ABSTRACT

Title of Dissertation:	REAL-TIME DISCOURSE COMPREHENSION ABILITIES IN HEALTHY ADULTS AND ADULTS WITH TRAUMATIC BRAIN INJURY
	Kelly Marshall, Doctor of Philosophy, 2026
Dissertation directed by:	Dr. Jared Novick, Hearing and Speech Sciences Dr. L. Robert Slevc, Psychology

Much of human social interaction, knowledge transmission, and cooperative action takes place through conversation. Although engaging in conversation is a common occurrence, the ability to understand conversation structure is a complex skill supported by dynamic interactions of linguistic and non-linguistic cognitive processes. When any of these processes break down, the social and occupational costs can be profound. This is often the case for individuals living with traumatic brain injury (TBI), as more than 75% of this population suffers from cognitive-communication deficits secondary to their brain injuries.

Gaps in knowledge about how best to facilitate rehabilitation with this population persist because the specific cognitive mechanisms underlying conversational discourse abilities are unknown, even in healthy adults. One major hurdle to identifying these mechanisms is that much of the study of discourse for healthy adults and patients with TBI has taken the form of descriptive discourse

analysis that considers conversations as a whole. Because a key feature of conversation is that it dynamically unfolds over time, new approaches must consider how conversation structures (i.e., topics) emerge, change, and interact with ongoing language comprehension and production processes. This work examined (1) sensitivity to topic structure shifts during naturalistic language tasks in healthy adults versus individuals with TBI (2) whether specific cognitive processes that may support understanding of topic transitions differ between healthy and TBI groups and (3) what effect topic structure has on online language comprehension processes. We found that shifts in topic structure affect both online language comprehension and production. Both healthy adults and adults with TBI were similarly sensitive to topic structures as they emerge. However, individuals with TBI may have differences in lower-level comprehension and production that become apparent during discourse tasks when fine-grained temporal processes can be measured.

Taken together, the results of these studies can both inform theories of healthy discourse processing and narrow the scope of possible reasons for conversational breakdown following brain injury, ultimately driving new directions in clinical research for cognitive-communication disorders.

REAL-TIME DISCOURSE COMPREHENSION ABILITIES IN HEALTHY ADULTS AND ADULTS WITH TRAUMATIC BRAIN INJURY

by

Kelly Marshall

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2026

Advisory Committee:
Dr. Jared Novick, Professor, Chair
Dr. L. Robert Slevc, Associate Professor, Co-Chair
Dr. Stefanie Kuchinsky, Research Associate Professor
Dr. Yasmeen Faroqi-Shah, Professor
Dr. Ellen Lau, Associate Professor

By Kelly Marshall
2026

Dedication

This work is dedicated to the clients and research participants living with cognitive-communication challenges who have shared their struggles and hopes with me throughout the years. Thank you for a great conversation.

Acknowledgements

I want to sincerely thank my advisors, Jared Novick and Bob Slevc, for their mentorship and support throughout this dissertation.

I also want to thank the members of the Audiology and Speech Pathology Center and the National Intrepid Center of Excellence at Walter Reed National Military Medical Center. I want to give special thanks to Stefanie Kuchinsky her mentorship, statistical guidance, and effort in accommodating many administrative challenges to completing this work. I also want to give special thanks to Kate Sullivan for her clinical expertise and clinical research mentorship.

Thanks to all of the research assistants, Gabriela, Heather, Sofia, Tanvi, Kimberly, Harry, James, Gurnoor, Anna, and Chinomso, and lab members, Valerie, Zach, Mine, Tal, Zoe, Anna, Rachel, Ren, and Yi, who have supported this work. I also want to thank my Honors/REACH mentees, Avery Vess and Erica Wu, for your interest in and work on these projects. Special thanks to Zoe Ovans, Valerie Langlois, Erica Wu, and Harry Mayilsamy for providing or assisting with stimuli for these projects.

I would like to give heartfelt thanks to everyone who participated in these studies. I would especially like to thank the service members and veterans at Walter Reed National Military Medical Center for volunteering their time and for their sincere commitment to improving our understanding of a complicated set of symptoms to enhance rehabilitation for anyone who finds themselves in a similar situation.

Finally, I would like to thank my family Mary, Peter, and Matthew Sharer, and my husband Andrew Marshall for their support.

Foreword

The content included in Chapter 2 of this dissertation was published in *Language, Cognition, and Neuroscience* by Marshall, Vess, Novick, and Slevc (DOI:10.1080/23273798.2026.2637577). The student was first author of this paper due to her substantial contributions to the work (conceptualization, data processing, analysis, writing). The committee approves of the use of this jointly authored work.

Disclaimer: This material is based upon work supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Maryland, College Park and/or any agency or entity of the United States Government/or WRNMMC.

UNIVERSITY OF
MARYLAND
DEPARTMENT OF HEARING AND SPEECH SCIENCES
PROGRAM IN NEUROSCIENCE AND COGNITIVE SCIENCE

0100 LeFrak Hall
College Park, Maryland 02742
Email: jnovick1@umd.edu
Phone: (301) 405-1288

March 19, 2026

Dear Dr. Roth,

As Co-Chairs of Kelly Marshall's Dissertation Committee in Neuroscience and Cognitive Science (NACS), Bob Slevc and I are writing to seek your approval for the inclusion of her published work as part of her final dissertation, as per the guidelines set forth by The Graduate School.

Kelly has made significant contributions to the field, including an in-press manuscript that aligns closely with the dissertation's focus. Although she collaborated on the paper with us, Kelly was not only substantially involved in this joint work, but she also spearheaded it. We have certified this in a letter to the Director of the NACS Program and to the Director of Graduate Studies, who both approved its inclusion. Kelly will submit the current letter to you along with her dissertation, as per the guidelines. This work has been appropriately cited within the dissertation, indicating where it was previously published.

Bob and I approved the inclusion of this paper in the dissertation, and we are requesting your endorsement for this inclusion as well.

Please let us know if you need any else from us. Thank you for your consideration.

Sincerely,

Jared Novick, PhD
Professor
Department of Hearing and Speech Sciences
Program in Neuroscience and Cognitive Science
University of Maryland

L. Robert Slevc, PhD
Associate Professor
Department of Psychology
Program in Neuroscience and Cognitive Science
University of Maryland

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1. TBI – Traumatic Brain Injury
2. NBI – Non-Brain Injured
3. PTSS – Post-traumatic Stress Symptoms
4. PTSD – Post-traumatic Stress Disorder
5. NSI – Neurobehavioral Symptom Inventory
6. SSS – Standford Sleepiness Survey
7. TBI-QOL – Traumatic Brain Injury Quality of Life survey
8. UMD – University of Maryland
9. WRNMMC – Walter Reed National Military Medical Center

Chapter 1: Introduction

Motivation

A large proportion of day-to-day communication takes place during conversation. Most healthy individuals can plan upcoming projects with co-workers, catch up with friends, or discuss a medical diagnosis with their doctor during conversational exchanges with ease. However, this is not always the case for individuals who have sustained traumatic brain injuries (TBI)—when a blow to or penetration of the head impairs brain structure or function. In fact, a hallmark of communication deficits after TBI is difficulty participating in these social and occupational roles due to cognitive-communication impairments (Coelho, 2007; Hill et al., 2018; Lê et al., 2022; Levin et al., 2014; Salas et al., 2018; Togher et al., 2023).

