		Group*Surprisal	0.00381	0.01028	-0.370	0.71127
		Cosine Distance	0. 09890	0.04049	2.423	0.01555*
		Word Frequency	0.05132	0.01625	-3.159	0.00174**
Total Fixation Duration (TFD)	1818	Group	1.13000	0.21680	5.226	0.00001***
		Surprisal	-0.00988	0.00490	-2.015	0.04400*
		Group*Surprisal	0.00597	0.00958	0.623	0.53000
		Cosine Distance	0.01647	0.03810	4.323	0.00002***
		Word Frequency	-0.07284	0.01486	-4.900	0.00001***

Key: β = Beta Coefficient, SE= Standard Error, z = Z-Score, pr(<|z|) = P-Value

Significance Levels: *= p<0.05, **=p<0.01, ***=0.001

Note: PR and GD did not have an equal number of observations due to filtering by the presently scripted models

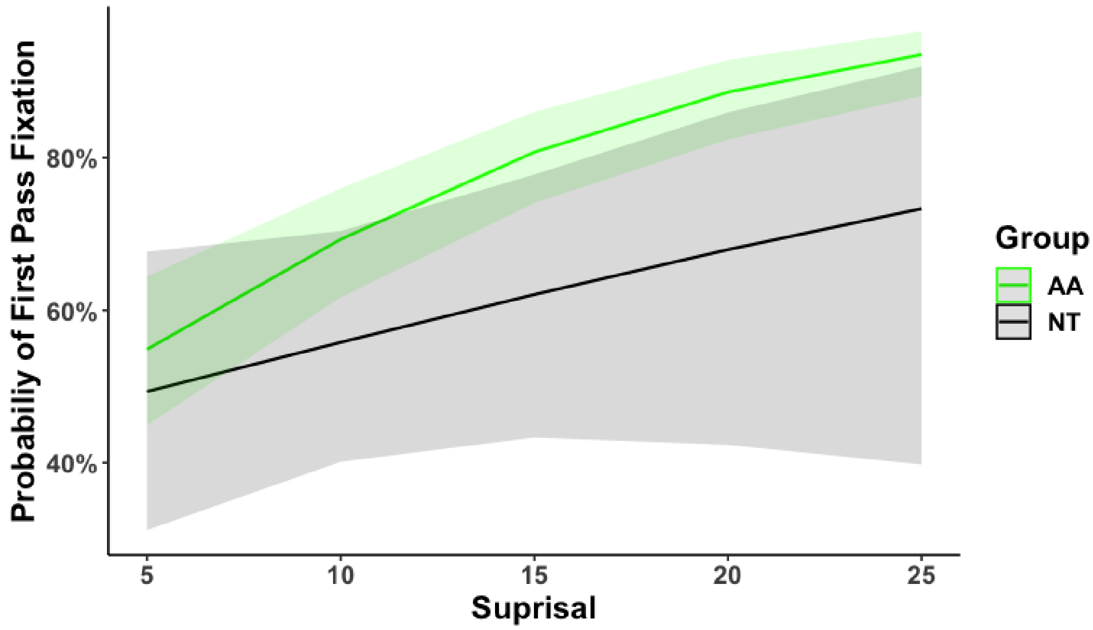

Figure 2. Graph representing the relationship between surprisal and probability of first pass fixation for neurotypical adults (NT) and people (adults) with aphasia (AA)

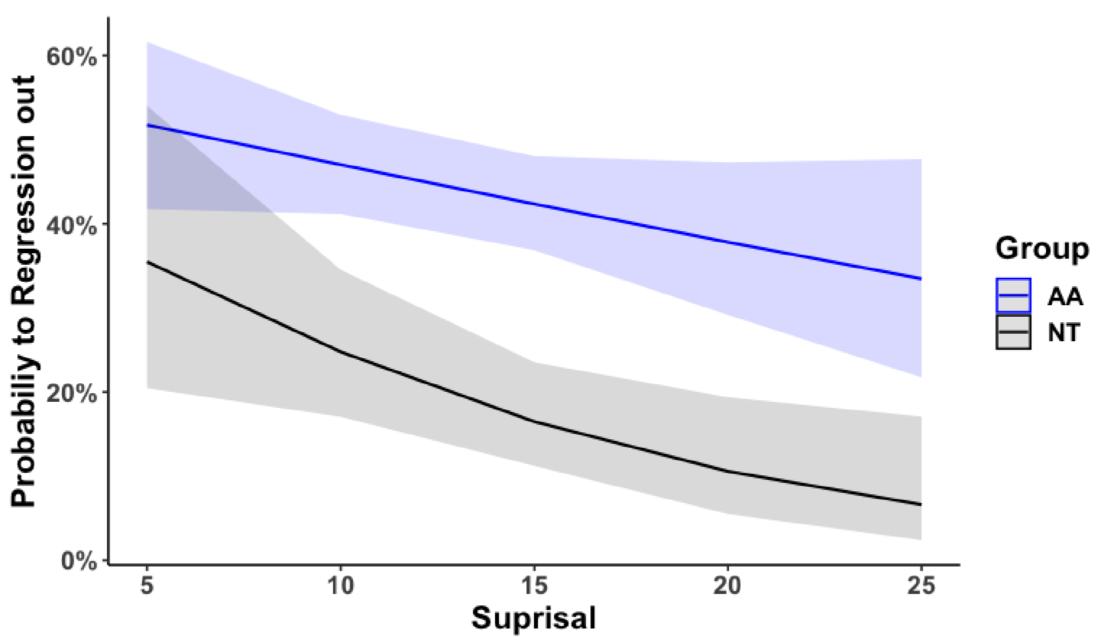

Figure 3. Graph representing the relationship between surprisal and probability of first pass regression for neurotypical adults (NT) and people (adults) with aphasia (AA)

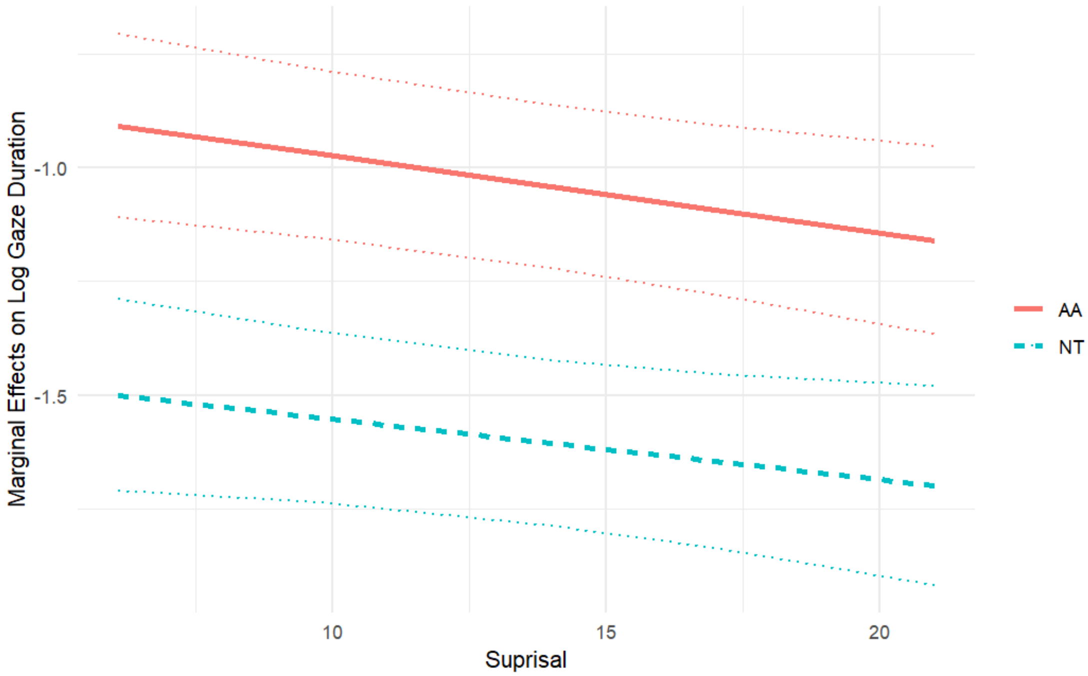

Figure 4. Graph representing the relationship between surprisal and gaze duration for neurotypical adults (NT) and people (adults) with aphasia (AA)