Many of the consequences of TBI are chronic and, for a large proportion of individuals who sustain a TBI (about 43%), they contribute to long-term disability (Levin et al., 2014; Selassie et al., 2008). Two major contributors to post-injury quality of life are returning to work or school and social integration. Only about half of individuals with TBI return to their pre-injury work, and similarly, about half report an inability to maintain important social relationships. Both occupational and social consequences are highly associated with an individual's cognitive-communication competence, or their ability to dynamically apply a range of linguistic abilities in the appropriate contexts to achieve goals based on the interplay of cognitive (e.g., attention, memory, executive function) and linguistic knowledge (Bond & Godfrey, 1997; Levin et al., 2014; S. MacDonald, 2017; Snow et al., 1997; Togher et al., 2023). Social and occupational outcomes also depend on individuals' subjective experience of effort and fatigue related to compensating for cognitive-communication changes (Johansson et al., 2009; Salas et al., 2018; Turkstra et al., 2025). Given that more than 75% of individuals with TBI experience cognitive-communication changes following their injuries, understanding cognitive-communication disorders is

paramount in reducing disability and increasing life participation after injury (S. MacDonald, 2017; Togher et al., 2023).

Crucial to understanding cognitive-communication competence is the fact that complex communication events, such as conversations, involve the dynamic interplay of various levels of linguistic information (pragmatics, semantics, syntax, discourse structure), moderated by non-linguistic cognitive abilities (memory, prediction generation and updating, attention) (S. MacDonald, 2017). These interactions evolve over time as the event unfolds, making the temporal dynamics of communication events a key area of interest in understanding communication difficulties following TBI (Turkstra et al., 2006). In conversation, it is especially important that these processes are coordinated quickly, as linguistic information is only available for a limited time and appropriate responses must be generated in a timely manner to keep the conversation going (Meyer, 2023).

However, understanding the dynamics of conversational discourse represents a complex empirical challenge. The complexity that defines unstructured conversational discourse is difficult to capture in controlled experiments. This difficulty is particularly pronounced when considering representations of conversational discourse structure, or topics, which are emergent properties of the context, the language input, and the respective knowledge of each interlocutor (Brennan et al., 2018; van Dijk, 1998). Therefore, there is limited evidence about what specific cognitive factors contribute to moment-to-moment discourse structure representations, such as memory for discourse context or expectations about continuing or changing topics. Conversely, there is also limited evidence about how discourse topics influence moment-to-moment language comprehension and production, such as adjusting subsequent attention or language production towards topic relevant and away from topic irrelevant information. These open questions about typical discourse processing limit knowledge about how or why conversational discourse abilities may break down as a result of brain injury (Lê et al., 2022).

The fact that conversation takes place over time also means that individuals must employ strategies to avoid missing key information and effectively manage limited cognitive resources and levels

of arousal. Thus, studying discourse comprehension in real-time is also complicated by the need to account for potential differences in processing dynamics that may lead to increased effort or fatigue for individuals with TBI. In other populations, such as individuals with hearing impairment, studies of real-time language comprehension have revealed differences in processing strategies or in cognitive effort, even when behavioral performance is indistinguishable from healthy controls (McMurray et al., 2017; Winn & Teece, 2021, 2022). Recent hypotheses suggest that efficient cognitive resource allocation is key to reducing persistent cognitive symptoms after TBI (Turkstra et al., 2025). Thus, better understanding of the dynamics of discourse structure processing—not only in regards to particular cognitive processes per se, but also in regards to processing strategies and cognitive effort—is needed to characterize conversation performance following TBI.

Taken together, these factors make it difficult to explain why conversational deficits emerge. The lack of a mechanistic understanding of conversation precludes the development of speech-language therapies that target relevant discourse processes (Lê et al., 2022). More broadly, it also leads to confusion and distress in individuals with TBI and their families, who want to understand the reason for cognitive-communication changes in the same way they may understand how the brain controls memory or motor processes (Grayson et al., 2021). Therefore, the goals of this dissertation are to provide insight into mechanisms of discourse comprehension by examining discourse structure processing in real-time in healthy individuals and individuals with TBI and by exploring differences between these groups in abilities that may contribute to dynamic topic structure processing deficits in TBI. In the following sections, I will review what is currently known about conversation topic representations in healthy adults and individuals with TBI. I will describe the methods that have been used in the literature thus far, and describe what is missing from current approaches that has led to gaps in knowledge about real-time discourse structure comprehension. Then, I will describe current open questions about real-time conversation structure comprehension and explain how some available theories might inform ways to address these questions.

Typical Conversation and Conversation Following TBI

Topics in Conversational Discourse

Discourse describes language forms such as narrative, exposition, and conversation that have structural organization above the level of words and sentences. Thus, understanding discourse structure requires foundational language abilities such as semantic and syntactic processes, as well as remembering prior information from the discourse, tracking high-level organization and coherence with the structure, and integrating incoming information with knowledge of the discourse structure and content, social knowledge, and world knowledge (Coelho et al., 2002, 2012; Lê et al., 2022; Leaman & Edmonds, 2021; Levin et al., 2014; Mason & Just, 2006; Snow et al., 1997). This complex interplay of linguistic and non-linguistic abilities explains why discourse structure comprehension is a cognitive-communication ability. Different discourse types have distinct higher-level structural organization, and the particular cognitive demands of understanding a discourse are likely to depend on the discourse type.

There are a variety of perspectives on the organizing principles for conversational discourse. However, most accounts agree that the basis of conversation structure are topics — what conversation partners are talking about and what any contributions during the conversational turn should be related to. There may also be subtopics related to larger topics. Therefore, conversation can be said to have hierarchical structure (Asher & Vieu, 2005; Brinton & Fujiki, 1984; Hobbs, 1985, 1990; Van Kuppevelt, 1995; Roberts, 2012; Tyler, 2014; van Dijk, 1979). Being “off topic” or “on a tangent” occurs when speakers apply coherence (referring to the same idea) locally instead of globally, letting one utterance lead to another linearly without referring back to the overarching topic at the higher levels of the hierarchy (Hobbs, 1985).

While the idea of topic is intuitive on the surface, discrepancies arise regarding the nature of the representational form that a topic or subtopic takes. For example, topics may be propositions (an event or state and arguments that apply to the event or state; e.g. “The weather is hot.”), where subordinate information refers to or elaborates on some component of the proposition (Hobbs, 1985; van Dijk, 1979).

Topics have also been described as explicit or implicit questions under discussion that subordinate information must provide part of an answer to (e.g., What is the weather like?) (Van Kuppevelt, 1995; Roberts, 2012). Further, Centering Theory characterizes topics as "focuses of attention" that relate to links between information in the current utterance and salient information in previous utterances (Grosz et al., 1995; Taboada & Zabala, 2008; Walker, 1998). Some theorists have noted that many descriptions of topics are unsatisfactory, so the concept can be better understood by examining topic transitions in conversation versus defining the particular format of topics (G. Brown & Yule, 1983).

Given these discrepancies, the current work takes a generally agnostic approach to the representational format of topics. Instead, the primary focus is on *changes* in topic structure and the effects of these changes on ongoing language processing. Regardless of the representational format for topics, it should be the case that the effect of being "on a topic" or having a particular topic "active" is that certain subsets of information (e.g., particular memories or facts, particular lexical items) are more active in the minds/brains of speakers relative to others because only some types of information can be applied as subordinate under the topic or relate sufficiently to previous focuses of the discourse. Therefore, when topics change these relative activations should also change.

In contrast to narrative and expository discourse, where the structure is completely controlled by the speaker, conversation involves communication partners who can impose changes to the topic structure. Often topics will undergo smooth transitions, where utterances connect elements related to the first topic to the new topic, often referred to as topic drift, or topic shading. This drift can occur based on implicit relationships, or cohesion, between topics with similar properties (e.g., talking about the movie E.T. and then about Star Wars because they are both science fiction movies about space) or via explicit explanation (e.g., bringing up a memory and explicitly explaining how the previous conversation jogged that memory; "Speaking of E.T., that reminds me of Star Wars") (Gardner, 1987; Hobbs, 1990; Leaman & Edmonds, 2020). Conversation partners may also use explicit connectives or bids for news as discontinuity devices (e.g., "anyway…", "now tell me about when…") to initiate new topics that are not

specifically related to the previous topic (Leaman & Edmonds, 2020). Previous topics can also be returned to throughout the conversation using similar strategies (Gardner, 1987). Across all types of topic change, interlocutors must update representations of the current topic, which changes the set of potentially relevant information. These externally imposed, moment-to-moment changes present a unique computational challenge for engaging in conversation relative to other types of discourse.

Topic Related Abilities Following TBI

In contrast to the more focal lesions caused by strokes, brain tumors, and some diseases, traumatic brain injuries are often associated with diffuse damage as a result of diffuse axonal injury, inflammation, and changes in cerebral blood flow. In particular, damage to axons can disrupt connectivity between brain regions, affecting processes that require coordination of many cognitive systems (Hayes et al., 2016). Further, the acceleration-deceleration forces applied to the brain during traumatic injuries disproportionally affect frontal lobe regions associated with managing complex tasks via executive functions (Levin et al., 2014; Stuss, 2011). Taken together, these injury characteristics are associated with difficulties across a variety of tasks that require managing complex informational structures efficiently due to difficulties attending to relevant information, remembering prior context, applying top-down predictions, identifying information structure changes, and shifting to new cognitive sets (Bamdad et al., 2003; Barlow et al., 2018; Gilmore et al., 2018; Kavé et al., 2011a; Macdonald & Johnson, 2005; Ord et al., 2010; Perianez et al., 2007; Solbakk et al., 2002; Zakzanis et al., 2011).

Within conversational discourse, the effects of changes to brain structures and functions following TBI manifests as a characteristic set of changes to conversation abilities. This includes difficulties managing information (e.g., supplying too much or too little), providing off-topic or tangential information, and difficulty determining how to continue the conversation beyond direct responses to the conversation partner (Bond & Godfrey, 1997; Coelho et al., 1993, 2002; Hill et al., 2018; Mentis & Prutting, 1991; Snow et al., 1997). Because of these characteristic changes, individuals with TBI and their conversation partners report reduced conversational success. Conversation partners also report negative

perceptions about conversing with individuals with TBI, such as reduced desire to participate in conversations again and frustration at lack of flexibility, reduced initiation, and/or verbosity (Bond & Godfrey, 1997; Coelho et al., 2002; Grayson et al., 2021; Togher et al., 2010). Individuals with TBI also note that difficulty with remembering and managing conversation content and difficulty dealing with misunderstandings with and frustration from conversation partners lead to social withdrawal (Salas et al., 2018).

While the previous literature on this population is largely descriptive, some inferences can be made about why these patterns could emerge. Notably, the latter two of the main deficit areas may relate to the topic structure of conversation. Providing tangential information suggests that the speaker has not established a cohesive tie between the current utterance and the global topic structure of the conversation. This could be due to failure to construct or maintain a global structure, an issue with cohesion even if the global structure exists, or difficulty accessing information that would be coherent with the current topic because of discrepancies in what subset of information should be relatively more active. A failure to provide new information in the conversation could have a variety of causes, such as interest, attention, or pragmatic understanding of the goals of the conversation. However, as related to topic structure, it may be the case that without a strong representation of the current topic, it could be unclear to the speaker what kind of new information would be appropriate to add. If no sets of topic-relevant information are relatively more active than others because of poor global topic representation, speakers may rely on providing responses to the only level available to them, which is the local level of the conversation partner's previous utterance. Taken together, this suggests that better understanding of the cognitive mechanisms related to topic structure and the ability to move from topic to topic may be particularly important for addressing persistent conversation difficulties in people with TBI.

Methods for Studying Topic in Conversational Discourse

Discourse Analysis Approaches

The literature examining topics in conversational discourse has taken two primary approaches. One approach is to analyze various components of naturalistic, unstructured conversations post-hoc. These approaches generally fall under the broad scope of discourse analysis, which focuses on studying language use in natural contexts to gain insight into sociological, psychological, and linguistic processes (Brennan et al., 2018; G. Brown & Yule, 1983). For example, many of the works discussing the representational format of topics as propositions versus questions under discussion take examples of real conversations and use logical and linguistic principles to draw conclusions about why those conversations take the form that they do (Hobbs, 1985, 1990; Van Kuppevelt, 1995; van Dijk, 1979). Some work also applies formal systems to examples of conversation in order to reason about how topics relate to the actual utterances in the conversation (Grosz et al., 1995; Roberts, 2012). The field of computer science has also been interested in analysis of naturalistic discourse with respect to extracting main topics via semantic similarity modeling for applications such as improving conversational agents, information categorization, and search engine optimization (Boyd-Graber et al., 2017; Yeh et al., 2016). These approaches have been good for addressing questions like *What are topics?* and *How could the information structure of topics relate to the information structure of individual utterances?* However, the reliance on analysis by individuals who were not part of the conversation and the availability of all of the information in the discourse means it is difficult to consider what individual participants in the conversation represent about the topic structure at any given time as the conversation unfolds.