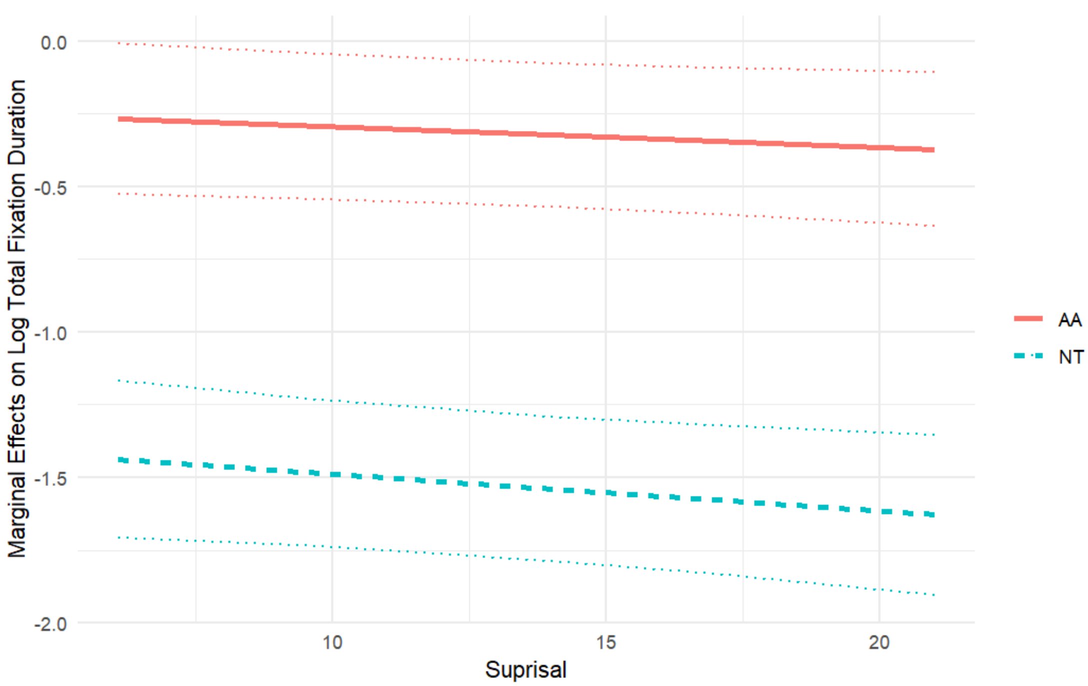

Figure 5. Graph representing the relationship between surprisal and total fixation duration for neurotypical adults (NT) and people (adults) with aphasia (AA)

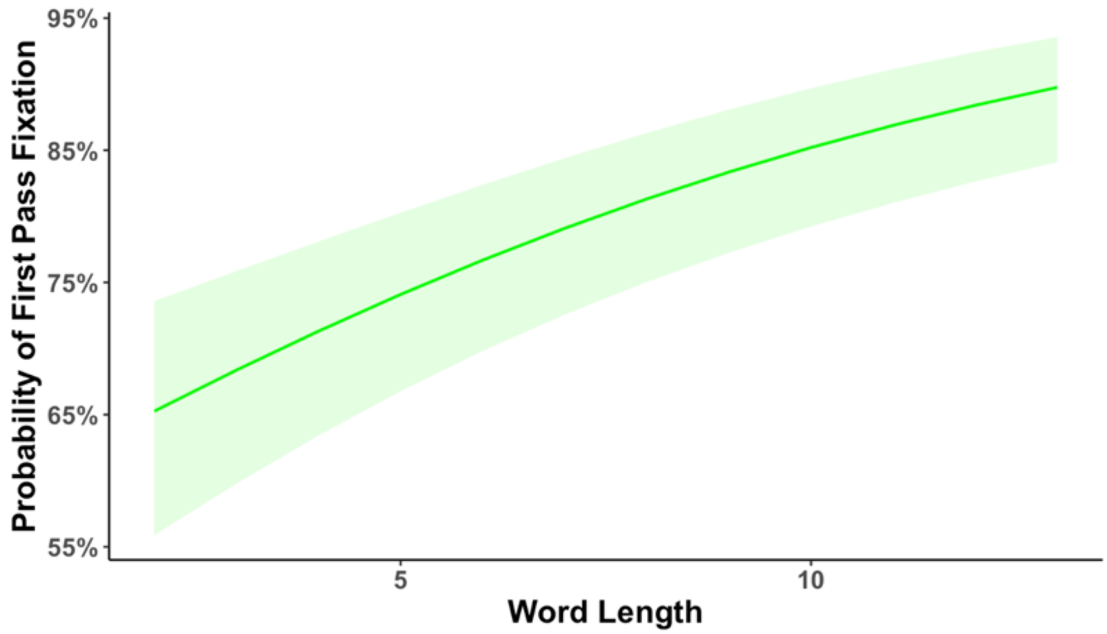

Figure 6. Graph representing the relationship between word legnth and probability of first pass fixation across participant

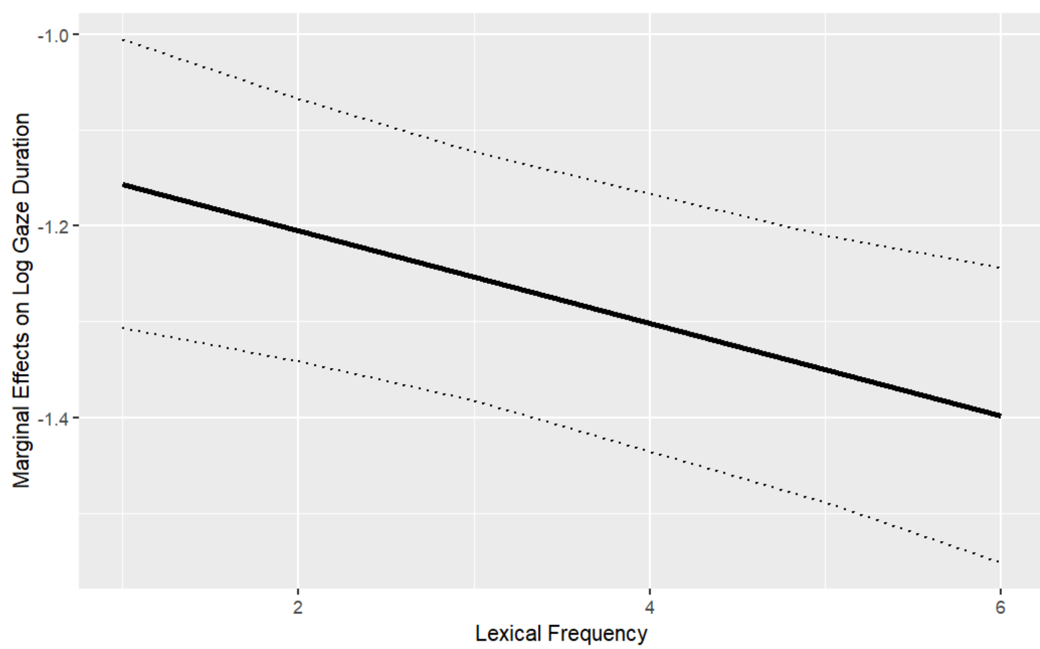

Figure 7. Graph representing the relationship between word frequency and gaze duration across participants