Discourse analysis has also been used to assess how abilities related to conversational discourse structure may change in neuropsychological populations, such as individuals with acquired neurological damage like stroke or traumatic brain injury (Azios et al., 2022; Coelho et al., 1993, 2002; Leaman & Edmonds, 2020, 2021), or individuals with psychiatric conditions such as schizophrenia (Boudewyn et al., 2012; Gernsbacher et al., 1999). While these populations may differ in several ways from the TBI

population, general principles across neuropsychological populations can provide insights about methods for studying conversation. These studies have addressed difficulties in topic maintenance, such as production of off-topic comments and the coherence between individual utterances and current topics, topic transitioning mechanisms, and over- or under-reliance on conversation partners to continue or introduce new topics. Analysis typically requires trained raters, who were not part of the original conversation, to assess the overall context of the conversation, decide what topics are being discussed and when, and rate or identify instances of non-coherence or topic shifting mechanisms. One benefit of using discourse analysis approaches with neuropsychological populations is that identifying differences between healthy and patient populations can provide a basis for making some inferences about what mechanisms could lead to the behavioral presentation differences during conversation, particularly when information is known about what cognitive processes have been altered by a particular biological factor (e.g., differences in neurotransmission, white matter connectivity, or lesion location). In other words, this approach has been useful for answering questions like *How does conversation present differently when particular cognitive or linguistic systems are impaired?*

However, this approach is observational in nature and requires additional studies to evaluate the proposed mechanisms. Boudewyn et al. (2012) note that some work has emerged aiming to address possible mechanisms, at least in the schizophrenia literature. However, this work has primarily addressed word- and sentence-level processes and may not fully reflect the demands of naturalistic discourse. Discourse analysis approaches in neuropsychological populations also have the same limitations as those with healthy adults. Methods that rely on third-party raters often report acceptable inter-rater reliability, suggesting that raters are able to pick up on some signal in the available information from the discourse (Azios et al., 2022; Coelho et al., 2002; Leaman & Edmonds, 2020). However, it cannot be known with these methods whether the topics identified by the raters are the same topics that were represented by individuals who were actually conversing, either due to differences in judgements because all of the conversation content is available at once or differences between raters and participants in world

knowledge or experiences that affect segmentation of utterances into larger topic units. For the same reasons, it is also unclear whether the utterances produced are coherent to the "actual" (i.e., real-time) topic representations. Some researchers have tried to account for this by using rating scales administered to individuals who were engaging in the conversation (Togher et al., 2010). While this eliminates the issues associated with third-party raters and is beneficial for understanding patient and communication partner experiences, it still assesses conversation after-the-fact and relies on subjective interpretations.

Experimental Approaches

In contrast to discourse analysis approaches that draw conclusions from conversational exchanges as they happen naturally, many other studies have explored language comprehension and production during conversational discourse by manipulating features of the linguistic or environmental context (Brennan et al., 2018). Many experimental approaches come from the field of psycholinguistics, which is interested in the dynamic interaction of language knowledge from prior experience, linguistic rules and constraints, and cognitive processes such as memory, attention, and conflict resolution that allow language input to be understood or produced by individual speakers or listeners. Interest in the influence of language context beyond isolated words or sentences led to several lines of research that examine the effect of discourse context on incremental word and sentence processing. For example, this has included interest in the constraints on incremental word and sentence comprehension posed by the physical context of real-world tasks (Brown-Schmidt et al., 2015; Brown-Schmidt & Tanenhaus, 2008; Tanenhaus et al., 1995) and by information in the previous discourse (Arnold, 2016; Arnold et al., 2004; Hagoort et al., 2009; Nieuwland & Van Berkum, 2006). While this kind of work has taken important steps to consider the role of communication contexts on language processing abilities in real-time using behavioral, eye tracking, and electroencephalography (EEG) techniques, it has typically used constrained discourse contexts (e.g., two sentences, short narrative vignettes) and has not addressed ongoing naturalistic conversation per se, including all of the complexity inherent to building topic structure together with other conversation partners.

Studies that are focused on understanding conversation or dialogue per se have focused on how interlocutors establish and use common ground—knowledge of what ideas and environmental factors in the current context are shared or not shared between interlocutors (Brown-Schmidt, 2012; Brown-Schmidt et al., 2015; Brown-Schmidt & Hanna, 2011; Clark, 2021)—or audience design—how individuals might change language production to account for the knowledge or expectations of listeners (Brown-Schmidt et al., 2015; Clark & Murphy, 1982). Many of these studies manipulate the information in the environment available to participants and examine the effects on subsequent language production. While these factors are essential to understanding language processes during conversation, they focus on mental representations about conversation partner knowledge, rather than directly addressing topic structure or dynamic changes in topic structure. A representation of common ground can help determine how to talk about a particular idea or thing if it is relevant to discuss at the moment, but it cannot determine whether talking about that idea or thing is appropriate given the current topic. There has been some work examining topics in written expository discourse in real-time using eye tracking. While this approach is promising, it has not yet been extended to conversation or spoken discourse (Lorch et al., 1985). Taken together, experimental approaches have been able to answer questions about the effect of discourse, in general, on real-time language processing, but have not necessarily addressed understanding of conversation topic changes over time.

Implications for the Current Work

Taken together, current approaches to studying conversation topic representations have left key gaps in knowledge. Although discourse analysis approaches have approached questions directly related to topic structure and are able to capture naturalistic interactions in their full complexity, they are limited by their observational nature and reliance on post-hoc analysis. Conversely, while experimental approaches could be beneficial to understanding mechanisms of discourse structure because of their ability to examine language comprehension in real-time and ability to manipulate specific factors, these approaches have not yet been used to explore dynamic changes in topic structure representations. The current work

aims to take several approaches to mitigate these limitations. First, it will use methods from both discourse analysis and experimental disciplines to balance the strengths and limitations of each approach. Second, it will introduce novel methods that allow a slightly different focus than previous work in both approaches. Discourse analysis work will use rating procedures that better capture information flow over time, while experimental work will aim to introduce more naturalistic stimuli that capture changes in topic focus over time. These approaches will be able to better address key open questions:

1. Do individuals in the conversation represent topic structures in real-time, given the constraints of memory, attention, and real-time language processing?
2. What processes contribute to updating topic structure representations?
3. Are these processes different in quality or time course due to particular brain injury characteristics?