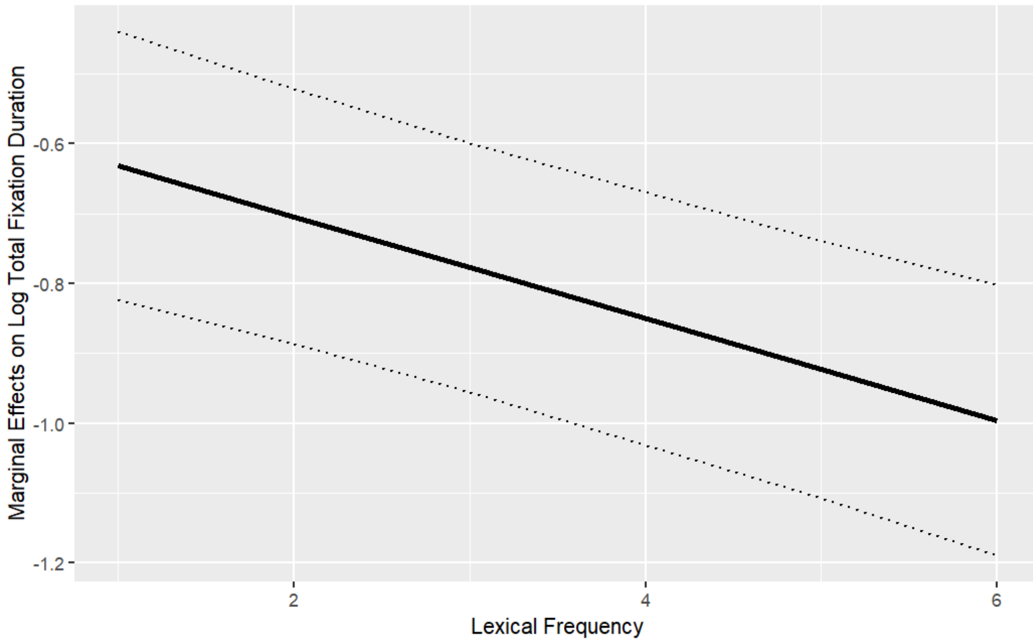

Figure 8. Graph representing the relationship between word frequency and total fixation duration across participants

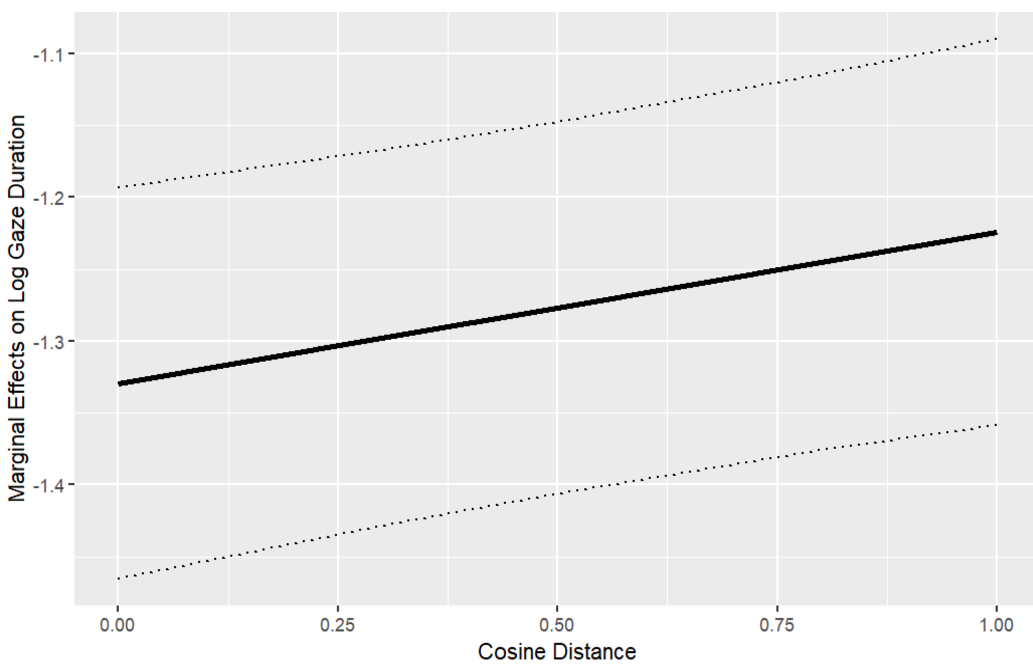

Figure 9. Graph representing the relationship between target-prediction cosine distance and gaze duration across participants

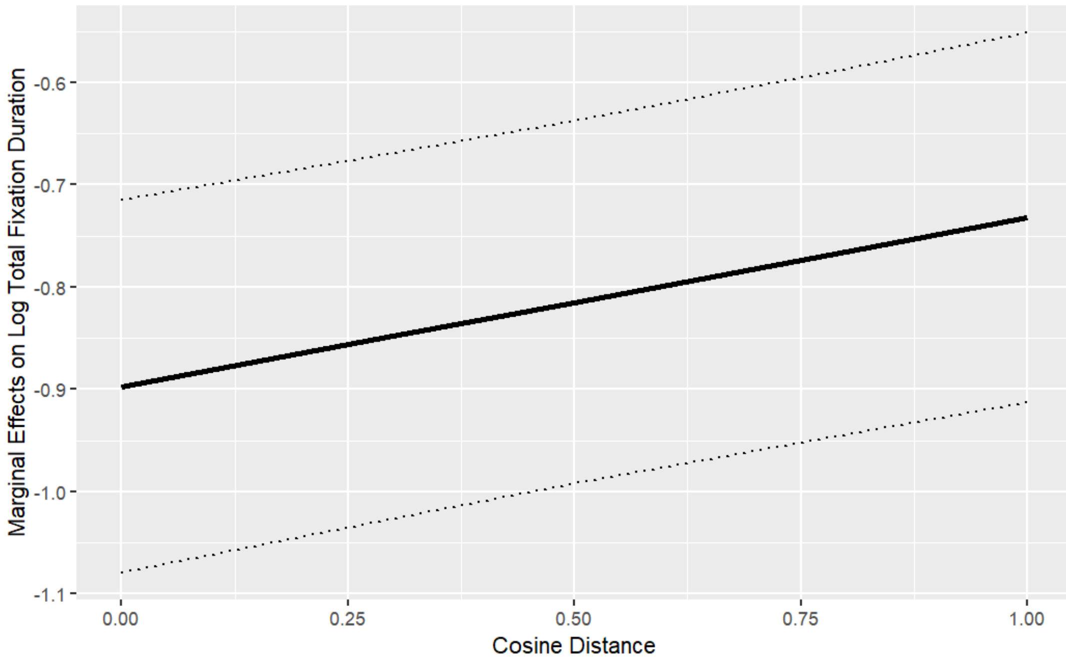

Figure 10. Graph representing the relationship between target-prediction cosine distance and total fixation duration across participants

3.2 Random Effects

An analysis of random effects evaluated the contribution of assigned story and individual participant to the distribution of data. For all four outcome variables (PF, PR, logGD, logTFD), story accounted for minimal to no variance in the dataset, and participant accounted for minimal variance in the data. Table 5 summarizes the variance in fixed effect given random effect.

Table 5 Summary of Variance in Eye Movement by Story and Participant Reader

Measure	**Observations**	**Random Effect**	**s^2**
Probability of First Pass Fixation	2316	Story (n=20)	0.21675
		Participant (n=18)	0.11603
Probability of First Pass Regression	1537	Story (n=20)	0.03043
		Participant (n=18)	0.10080
Gaze Duration	1574	Story (n=20)	0.00145
		Participant (n=18)	0.07192
Total Fixation Duration	1818	Story (n=20)	0.00000
		Participant (n=18)	0.13920

Key: s^2 = variance

4.0 Discussion

The present study investigated the effects of computer-estimated predictability measure (surprisal) on salient eye tracking variables (probability of first pass fixation, probability of first pass regression, gaze duration, total fixation duration) during story reading in neurotypical adults and people with aphasia. The study findings were both consistent and inconsistent with expectations from previous literature. The goal of this discussion is to review replicated and contradictory research outcomes, reason with methodological limitations, and outline theoretical and clinical implications of the research findings.