Models That Inform Hypotheses About Real-time Topic Processing

Bottom-up Discourse Comprehension Models

While there are currently gaps in empirical knowledge about real-time construction, updating, and use of topic structures during conversation, there have been several theoretical proposals that provide a starting point for hypotheses about how these processes may occur. Perhaps most explicitly related is Structure Building Framework (Gernsbacher, 1991, 1997). This framework addresses discourse in general, without regard to the specific genre (i.e., narrative, conversation, exposition). Language comprehension during discourse is posited to involve three main processes: laying a foundation, mapping new information onto the framework, and shifting to initiate new substructures. At the beginning of a discourse event, the earliest available information is used to initiate a discourse structure representation, or lay a foundation. Evidence that individuals spend longer reading or looking at the first sentence or picture element of a written or cartoon story and that they have a memory advantage for the first

mentioned or viewed information suggest that the first available information determines the initial discourse structure (Gernsbacher, 1997).

As the discourse continues to unfold, incoming information is evaluated according to its coherence with the initial foundation. Coherence is determined by referential links (i.e., referring to the same concept again, anaphora), temporal or location consistency with the foundational concept, or a causal relationship between the incoming information and the foundational concept. Concepts or ideas that are coherent should activate similar knowledge (of the previous discourse or world knowledge) to the foundational concept, while concepts or ideas that are less coherent will activate different information (i.e., knowledge representations, concepts, memories) that is not already active because of the foundation. Memory benefits for information that shares a coherent link through conceptual consistency, anaphora, or causal relationship relative to information that does not share such a link has provided support for the mapping stage of the Structure Building Framework (Gernsbacher, 1997).

At some threshold of non-coherence, or mismatch between the representations currently active because of the foundation and the representations activated by incoming information, comprehenders decide that the incoming information is not similar enough to be integrated into the current discourse structure. This triggers the final step: shifting. Comprehenders create a new discourse substructure, which involves inhibiting representations associated with the previous discourse and increasing activation of representations that are associated with the incoming new information. Studies using both linguistic and non-linguistic (i.e., picture/cartoon) narratives have demonstrated worse memory for information presented prior to a non-coherent boundary than for information presented following the boundary during the most current discourse focus. Further, bridging inferences are less accurate when the two pieces of information required for the inference are across a coherence boundary than when they are within the same coherent structure. Additionally, reading or looking dwell times have been shown to be slower when non-coherent information is expected or actually occurs at narrative or pictorial event boundaries (Gernsbacher, 1997; Hard et al., 2011a). Taken together, these results suggest that encountering event

boundaries requires some costly cognitive process, perhaps the active inhibition of previous discourse information (Gernsbacher, 1997).

While a significant amount of evidence is consistent with this framework, the empirical work addressing structure building processes has primarily addressed short narratives (i.e., a few connected sentences). It often relies on after-the-fact memory performance or eye tracking measures of written discourse or non-linguistic actions as dependent measures. As described above, monologic discourse may not capture specific interactive qualities of conversation that contribute to structure building requirements, and these types of measures cannot fully describe topic representations in real-time. Thus, Structure Building Framework is taken as a useful proposal for how conversational discourse structure might emerge, but empirical support regarding conversation, specifically, is still needed.

Structure Building Framework emerged alongside a similar proposal, the Construction-Integration model, which aimed to better account for online processing (van Dijk & Kintsch, 1983). This model also proposes that information from the discourse activates a set of associated memory nodes from an individual's knowledge base. As new information unfolds in the discourse, it either leads to similar patterns of activation and the new information is integrated into the current discourse model, or different patterns of activation emerge and inhibit previous sets to create a new discourse understanding. What is considered relevant to the discourse is determined by spreading activation to knowledge related to the initial input (Kintsch, 1988). While the Construction-Integration model was focused on smaller units of discourse (a few sentences) versus narratives or sets of events, its focus on how bottom-up information can drive the creation of coherent discourse units provides an important complement to Structure Building Framework in explaining how discourse understanding might emerge during a continuously unfolding language event.

However, others have argued against strict bottom-up accounts of discourse structure organization, citing issues with applying the principles beyond adjacent sentences/utterances in the discourse (Giora, 1996). Instead, it is proposed that coherence operates between individual utterances and

topics/themes, rather than between adjacent utterances. It also proposes that informativeness about the topic/theme, rather than local coherence, guides what is appropriate to include in a subsequent utterance. At the extreme of Structure Building Framework proposals, discourses could be completely locally coherent but have no global structure, similar to "stream of consciousness". Written discourses often include topics or themes relatively explicitly (e.g., in an initial topic sentence). However, it remains unclear how the top-down topic structures in conversational discourse might emerge, particularly in unstructured conversations where there may be no specific a priori goal and where the content of the contribution from a conversation partner is often unknown. However, these counterarguments raise an important point about the need to consider the scope of comparison between previous information and currently unfolding information. In other words, is current information compared only to immediately preceding information, or to some emergent discourse understanding from many previous utterances? This critique also raises the need to consider additional means of indicating structure shifts that do not rely on degrees of match or mismatch between previous and current knowledge activations, such as explicit transition markers (e.g., "however", "anyway").

Event Cognition Models

A key proposal of Structure Building Framework and of models of cognitive-communication competence is that discourse structure building processes are mediated by domain general cognitive mechanisms (memory, attention, inhibition/set shifting) that contribute to organizing information that unfolds over time. There is a rich literature examining how individuals perceive events and how event structure influences memory and attention processes in turn (Zacks & Tversky, 2001). Under the assumption that non-linguistic events (e.g., cooking dinner) and linguistic events (e.g., conversation) are subserved by similar mechanisms, understanding the principles of event cognition can also inform what kinds of processes are also relevant to examine in conversational discourse.

Although events exist in the minds of observers rather than directly in the world, individuals are able to relatively consistently determine when events start and end and how they relate to each other.

When asked to identify boundaries between events, individuals tend to agree on the segmentation of unfolding actions in the world into particular events (Zacks & Tversky, 2001). Various proposals have been put forward about what elements signal a change in event, such as changes in setting, tempo, object, direction of motion, or participants in an activity or the points at which there is maximal physical change along some dimension (Newtson et al., 1977; Zacks & Tversky, 2001). Additionally, events can combine to form hierarchical structures. For example, individuals have relatively good agreement about what sub-events make up daily activities (Bower et al., 1979). Neuroimaging evidence from movie viewing paradigms also suggests that as information unfolds over time, different levels of event detail are processed at different time scales across different brain regions and this hierarchical representation is similar across participants (Baldassano et al., 2017).

Zacks et al. (2007) proposed a series of processes that allow for this event segmentation ability. Individuals continuously make predictions about upcoming perceptual information based on their model of the current event (perceptual information from the recent past bound into an "event model") and knowledge of event schema from long term memory. They compare the predicted information to what actually unfolds, and when the prediction error is large enough, this signals the need to segment the perceptual information into a new event. Instantiating a new event requires eliminating the bound information in the current event model and replacing it with information about the new incoming stimuli. It also involves temporarily increasing attention to incoming sensory input to better inform the new event model, with a commensurate decrease in attention to internal processes such as long term memory retrieval and prediction generation. The temporary increase in incoming sensory information means that the event model undergoes a quick change in content, but once this temporary increase is over, the event model remains more stable (Zacks et al., 2007). More recent evidence suggests that changes in internal states and changes in perceived goals and intentions, in addition to error signals arising only from low-level perceptual information, can also drive event segmentation (Nolden et al., 2024; Su & Swallow, 2024).