4.1 Group Effects

The first objective of the present study was to investigate the effect of participant group (neurotypical adults, people with aphasia) on probability of first pass fixation (PF), probability of first pass regression (PR), gaze duration (GD), total fixation duration (TFD) during story reading. Contrary to the original hypothesis, people with aphasia did not demonstrate significantly higher PF and PR compared to neurotypical adults. This finding also contrasts previous evidence regarding sentence-level reading in neurotypical adults and people with aphasia, where people with aphasia demonstrated significantly higher PF and PR on target words in sentences (Huck et al., 2017). In contrast, the group analysis also showed that people with aphasia demonstrated significantly greater GD and TFD compared to neurotypical adults, which is both consistent with the original hypothesis and the Huck et al. (2017) findings.

Overall, the results inconsistently support the hypothesis that people with aphasia would demonstrate higher values in all four sensitive eye tracking measures (PF, PR, GD, TD) compared to neurotypical adults. There are several possible explanations for incongruency with previous findings regarding sentence-level effects of predictability in neurotypical adults and people with aphasia. One factor that may have influenced the current findings is the characteristics of the people with aphasia who participated in the present study. Of note, the people with aphasia included in our analysis completed the paragraph-factual reading subtest of the Reading Comprehension Battery for Aphasia—Second Edition (RCBA-2; LaPointe & Horner, 1998). This tests an ability to match pictures to short paragraph, and the majority of participants demonstrated a high level of accuracy on this subtest (see Appendix C for a review of their performance). Our participants also demonstrate a high level of accuracy on gist questions that followed each story in the experiment. Since similar measures were not adopted by Huck et al. (2017) and since their experiment demanded a lower level of reading, it is possible that the participants in this previous work exhibited different reading characteristics from the present study participants. Future research may be able to explore the consequence of aphasia profile versus task demand by testing both sentence- and paragraph-level reading in the same sample of people with aphasia.

The recruitment of neurotypical adults who were age-matched to people with aphasia, must also be considered because this resulted in a generally older demographic than the samples typically recruited for study of predictability effects on eye movement. Review of these studies reveal that participants were typically young adults (e.g., Vitu et al., 1991) and/or college undergraduates (e.g., Rayner et al., 2004). It is true that inclusionary criteria of the present study ensures that neurotypical adults had sufficient vision, hearing, and cognition to meet the demands of the task, ruling out age-related changes that could impair participant ability to engage in the task. However, it cannot be assumed that younger and older adults are affected by predictability to the same degree. Perhaps the nonsignificant group differences in PF and PR could be explained by a healthy aging effect, in which neurotypical the PF and PR of older adults resemble that of people with aphasia and differ from that of neurotypical younger adults (DeDe, 2014; Rayner et al., 2006). However, there is no research directly comparing eye movements during reading in older adults, younger adults, and people with aphasia, or examining the effects of predictability on reading across these groups. Further research is required to test this possible explanation for the absence of group effects in the current study.

Finally, the use of a simple comprehension task (i.e., gist questions) following story readings may have influence participant reading behavior, and ultimately change the eye movements measured for neurotypical adults and/or people with aphasia. There is evidence to suggest that use of different reading tasks can yield different patterns of eye movement by readers (Heller, 1982; Radach, 2007, Tinker, 1958). Specifically, these authors highlight that a skilled reader will change their pace of reading, as reflected in word-level fixation durations, based on the reading demands and purpose (i.e., unstructured or superficial reading versus reading for comprehension). The present research study was neither designed to test reading comprehension rigorously, nor was it an unstructured reading task. This specific purpose of reading could have had a top-down influence on fixation duration measures.

4.2 Surprisal Effects

The second objective of this work was to investigate the effect of surprisal (operationalized by computing the negative logarithm of GPT-2 estimated probability of a word given context), on relevant eye tracking measures (PF, PR, GD, TFD) during story reading. The results demonstrated that increased surprisal (indicating reduced predictability) was associated with a significant increase in the probability of PF. This finding is consistent with the original hypothesis and previous studies measuring predictability effects on PF sentence-level reading (Balota et al. 1985; Binder et al., 1999; Rayner et al., 2011, Rayner & Well, 1996) and paragraph-level reading (Ehrlich & Rayner, 1981; Schustack et al., 1987). In contrast, increases in surprisal were related to decreases in PR, GD, and TFD. These findings were inconsistent with the original hypothesis and opposite from previous knowledge about predictability effects on PR (Boston et al., 2008; Kliegl et al., 2004), GD (Balota et al. 1985; Binder et al., 1999; Ehrlich & Rayner, 1981; Rayner, 2004; Rayner et al., 2011; Rayner & Well, 1996; Schustack et al., 1987 Vitu, 1991) and TFD (Balota et al. 1985; Binder et al., 1999; Calvo & Meseguer, 2002; Ehrlich & Rayner, 1981; Kliegl et al., 2004; Rayner, 2004; Rayner et al., 2011, Rayner & Well, 1996; Schustack et al., 1987).

Taken together, the results of this research generally refuted the original hypothesis, which expected increases in surprisal to yield increases in relevant eye-movement measures during reading. With the exception of PF, the present study revealed that increases in word surprisal were associated with decreases in these eye tracking measures. While the findings of this work are inconsistent with previous research studies, there may be theoretical rationale for the observations noted in the present study. One explanation, which could account for the PR and

GD conflicts, may be a strategic use of subsequent contextual information to resolve ambiguity in the text, in either people with aphasia or the collective sample. Evidence for use of such a strategy can be derived from an experiment where the eye movement of neurotypical adults was tracked while reading temporarily ambiguous phrases contained within a sentence (Frazier & Rayner, 1982). The authors found that rather than consistently “backtracking” to resolve sentence ambiguity, readers instead often exploited information in the post-ambiguous region of the sentence. This pattern might extend to the current findings for hard-to-process, high-surprisal words, where participants may have responded to highly unpredictable words by strategically moving ahead to subsequent material with the goal of finding additional helpful information. Further analysis would be needed to test this explanation, such as testing whether high surprisal words had more progressive fixations than low surprisal words (meaning people were more likely to move past them to read subsequent words).

An alternative explanation may be oversight of a word position in text effect, which is known to decelerate reading time at the end of lines and passages, regardless of the linguistic object encountered (Kuperman et al., 2010). The selection of noun targets, in stories with high conformity to active sentence structures, may have resulted in a large sampling of object nouns occurring at the ends of sentences, and potentially lines and stories. Since surprisal values are frequently lower for words that come later in a sentence, these late-occurring words may have had lower surprisal values. This coincidence may be a reason that words with lower surprisal values were read slower than words with higher surprisal values, contrary to previous findings (Balota et al. 1985; Binder et al., 1999; Ehrlich & Rayner, 1981; Rayner, 2004; Rayner et al., 2011; Rayner & Well, 1996; Schustack et al., 1987 Vitu, 1991). One way to investigate this possibility further would be to include word position metrics (e.g., word position in line or sentence) as a covariate in the models to assess a partial or supplemental contribution of word position in text to GD and TFD.