Consistent with this model, across a variety of studies it has been shown that when asked about what will happen next, participants are more accurate within events than across event boundaries. Additionally, memory for details near event boundaries is better than for details that occur in the middle of events, perhaps due to increased attention to the incoming perceptual stimuli (Nolden et al., 2024). Other studies have demonstrated that participants are sensitive to events with different kinds of event boundaries. When event boundaries are predictable, participants have difficulty identifying distracting stimuli that occur near expected boundaries. This suggests that participants allocate their attention to expected boundaries in real-time, such that predictive boundary processing interferes with distractor detection (Ji & Papafragou, 2022).

Integrating Discourse and Event Structure Building Theories

Although arising from different fields of study, Structure Building Framework and models of event segmentation address questions of how individuals might be able to impose higher-order structures on information that unfolds continuously in time and what processes are involved in this segmentation and ordering. Both types of models aim to account for how a continuous stream of bottom-up information can be related or integrated with information that has come before. Both propose that error signals that arise from comparison between some features of incoming information and previous information are key mechanisms for triggering updating of the current discourse structure or event model. And both include a mechanism for downregulating (inhibiting or removing) information from previous events in order to upregulate (increase activation or maintain) information that relates to the current event component.

The two types of models do differ in regards to how error is conceptualized—lack of overlap with semantic or associative networks in Structure Building Framework versus degree of difference from a specific prediction in event models. Event models also provide more specific predictions about how attentional control is affected by perceived changes in event structure. However, taken together, an understanding of these models can provide a basis for how to conceptualize real-time conversational discourse comprehension processes.

Implications for Studying Topic Changes

Taken together, previous evidence supporting Structure Building Framework and event models demonstrates that individuals can impose structure on continuous events in real-time. Like monologic discourse and continuous action sequences, dialogues in conversation involve a continuous stream of bottom-up information that individuals seem to be able to segment into meaningful higher-order structures (i.e., topics). Therefore, it is reasonable to think individuals could impose topic structure on continuously unfolding language input during real-time conversation processing. However, there is additional complexity that arises from social interactions and the dynamic interplay of multiple interlocutors who can all influence how a conversation event will unfold. It could be the case that topic changes are difficult or impossible to predict, given high variability in conversation partners' language and social behavior as well as differences in what interlocutors know about each other's knowledge or communication style (Dunbar et al., 1997; Hakulinen, 2009; Tagliamonte, 2012). Therefore, there may be no benefit to topic structure building in real-time or, if topic structure building relies on active prediction and identification of prediction error like event models, it may not then be possible to engage conflict detection and updating mechanisms. Instead, individuals might wait to construct topic representations until enough information is available as an "offline" memory consolidation process. Thus, real-time topic structure building in conversation remains to be explored empirically.

If topic structure building does occur in real-time, evidence from studies addressing Structure Building Framework and event segmentation models suggest that relevant processes would include maintaining information about previous information from the topic, predicting when topic changes are likely to occur and, subsequently, downregulating irrelevant information and upregulating relevant information for the current topic. In conversation it is also important to consider that individuals have dual roles as speakers and listeners. Therefore, real-time language comprehension and production processes must both be considered.

If these are the relevant processes, this can suggest what factors may be different following TBI and may thus contribute to conversation deficits. Individuals with TBI may have an inability to maintain previous information from the conversation context and use it to guide predictions about whether incoming information is going to be coherent (the same topic) or incoherent (a new topic). This would preclude making comparisons with new incoming information in order to trigger structural boundaries. If conversation structure processing breaks down at this level, individuals would not be able to generate topic structures, or not be able to generate them efficiently, and would not be able to use them to guide subsequent language production or comprehension. If previous information could be maintained, there may still be issues with responding to mismatch between previous and incoming information. Thus, topic boundaries may not be applied consistently, leading to differences in the understanding of the topic structure between conversation partners and potentially to comments that appear to be "off topic" or lingering on a previous topic. In addition to identifying conflict per se, individuals may struggle to use knowledge about features of the language input that signal when the topic is likely to change. Then, updating topic structure representations may occur slower than expected, or may not be able to proceed if new information continues to unfold in the conversation, leading to similar issues of incongruent topic representations between partners. Further, even if mismatch identification could proceed efficiently, it could still be the case that there are problems related to downregulating information from the previous topic and upregulating information relevant to the new topic. The failure to downregulate information from a previous topic could lead to contributions to the conversation that are no longer relevant. It could also lead to unnecessary competition between representations or distracting representations if old representations interfere with representations relevant to incoming information about the current topic, potentially affecting online language comprehension.

Alternatively, or in addition to deficits in particular component cognitive processes, speed or effort factors may affect conversation ability after TBI. All of these component processes may be intact, but deficits in information processing speed may not allow the processes to be deployed efficiently

enough to keep up with the rapid flow of incoming linguistic input (Levin et al., 2014). Or, individuals with TBI may have a more limited reserve of cognitive resources that affect allocation to processing topic structure changes (i.e., differences in effort) or that are more likely to be depleted and lead to fatigue over the course of the conversation (Peach & Hanna, 2021; Pichora-Fuller et al., 2016; Turkstra et al., 2025).

In addition to suggesting important elements of how real-time conversation processing might occur and break down subsequent to injury, this previous work provides insights into what kinds of evidence are needed. The previous literature first identified behavioral indices that demonstrated incremental or real-time comprehension, like consistent boundary segmentation across individuals. Once it is clear that individuals are sensitive to topic structures in real-time, experiments that manipulate particular features of the conversation, such as the type of prior discourse information or cues to upcoming topic shifts, can examine the relevant maintenance, boundary identification, upregulation or downregulation processes, or changes in attentional control. Using techniques like eye tracking during continuous conversations can provide key information about both mechanisms and the time course of these relevant processes.

Hypotheses

Based on the current gaps in knowledge and insights from models of structuring continuous temporal events, the current work will address the following hypotheses:

1. Individuals track discourse topics in real-time. Updating topic representations requires cognitive/neural resources that interact with concurrent or immediately subsequent language comprehension or production processes.
2. Individuals with TBI exhibit differences in abilities related to real-time topic structure tracking relative to healthy adults. The current work will test several possible loci of breakdown in the topic structure building process: a) maintaining information about previous discourse content, b) using previous discourse information to predict when incoming language input is consistent or inconsistent with a new topic, c) updating the model of what is

currently relevant in the discourse when new information occurs and/or, d) making inferences about when new topics are likely to occur based on cues from conversation partners.

3. Once established, current topic representations guide/constrain ongoing language comprehension and production processes.

Significance

Theoretical Significance

Humans impose structures on continuously unfolding information over time for a variety of different kinds of stimuli —written or spoken narratives, movies, natural action sequences, and even music (Gernsbacher, 1997; Ji & Papafragou, 2022; Krumhansl & Kessler, 1982; Sankaran et al., 2020; Zacks et al., 2007). Although the types of structures imposed differ across different kinds of information and across different task demands, deriving structure from continuous incoming information seems to share some common computational demands, such as maintaining prior information, comparing prior information to incoming information to detect some degree of mismatch or conflict, and when the level of conflict is great enough, creating a new higher order structural element. Conversation constitutes another type of continuously unfolding event, where the parameters of its real-time comprehension are yet to be explored. Conversation also differs from these other domains in important ways that make it an interesting case to study, as conversation structure is dependent on moment-to-moment changes from multiple sources as conversation participants all contribute to the unfolding input. Thus, understanding real-time topic processing in conversation not only broadens the current understanding of similar events, but may reveal how other cognitive processes may interact with real-time structure building to accommodate the added complexity imposed by multiple conversation partners. Further, because conversation is one of the most common modalities of communication and facilitates not only the transfer of ideas but also important social connections, understanding the mechanisms of its important components, like topic structure building, addresses an ability integral to the human experience.

Clinical Significance

Because the ability to participate in conversation is so important to maintaining the connections that foster healthy social interaction and participation in professions or the community, improving treatments that address cognitive-communication impairments in conversational discourse is a key objective for speech-language pathologists. By examining the cognitive processes relevant to real-time topic structure building in healthy adults, the current work provides a foundation for determining which of those processes may be related to conversational difficulties due to traumatic brain injury. Current treatment recommendations focus on providing conversation practice in individualized, realistic scenarios, often with frequent communication partners (Lê et al., 2022). However, without knowledge of which elements of discourse structure processing may be impaired, it is difficult for clinicians to know how to structure or modify conversation practice to address the underlying issues that lead to conversation difficulty. For example, if clinicians knew that individuals with TBI had difficulty tracking prior information from the discourse, they may modify the memory load or amount of preceding information before any given topic shift. If individuals with TBI instead only had difficulty with generating expectations about upcoming topic changes, clinicians may guide attention to relevant cues from the conversation partner, but not need to change anything about the content of the conversation practice. If the problem were instead that individuals with TBI required more time to downregulate topic-irrelevant and upregulate topic-relevant representations, clinicians may train conversation partners to pause and provide more time specifically around topic changes. While these are not the only possibilities for treatment interventions, they demonstrate that a better understanding of what specific topic-related processes are affected in individuals with TBI can lead to differentiated treatment approaches. Further, the application of research methods that address real-time language processing to studying conversation structure provides an important first step not only towards understanding how clinicians might intervene

to address processes as they unfold, but also how moment to moment effort due to processing changes may affect fatigue and eventual self-isolation.

It should be noted that the current work is several steps away from providing evidence about treatment implementation. Heterogeneity in the TBI population means that the processes affected in one individual may be different from another individual (Covington & Duff, 2021). Additional work will be needed to be able to identify deficit patterns in any given individual. However, the current work can help the field move towards understanding at the individual patient level by constraining hypotheses about possible types of topic-related deficits by investigating each process at the group level and by demonstrating the possibility of applying new methods to study individuals with TBI. Further, while the current proposal addresses TBI, insights regarding the nature of topic processing and possible breakdowns in topic-related abilities as well as research methods could be applied to other populations with cognitive-communication deficits, such as individuals with post-stroke aphasia, dementias, multiple sclerosis, schizophrenia, or other conditions.

Chapter 2: Topic Shift Costs in Naturalistic Discourse

Rationale

Although conversation is a routine part of daily life, navigating its structure requires the coordination of complex linguistic and non-linguistic cognitive processes (S. MacDonald, 2017). Chapter 1 outlined various proposals about the nature of topic representations and how interlocutors may impose higher-level discourse structures on incoming linguistic information. Across these proposals, two key components of topic structure in conversation are (1) that topics constrain the types of information that are appropriate to expect or include and (2) that these constraints are subject to frequent change. In contrast to narrative and expository discourse, where the structure is completely controlled by the speaker, conversation involves communication partners who can impose changes to the topic structure through a variety of explicit (e.g., "Speaking of …that reminds me of…") and implicit (e.g., through coherence between referents, time, or space) mechanisms (Gardner, 1987; Hobbs, 1990). Across all types of topic change, interlocutors must update representations of the current topic in order to understand what communication partners will say next and to plan an appropriate response.

Yet, the cognitive processes that allow interlocutors to use evolving discourse-level structures to guide their language comprehension and production remain underexplored. This gap is especially important given that difficulty maintaining or navigating topic structure is a hallmark of cognitive-communication disorders, such as those commonly observed in individuals with traumatic brain injury (TBI). In these cases, communication impairments are often attributed to deficits in attention, memory, or executive functioning—core cognitive systems that support the regulation, planning, and monitoring of language—rather than a loss of linguistic knowledge per se (Bond & Godfrey, 1997; Coelho et al., 2002; S. MacDonald, 2017; Mentis & Prutting, 1991; Snow et al., 1997). Investigating how topic structure processing breaks down in TBI not only advances our understanding of these cognitive-communication disorders but also sheds light on the cognitive mechanisms that support effective discourse in neurologically healthy individuals (Lê et al., 2022).

As an initial step towards these broader goals, this study examines how people respond to shifts in topic structure—specifically, whether such shifts influence ongoing language production and whether these effects differ between individuals with TBI and healthy adults.

Topic Shifts in Conversation

Previous work on conversational topic has largely focused on structural questions—such as how topics and subtopics are organized, how topic structure can be inferred from surface-level utterances, and what types of contributions are appropriate given the current topic structure (Asher & Vieu, 2005; Brinton & Fujiki, 1984; Grosz et al., 1995; Hobbs, 1985, 1990; Van Kuppevelt, 1995; Roberts, 2012; van Dijk, 1979). However, open questions remain about the cognitive processes that allow topic structures to be built, updated, and used during conversation. In order to begin to examine specific cognitive processes that support the ability to track conversation topics, an important initial question is whether and how we can detect cognitive demands that arise when individuals experience topic shifts.

In other words, while previous work has addressed *what* the structure of conversation topics looks like, we still know little about *how* individuals respond to changes in topic structures and the cognitive-linguistic consequences for discourse understanding and production. Addressing these questions is critical not only for developing more complete models of language processing that incorporate discourse-level organization, but also for understanding *why* communication abilities are vulnerable to disruption from cognitive deficits, such as those seen in individuals with TBI, even when core language representations remain intact.