While there are possible theoretical explanations for these behaviors, aspects of the methodology had the potential to influence any or all outcomes, and more specifically the current application of a surprisal measure. The majority of previous research on predictability and eye movement during reading in neurotypical adults (Balota et al. 1985; Binder et al., 1999; Calvo & Meseguer, 2002; Ehrlich & Rayner, 1981; Huck et al., 2017; Kliegl et al., 2004; Rayner et al., 2004; Rayner et al., 2009; Rayner & Well, 1996; Schustack et al., 1987; Vitu, 1991), along with the recent study of predictability and eye movement in people with aphasia (Huck et al., 2017), used cloze predictability to measure the expectedness of a word in context. Surprisal, which is a fairly new measure of predictability effects in the eye tracking literature (Boston et al., 2008; Demberg & Keller, 2008), has key differences from cloze predictability. Most notably, cloze predictability is frequently used as a categorical variable, where words may be classified as "predictable" or "unpredictable" (e.g., Huck et al., 2017) or "low-, medium-, and high-predictability" (e.g., Rayner & Well, 1996), through reference of normative thresholds for varying degrees of cloze predictability. Surprisal, in contrast, is treated as a continuous variable, and this is important to note because this work did not selectively assess extreme levels of predictability. In order to make the current surprisal measure more comparable to previous studies of cloze predictability, I could consider transforming the data and selectively analyzing target words which fell into normative "high and low" or "high, medium, and low" ranges of cloze predictability, and then see if this alters eye movement patterns.

4.3 Interaction Effects

The final objective of this work was to explore potential differences in surprisal effects between participant groups (neurotypical adults, people with aphasia) on eye tracking measures of interest (PF, PR, GD, TFD) during story reading. While there was a marginal interaction of group and surprisal on PF (i.e., the difference in PF for neurotypical adults and people with aphasia was more pronounced under the condition of higher target word surprisal), there were no other significant interactions between surprisal and group. There is only one known paper which explored the comparison of predictability effects in neurotypical adults and people with aphasia, and this paper revealed no significant interactions of group and predictability for measures of PF, PR, and GD in sentence level reading (Huck et al., 2017), just as the present study did. Huck et al. (2017) also found that predictability effects on TFD were exaggerated in people with aphasia compared to neurotypical adults. Specifically, they found a greater TFD disparity for unpredictable and predictable words in people with aphasia than neurotypical adults, and this was replicated in the current work. While the current study presents evidence that is inconsistent with this interaction finding, there is ultimately no other known work to provide evidence of the interaction of aphasia and predictability during reading.

Another relevant body of evidence for interpreting the current interaction effects is the literature on predictive processing in aphasia using visual world eye tracking paradigms and self-paced reading. The majority of these studies found that people with aphasia demonstrated similar responses to predictable stimuli as neurotypical adults (Hanne et al., 2015; Hayes et al., 2016; Mack et al., 2017, Warren et al., 2016). This suggests that people with aphasia may have spare predictive processing skills, making a nonsignificant interaction of predictability and group evidence of a similar phenomenon.

4.4 Covariate Effects and Random Effects

The present study objectives were investigated with consideration of other relevant variables which would partially influence reading outcomes. These variables were word length, word frequency, and target-prediction cosine distance. Consistent with the literature, word frequency had a significant effect on measures of GD and TFD, with longer GD and TFD for lower frequency words. Furthermore, word length was associated with a significant increase in PF.

While target-prediction cosine distance had nonsignificant effects on PF and PR, it had a significant positive relationship with GD and TFD. That is, we found longer GD and longer TFD on words that were more semantically distant from (less semantically related to) the most expected word generated by GPT-2 (Radford et al., 2019). This pattern compliments results from a study on the effects of predictability on event-related potentials (ERP), which found that the that the N400 penalty for an unexpected (surprising) word was reduced when this unexpected word exhibited greater semantic similarity to the expected (predictable) word in the same context (e.g., reduced N400 for the pair PALMS-PINE than PALMS-TULIPS) (Federmeir & Kutas, 1999). Their interpretation of this finding was that people anticipate and pre-activate a word during context reading, which then spreads to other related words in the lexicon and possibly lessens the difficulty of a word that is unexpected but semantically similar to the word pre-activated in context. If this hypothesis were correct, it would possibly explain why eye movement for surprising words in their context could be read for shorter amounts time, given control of the semantic distance between a “pre-active” (predicted) words and true text.

One principle that is clear from this analysis is that the effects of target-prediction cosine distance are independent of surprisal effects, where the effect of cosine distance (a measure of semantic similarity) was significant when included in models assessing surprisal effects. Furthermore, our results highlight that target-prediction cosine distance and surprisal had an opposite effect on GD and TFD, despite having a weak positive correlation. The implication of this finding is the possible discovery of a robust factor (semi-correlated to predictability) which influences eye movement during reading in neurotypical adults and people with aphasia during story reading. The fact that cosine distance had a different effect from surprisal suggests that further analysis is needed to determine how these factors interact or independently contribute to eye movements and reading behavior.

4.5 Additional Considerations and Limitations

The current research outcomes may be partially attributable to the retrospective approach to the present research question, and this should be considered during the interpretation of the results. One factor with a large potential to affect research findings is that the current study was a retrospective analysis of stimuli that were not designed to answer questions about predictability effects. Cloze predictability research often treats predictability as a categorical variable, where participants are presented with target words that represented the extremes of predictability (see Huck et al., 2017). In the present study, stories were not constructed include a balanced number of words which exhibited high versus low predictability, instead requiring retrospective estimates of word predictability. These estimates were found to be normally distributed, rather than being grouped into extremes of high versus low predictability. There were also an uneven number of target words per story and readers per story.

Given these constraints, the data did not have the structure necessary for more commonly-used analyses of predictability effects, such as analysis of variance testing (see Huck et al., 2017 for example): there were unequal numbers of observations per person and story, and the main independent variable (predictability) could not be split into clearly-separated levels. The analysis approach used in the current study (linear and generalized linear mixed effect regression) does not require that these assumptions be met. Given these differences between the present study and previous research, the potential contribution of analysis approach and independent variable distribution must be considered.

While there were some limitations due to the retrospective nature of this study, our approach did provide opportunity to assess predictability effects in passages that were not highly controlled to address an original research question. The choice to use passage-level stimuli with a range of predictability values made the current study's examination of predictability effects more naturalistic and ecologically valid than traditional sentence stimuli used to study these effects. The variability associated with less-controlled texts also made it possible to perform a true continuous analysis of predictability and include intermediately-predictable target words in the analysis.

The present study was also vulnerable to issues of sampling bias. The participants included nine neurotypical adults and nine people with aphasia who were all white, had a minimum of a high school level of education, and used English as a primary language. Considering the influence of personal knowledge on contextual predictability (Balota et al., 1985), it is important to represent diverse readers in the predictability literature, and also think critically about the construction of stimuli and recruitment of strategies to compute predictability. One other consideration related to sampling bias is that the present study only featured data from individuals who could tolerate the physical demands of a rigorous eye tracking study (e.g., four sets x five stories = twenty stories total), even when accommodations available to participants (e.g., optional breaks between story sets). Future studies would also benefit from modifications to the present methodology (e.g., reducing number of consecutive or total stories) to mitigate barriers to research participation and provide opportunity for more inclusive sampling of the population this research hopes to serve.

4.6 Clinical Implications

The present study helped to characterize the effect of predictability on eye movement during story reading in people with aphasia. The present research findings, along with the robust literature on predictability effects on eye movements (Balota et al. 1985; Boston et al., 2008; Binder et al., 1999; Calvo & Meseguer, 2002; Demberg & Keller, 2008; Ehrlich & Rayner, 1981; Huck et al., 2017; Kliegl et al., 2004; Rayner et al., 2004; Rayner et al., 2009; Rayner & Well, 1996; Schustack et al., 1987; Vitu, 1991), and recent evidence of spared predictive processing in people with aphasia (Hanne et al., 2015; Hayes et al., 2016; Mack et al., 2017; Warren et al., 2016), collectively justify the continued exploration of predictability as a clinical tool. One immediate application to explore is the use of computational tools, such as GPT-2 (Radford et al., 2019), to strategically develop assessment tools, treatment materials, and treatment stimuli to support reading intervention designed for people with aphasia. Manipulation of word predictability in these materials may allow clinicians to isolate the strengths and challenges of their clients, which can inform the development of a hierarchical reading program that supports progression from highly predictable to highly unpredictable target words in contexts. This warrants future research, which investigates how neurotypical adults and people with aphasia respond to computer-generated stimuli, varied in terms of target word predictability.

This line of research can also inform a future interventional strategy which leverages residual predictive processes in people with aphasia. Specifically, if people with aphasia demonstrate a preserved sensitivity to the predictability of words in a reading context, then people with aphasia may benefit from "context reading" (Huck et al., 2017), or purposeful attention to context clues to shape predictability effects. While context is consistently available to all readers, people with aphasia may require explicit instruction, prompting, and/or practice to exercise context reading, as people who were highly skilled readers prior to their aphasia onset may not have been as influenced by the predictability of words during reading (Ashby et al., 2005).

4.7 Future Research Directions

Future research should incorporate a range of methods, tasks, materials and variables of interest to expand knowledge on predictability effects as a clinical tool. One immediate expansion of the current work would be the development of models with other potentially influential covariates. For instance, the present study may be able assess an WAB-R AQ scores (i.e., index of aphasia severity; Kertesz, 2007) as a covariate which may influence predictability effects on eye movement. This information would be relevant because no known research has assessed the relationship between aphasia severity and predictability processing advantages, and the nature of this relationship would be relevant in the tailoring of future interventions.

It would also be beneficial to account for the robust influences of word position in text and word position in sentence on fixation time outcomes to test if the discrepancy between the present research and previous findings is attributable to an overlooked confound rather than predictability (surprisal). As noted above, fixation times tend to be higher for words at the end of sentences and stories, so the higher reading times for words with lower surprisal values might be due to their position in the sentences or stories (see Warren, 2009 for review). The relative position of words in the lines of the passage may also be contributing to the reading time patterns, as some target words with low surprisal values may have also occurred at the end of a line, and this could have resulted in higher fixation times (see Kuperman et al., 2010 for review). Addressing these potential confounds would require retrospective coding of the variables, followed by implementation of target word positions as covariates in a post-hoc model.

It may also be beneficial to tag and analyze verbs from the same stories to investigate if the effects of verb predictability on eye movements is consistent with the effects on noun predictability in both neurotypical adults and people with aphasia. One rationale for this analysis is evidence of a word class effect for people with aphasia, where verb processing is typically more difficult than noun processing for people with aphasia (see Alyahya et al., 2018 for review). Assessing verb predictability may also address the open research question of how to distinguish word position in text effects from true noun predictability effects. The stories have a high conformity to active sentence structure, making it highly likely that sentences and lines ended with nouns. In contrast, verbs are more likely to appear in the middle of sentences and lines, indirectly controlling for potential word position in text or wrap-up effects.

Future research on predictability and reading comprehension would enrich our understanding of how predictability should be utilized in future reading interventions. The present study uses gist questions, but as inclusion criteria rather than as measurement of reading success. Future research may benefit from strategic development of reading comprehension questions or tasks (e.g., passage summarization) that directly relate to predictable and unpredictable words used in the text, as this would allow us to directly assess the benefit of predictability on comprehension. If these questions are featured in the context of eye tracking paradigm, this also provides researchers with the opportunity to couple eye movement behaviors and reading comprehension outcomes. It is important to isolate and compare the effects of fixation behavior and reading comprehension to understand the ability for predictive processing intervention to support different client goals in reading interventions.

There is also a benefit of conducting studies which utilize self-report measures, such as interview (see Hux et al., 2021 for example), to investigate conscious and unconscious awareness of predictability among readers, including neurotypical adults and people with aphasia. There is value to understanding an awareness of predictability effect in reading, especially given evidence of implicit learning challenges in people with aphasia (e.g., Christiansen et al., 2010). In the emergence of intervention, it would be insightful to understand if predictability is something that must be explicitly taught to afford the benefit of text predictability to our clients. In support of this idea, it has been shown that people with aphasia improve their use of agent context to complete sentence-picture matching when they had the participate in a structured syntactic intervention program (Mack & Thompson, 2013).

A final line of follow up work would investigate the weight of specific contextual factors (i.e., word relatedness, preceding syntactic constraint, reader knowledge) on the eye movements observed in neurotypical adults versus people with aphasia. While both populations demonstrate similar eye movement responses to predictable and unpredictable words in text, there is no known word which compares aspects of context that support or hinder predictability effects on eye movement in neurotypical adults versus people with aphasia. To address this with the current dataset, covariates measuring properties of the context preceding target words (e.g., preceding verb restrictiveness; see Mack et al., 2017 for review) can be used to enrich the models and isolate faciliatory and inhibitory aspects of context. Future experiments may also be able to control the level of semantic, syntactic, reader knowledge available to a reader before predictable and unpredictable target words are encountered. This would allow us to investigate the role of different contextual factors on the resulting eye movement effects across groups.

4.8 Conclusion

There is a present need to identify tailored reading interventions for people living with aphasia. Through study of the factors affecting reading, researchers can learn to manipulate materials or behavior to optimize reading. The present study aimed to contribute to this literature by investigating the effects of computer-estimated word predictability (surprisal) on eye movement during story reading in neurotypical adults and people with aphasia. The results demonstrated significant differences in the first pass fixation times (GD) and total fixation times (TFD) during story reading in neurotypical adults and people with aphasia, while illustrating nonsignificant differences in PF and PR measures at this story level. This study also replicated an effect of higher PF given increases in target word surprisal (decreases in target word predictability), but demonstrated a reversal of previously cited predictability effects through a finding of lower PR, GD, and TFD in response to increases in target word surprisal. Finally, this work found generally comparable effects of surprisal on neurotypical adults and people with aphasia.

One key takeaway from this study is that people with aphasia and neurotypical adults demonstrated both unique and comparable eye movement patterns in story reading to participants enrolled in previous sentence-level and paragraph-level reading studies. Since no previous studies have assessed reading behaviors of people with aphasia and age- and education-matched neurotypicals in story context, findings of unique eye movement patterns in higher-level (story) reading could imply possible reading strategies that either the people with aphasia or the collective sample engaged in during the experiment. This raises questions about how reading material (sentence, story) may alter the effect of predictability and surprisal during reading in both neurotypical adults and people with aphasia given multi-sentence stories. For instance, researchers may investigate if PR is modulated by the availability of subsequent contextual information in stories that does not exist in single-sentence stimuli.

Another important takeaway from this research is the comparable effect of predictability on several eye tracking measures (PR, GD, TFD) for neurotypical adults and people with aphasia during story reading. This adds to evidence that people with aphasia respond to predictability similarly to neurotypical adults, and points to a relative strength of people with aphasia that could be leveraged in future reading interventions. Example applications of this work include the use of computational tools to manipulate word predictability in reading treatment materials, or the training of context reading as a compensatory reading strategy for people with aphasia.

Taken together, this research sets a foundation necessary to center assessments and interventions on well-evidenced predictability effects. This evidence will help shape more tailored intervention plans through future investigations of predictability effects on sentence and story reading in diverse individuals with aphasia (e.g., people mild versus severe reading deficits) and in relation to different clinical outcomes (e.g., eye movement versus reading comprehension).

Appendix A

A Cat Story

Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the neighborhood for months without success. When winter came, Sam's uncle surprised him and brought him a new kitten. Sam lovingly held his new friend. Then, he looked up and saw Jack sitting outside on the windowsill, eagerly looking inside.

Appendix Figure 1. Sample Story Drafted by Knollman-Porter et al. (2022)

Appendix B

Story:	Escaping the Storm							
Fixation #	Fixation word	Word length plus preceding space	Letter/space number	Fixation duration	Word number	Text line	Fixation number	Fixation number w/o interruption
1		=LEN(V10)+1						
2								
3								
4								
5								
6								
7								
8								
9								
10								

Appendix Figure 2. Sample Template for Raw Data Transcription

Appendix C

Appendix Table 1. Standardized Testing Results for Participants with Aphasia

Participant	WAB-R *Aphasia Quotient (AQ)* (max=100)	CAT *Comprehension of Spoken Paragraphs Subtest Score* (max=4)	RCBA-2 *Paragraph—Factual Subtest Score* (max=10)
A	37.30	2	9
B	91.10	3	10
C	61.10	2	6
D	71.50	4	10
E	73.50	4	10
F	71.40	3	9
G	60.47	3	8
H	82.20	4	9
I	97.80	4	10
MEAN	*71.82*	*3.44*	*9.00*
SD	*18.00*	*0.83*	*1.32*

Key: WAB-R= Western Aphasia Battery—Revised, CAT= Comprehensive Aphasia Test (CAT), RCBA-2=Reading Comprehension Battery for Aphasia—Second Edition

Appendix D

Appendix Table 2. Microsoft Excel© Algorithms for Eye Tracking Measures

Variable	**Process for Excel Computation**		
Given fixation order (Row A), fixation duration (Row E), word number in story (Row F)…			
	First	Then (if applicable)	Then (if applicable)
Probability of First Pass Fixation (PF)	=IF(AND((F_{Row}>F_{Row-1}),(F_{Row}>(MAX ($F\$_{Row}$: F_{Row-1})))),"YES","NO") *Computed in Row I*	Use replace function to replace all "YES" values with "1", and all "NO" values with "0" *Identified in Row M*	Pivot table organized for PF of each fixated word using "minimum value" directive. Table copy created to manual insert 0 values for words not fixated in passage
Probability of First Pass Regression (PR)	=IF(F_{Row+1}>=F_{Row},0, IF(J_{Row}>0,1,0)) *Computed in Row L*	Pivot table organized for PR of each fixated word using "minimum value" directive.	-
Gaze Duration (GD)	=IF(I_{Row}="YES", E_{Row-1},IF(AND(F_{Row}=F_{Row-1}, J_{Row-1} >0),J_{Row-1}+E_{Row},0)) *Computed in Row J*	=IF(J_{Row}=0,0,IF(AND(I_{Row}="YES",F_{Row}<>$F_{Row\,1}$, F_{Row}<> F_{Row+1}),J_{Row}, IF(AND(J_{Row+1}>J_{Row}, I_{Row+1}="NO"),"NA", J_{Row}))) *Computed in Row K*	Pivot table organized for GD of each fixated word using "sum" directive of Row J values.
Total Fixation Duration (TFD)	Use pivot table to compute sum of all fixation duration recordings per word number	-	-

Note: Excel algorithms were developed and refined in collaboration with Dr. Michael Walsh Dickey and Dr. Tessa Warren

Appendix E

Appendix Table 3. R Scripts for Linear Mixed Effect Models

Outcome Variable	Model
Probability of First Pass Fixation (PF)	glmer(PF ~ 1 + Group*Surprisal + Group*WrdLngth + CosDist + (1+Suprisal|Person) + (1|Story), data = FinalDataStructure, family= binomial)
Probability of First Pass Regression (PR)	glmer(PR ~ 1 + Group*Surprisal + Surprisal + (1|Person) + (1|Story), data = FinalDataStructure.filtered2, family = binomial)
Gaze Duration (GD)	lmer(log_gd ~ 1 + Group*Surprisal + CosDist + WrdFrq + (1 | Person) + ## (1 | Story))
Total Fixation Duration (TFD)	lmer(log_tfd ~ Group*Suprisal + CosDist + WrdFrq + (1 | Person) + ## (1 | Story)

Key: glmer = generalized linear mixed effect model, lmer = linear mixed effect model, WrdLngth = word length, CosDist = target-prediction cosine distance, WrdFrq = subtitle word frequency

Note: PF and PR models were scripted with support of Candace van der Stelt, and GD and TFD models were scripted with support of Alex Swiderski

Appendix F

Appendix Table 4. Sample List of Target Nouns and Preceding Story Context

Context	Target
A Cat	Story
A Cat Story. Sam enjoyed spending	time
A Cat Story. Sam enjoyed spending time with Jack, his black	cat
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball	practice
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer	day
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the	neighborhood
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the neighborhood for	months
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the neighborhood for months without	success
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the neighborhood for months without success. When	winter
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the neighborhood for months without success. When winter came, Sam's	uncle
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer	kitten

day, Sam could not find Jack anywhere. He searched the neighborhood for months without success. When winter came, Sam's uncle surprised him and brought him a new	
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the neighborhood for months without success. When winter came, Sam's uncle surprised him and brought him a new kitten. Sam lovingly held his new	friend
A Cat Story. Sam enjoyed spending time with Jack, his black cat. After arriving home from basketball practice on a beautiful summer day, Sam could not find Jack anywhere. He searched the neighborhood for months without success. When winter came, Sam's uncle surprised him and brought him a new kitten. Sam lovingly held his new friend. Then, he looked up and saw Jack sitting outside on the	windowsill

Bibliography

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018). Noun and verb processing in aphasia: Behavioural profiles and neural correlates. *NeuroImage: Clinical*, *18*, 215–230. https://doi.org/10.1016/j.nicl.2018.01.023

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye Movements of Highly Skilled and Average Readers: Differential Effects of Frequency and Predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390. https://doi.org/10.1016/0010-0285(85)90013-1

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bianchi, B., Bengolea Monzón, G., Ferrer, L., Fernández Slezak, D., Shalom, D. E., & Kamienkowski, J. E. (2020). Human and computer estimations of Predictability of words in written language. *Scientific Reports*, *10*(1), 4396. https://doi.org/10.1038/s41598-020-61353-z

Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1162–1172. https://doi.org/10.1037/0096-1523.25.4.1162

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1). https://doi.org/10.16910/jemr.2.1.1

Brennan, A., Worrall, L., & McKenna, K. (2005). The relationship between specific features of aphasia-friendly written material and comprehension of written material for people with aphasia: An exploratory study. *Aphasiology*, *19*(8), 693–711. https://doi.org/10.1080/02687030444000958

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157–158*, 81–94. https://doi.org/10.1016/j.bandl.2016.04.008

Brookshire, C. E., Wilson, J. P., Nadeau, S. E., Gonzalez Rothi, L. J., & Kendall, D. L. (2014). Frequency, nature, and predictors of alexia in a convenience sample of individuals with chronic aphasia. *Aphasiology*, *28*(12), 1464–1480. https://doi.org/10.1080/02687038.2014.945389

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive Processes in Eye*

Guidance (pp. 53–78). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198566816.003.0003

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Calvo, M. G., & Meseguer, E. (2002). Eye Movements and Processing Stages in Reading: Relative Contribution of Visual, Lexical, and Contextual Factors. *The Spanish Journal of Psychology*, *5*(1), 66–77. https://doi.org/10.1017/S1138741600005849

Clifton Jr., C. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*(3), 317–334. https://doi.org/10.1016/S0749-596X(03)00070-6

Clifton Jr, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye movements*, 341-371.

Cook, A. E., & Wei, W. (2019). What Can Eye Movements Tell Us about Higher Level Comprehension? *Vision*, *3*(3), 45. https://doi.org/10.3390/vision3030045

DeDe, G. (2013). Reading and Listening in People With Aphasia: Effects of Syntactic Complexity. *American Journal of Speech-Language Pathology*, *22*(4), 579–590. https://doi.org/10.1044/1058-0360(2013/12-0111)

DeDe, G. (2014). Sentence Comprehension in Older Adults: Evidence for Risky Processing Strategies. *Experimental Aging Research*, *40*(4), 436–454. https://doi.org/10.1080/0361073X.2014.926775

DeDe, G. (2017). Effects of Lexical Variables on Silent Reading Comprehension in Individuals With Aphasia: Evidence From Eye Tracking. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2589–2602. https://doi.org/10.1044/2017_JSLHR-L-16-0045

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Education. Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Flesch, R. (1979). How to write plain English. University of Canterbury. https://pages.stern.nyu.edu/~wstarbuc/Writing/Flesch.htm

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Heller, D. (1982). Eye movements in reading. In R. Groner & P. Fraisse (Eds.), *Cognition and eye movements* (pp. 139–154).

Hanne, S., Burchert, F., De Bleser, R., & Vasishth, S. (2015). Sentence comprehension and morphological cues in aphasia: What eye-tracking reveals about integration and prediction. *Journal of Neurolinguistics*, *34*, 83–111. https://doi.org/10.1016/j.jneuroling.2014.12.003

Hayes, R. A., Dickey, M. W., & Warren, T. (2016). Looking for a Location: Dissociated Effects of Event-Related Plausibility and Verb–Argument Information on Predictive Processing in Aphasia. *American Journal of Speech-Language Pathology*, *25*(4S). https://doi.org/10.1044/2016_AJSLP-15-0145

Huck, A., Thompson, R. L., Cruice, M., & Marshall, J. (2017). Effects of word frequency and contextual predictability on sentence reading in aphasia: An eye movement analysis. *Aphasiology*, *31*(11), 1307–1332. https://doi.org/10.1080/02687038.2017.1278741

Hux, K., Wallace, S. E., Brown, J. A., & Knollman-Porter, K. (2021). Perceptions of people with aphasia about supporting reading with text-to-speech technology: A convergent mixed methods study. *Journal of Communication Disorders*, *91*, 106098. https://doi.org/10.1016/j.jcomdis.2021.106098

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, *23*(5), 612–624. https://doi.org/10.1016/S0022-5371(84)90382-7

Kertesz, A. (2006). *Western Aphasia Battery–Revised*. Pearson

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Knilans, J., & DeDe, G. (2015). Online Sentence Reading in People With Aphasia: Evidence From Eye Tracking. *American Journal of Speech-Language Pathology*, *24*(4). https://doi.org/10.1044/2015_AJSLP-14-0140

Knollman-Porter, K., Bevelhimer, A., Hux, K., Wallace, S. E., Hughes, M. R., & Brown, J. A. (2023). Eye Fixation Behaviors and Processing Time of People With Aphasia and Neurotypical Adults When Reading Narratives With and Without Text-to-Speech Support. *Journal of Speech, Language, and Hearing Research*, *66*(1), 276–295. https://doi.org/10.1044/2022_JSLHR-22-00298

Knollman-Porter, K., Brown, J. A., Hux, K., Wallace, S. E., & Crittenden, A. (2022). Reading Comprehension and Processing Time When People With Aphasia Use Text-to-Speech

Technology With Personalized Supports and Features. *American Journal of Speech-Language Pathology*, *31*(1), 342–358. https://doi.org/10.1044/2021_AJSLP-21-00182

Knollman-Porter, K., Wallace, S. E., Hux, K., Brown, J., & Long, C. (2015). Reading experiences and use of supports by people with chronic aphasia. *Aphasiology*, *29*(12), 1448–1472. https://doi.org/10.1080/02687038.2015.1041093

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *Quarterly Journal of Experimental Psychology*, *63*(9), 1838–1857. https://doi.org/10.1080/17470211003602412

LaPointe, L. L., & Horner, J. (1998). *Reading Comprehension Battery for Aphasia–Second Edition (RCBA-2)*. Pro-Ed.

Lowder, M. W., Choi, W., Ferreira, F., & Henderson, J. M. (2018). Lexical Predictability During Natural Reading: Effects of Surprisal and Entropy Reduction. *Cognitive Science*, *42*(S4), 1166–1183. https://doi.org/10.1111/cogs.12597

Mack, J. E., Ji, W., & Thompson, C. K. (2013). Effects of verb meaning on lexical integration in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics*, *26*(6), 619–636. https://doi.org/10.1016/j.jneuroling.2013.04.002

Mack, J. E., & Thompson, C. K. (2017). Recovery of Online Sentence Processing in Aphasia: Eye Movement Changes Resulting From Treatment of Underlying Forms. *Journal of Speech, Language, and Hearing Research*, *60*(5), 1299–1315. https://doi.org/10.1044/2016_JSLHR-L-16-0108

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. http://arxiv.org/abs/1301.3781

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S.,Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H.(2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Parr, S. (1995). Everyday reading and writing in aphasia: Role change and the influence of pre-morbid literacy practice. *Aphasiology*, *9*(3), 223-238. https://doi.org/10.1080/02687039508248197

Parr, S. (2007). Living with severe aphasia: Tracking social exclusion. *Aphasiology*, *21*(1), 98-123.

Purdy, M., Coppens, P., Madden, E. B., Mozeiko, J., Patterson, J., Wallace, S. E., & Freed, D. (2019). Reading comprehension treatment in aphasia: A systematic review. *Aphasiology*, *33*(6), 629–651. https://doi.org/10.1080/02687038.2018.1482405

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–688. https://doi.org/10.1007/s00426-008-0173-3

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (n.d.). *Language Models are Unsupervised Multitask Learners*.

Raney, G. E., Campbell, S. J., & Bovee, J. C. (2014). Using eye movements to evaluate the cognitive processes involved in text comprehension. *Journal of visualized experiments : JoVE*, (83), e50780. https://doi.org/10.3791/50780

Rayner, K. (n.d.). Eye Movements in Reading and Information Processing: 20 Years of Research. *EYE MOVEMENTS IN READING*.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The Effects of Frequency and Predictability on Eye Fixations in Reading: Implications for the E-Z Reader Model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732. https://doi.org/10.1037/0096-1523.30.4.720

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448. https://doi.org/10.1037/0882-7974.21.3.448

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 514–528. https://doi.org/10.1037/a0020990

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. https://doi.org/10.3758/BF03214555

Rose. (n.d.). *The effectiveness of aphasia-friendly principles for printed health education materials for people with aphasia following stroke*.

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, *118*(45), e2105646118. https://doi.org/10.1073/pnas.2105646118

Schustack, W., Ehrlich, S. E., & Rayner, K. (n.d.). *Local and Global Sources of Contextual Facilitation*.

Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. P. (2022). *Large-Scale Evidence for Logarithmic Effects of Word Predictability on Reading Time* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/4hyna

Sharma, S., Kim, H., Harris, H., Haberstroh, A., Wright, H. H., & Rothermich, K. (2021). Eye Tracking Measures for Studying Language Comprehension Deficits in Aphasia: A Systematic Search and Scoping Review. *Journal of Speech, Language, and Hearing Research*, *64*(3), 1008–1022. https://doi.org/10.1044/2020_JSLHR-20-00287

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation: Predictability and Eye Movements. *Language and Linguistics Compass*, *9*(8), 311–327. https://doi.org/10.1111/lnc3.12151

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test*. Psychology Press.

Taylor, W.L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30,* 415–433.

Tinker, M. A. (1958). Recent studies of eye movements in reading. *Psychological Bulletin*, *55*(4), 215–231. https://doi.org/10.1037/h0041228

Vitu, F. (1991). The influence of parafoveal preprocessing and linguistic context on the optimal landing position effect. *Perception & Psychophysics*, *50*(1), 58–75. https://doi.org/10.3758/BF03212205

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E–Z Reader. *Cognition*, *111*(1), 132–137. https://doi.org/10.1016/j.cognition.2008.12.011

Warren, T., Dickey, M. W., & Lei, C.-M. (2016). Structural prediction in aphasia: Evidence from either. *Journal of Neurolinguistics*, *39*, 38–48. https://doi.org/10.1016/j.jneuroling.2016.01.001

Webb, W. G., & Love, R. J. (1983). Reading Problems in Chronic Aphasia. *Journal of Speech and Hearing Disorders*, *48*(2), 164–171. https://doi.org/10.1044/jshd.4802.164

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., … Rush, A. M. (2020). *HuggingFace's Transformers: State-of-the-art Natural Language Processing* (arXiv:1910.03771). arXiv. http://arxiv.org/abs/1910.03771