Some research has begun to address related questions. Previous psycholinguistic studies have examined how language comprehension (Arnold et al., 2004; Arnold, 2016; Brown-Schmidt et al., 2015; Brown-Schmidt & Tanenhaus, 2008; Hagoort et al., 2009; Nieuwland & Van Berkum, 2006; Tanenhaus et al., 1995) and production (Brown-Schmidt, 2012; Brown-Schmidt et al., 2015; Brown-Schmidt & Hanna, 2011; Clark, 2021; Clark & Murphy, 1982) are shaped by discourse context. For example, discourse information that is shared between interlocutors (i.e., common ground) can constrain what

listeners consider as relevant or possible referents and cause speakers to adjust production to reflect what conversation partners mutually know (Arnold et al., 2004; Brown-Schmidt et al., 2015; Brown-Schmidt & Tanenhaus, 2008). However, much of this work has focused on brief exchanges involving a single topic, context, or task goal. To better understand language use in more dynamic and naturalistic settings, further research is needed on how processing adapts when the focus shifts within a broader, ongoing discourse—for example, in response to topic changes during extended conversation.

A Review of Models of Discourse Structure Building

As reviewed in Chapter 1, Structure Building Framework (Gernsbacher, 1991, 1997), offers a theoretical proposal for how interlocutors can generate higher-level structures over the course of an unfolding discourse. According to this framework, discourse comprehension involves three key processes: laying a foundation, mapping incoming information onto that foundation, and shifting to initiate new substructures when coherence is lost. At the start of a discourse event, comprehenders use the earliest available information to initiate a discourse structure representation, or lay a foundation. As new information is encountered, it is evaluated according to its coherence with this foundation. Coherence may arise through referential continuity, temporal or spatial consistency, or causal links with previously established content. When incoming information aligns with the foundation, it activates related knowledge and is integrated into the existing structure. In contrast, information that lacks coherence activates unrelated knowledge and fails to align with what is currently active in memory. When this mismatch reaches a certain threshold, comprehenders interpret the incoming content as unrelated to the current discourse structure. This triggers a shift, whereby the comprehender inhibits the previously active discourse substructure and constructs a new substructure for the incoming information.

Previous research investigating the Structure Building Framework has shown that listeners are sensitive to narrative discourse coherence boundaries and the shifting processes they trigger. For example, studies using both linguistic and non-linguistic (i.e., picture or cartoon) narratives have shown that memory is worse for information presented before a coherence-disrupting boundary (i.e., information

hypothesized to be related to the inhibited substructure) than for information presented after the boundary, which aligns with the current, active substructure (Anderson et al., 1983; Gernsbacher, 1997; Mandler & Goodman, 1982). Further, bridging inferences are less accurate when the two pieces of information required for the inference are across a boundary than when they are within the same coherent structure (Gernsbacher, 1997). Reading times are also slower when non-coherent information is expected or actually occurs at narrative or pictorial event boundaries (Gernsbacher, 1997; Hard et al., 2011b). Taken together, these results suggest that encountering discourse structure boundaries requires some costly cognitive process that affects subsequent comprehension, perhaps the active inhibition of previous discourse information (Gernsbacher, 1997).

Like narratives, conversations involve a continuous stream of information that has meaningful higher-order structures. Thus, individuals might impose topic structure on unfolding language input during real-time conversation, just as event boundaries appear to be imposed during real-time narratives. However, social interactions involve the dynamic interplay of multiple interlocutors, each influencing how the conversation will unfold. The inherent variable nature of social context, combined with individual differences in language use and social behavior, may make topic changes difficult to predict or detect reliably (Dunbar et al., 1997; Hakulinen, 2009; Tagliamonte, 2012). Therefore, although tracking topic structure shifts may offer benefits (e.g., constraining interpretation during comprehension or narrowing the range of possible responses during production), it may be impractical or even impossible to do so consistently given the variability of conversational contexts and communication partners. In this case, individuals may rely on "offline" memory consolidation to build topic representations after the fact. Thus, it remains an open question whether topic-structure building occurs in real-time during conversation and how such processes, if they occur, influence subsequent language comprehension and production.

Discourse Processing in Individuals with Traumatic Brain Injury

Across various types of discourse, individuals with TBI frequently show intact sentence-level organization and local coherence (micro-structure processes) but impaired organization at the broader discourse level (macro-structure processes). This pattern has led to a prevailing view that macro- and micro-structure processes are dissociable in discourse processing (both comprehension and production). In particular, macro-structural organization is thought to rely on domain-general cognitive abilities such as attention, working memory, and executive functions (e.g., planning and organization), while micro-structure organization relies more on other, domain-specific linguistic processes (Adornetti, 2014; Cosentino et al., 2013; Mozeiko et al., 2011).

Macro- and micro-structure processes, while operating on different levels of representation, may rely on some of the same capacity-limited, domain-general cognitive resources to construct or adjust those representations. Evidence from individuals with TBI suggests that micro-structural processes (e.g., within-sentence repetitions, retracings, and filled or unfilled pauses) are correlated with deficits in domain-general cognitive abilities like working memory and executive function, indicating that these cognitive systems may be involved in micro-structure production (Peach, 2013). A wide body of psycholinguistic research on healthy adults and those with other neurological disorders also supports the involvement of domain-general cognitive processes in sentence-level comprehension and production ((Boudewyn et al., 2025; Hsu et al., 2017, 2021; Hsu & Novick, 2016; Kim et al., 2025; Ness et al., 2024; Novick et al., 2009; Ovans, 2022; Sharer & Thothathiri, 2020; Slevc, 2011; Slevc & Martin, 2016; Thothathiri, 2018; Thothathiri et al., 2018). Further, in individuals with TBI, there is evidence of competition between macro- and micro-linguistic processes: devoting resources to discourse-level cohesion (e.g., producing more across-sentence ties, higher cohesion ratings) is associated with worse performance on sentence-level measures (Peach & Coelho, 2016; Peach & Hanna, 2021).

Taken together, these findings suggest that macro- and micro-structural production impose processing demands on at least some shared mechanisms (although which specific mechanisms is not yet

clear), which must be managed strategically to meet task demands. For example, depending on the task, it may be more beneficial to produce a coherent discourse or construct well-formed sentences. Given this, if individuals with TBI have difficulty appropriately allocating processing resources between macro- and micro-linguistic levels, then deficits may become apparent at either level, depending on the nature of the discourse task. This makes examining processing demands in more complex linguistic tasks with micro- and macro-linguistic demands, such as conversation, a beneficial method for examining sources of breakdown in cognitive-linguistic abilities.

The Current Study

Despite the theoretical proposals described above, it remains unclear whether tracking topic structures during conversation imposes cognitive demands, and whether the response to these demands differs in individuals who have difficulty with discourse processing. Key questions include whether topic shifts place measurable demands on comprehenders, and whether these demands influence ongoing language production in the next conversational turn. These questions are especially relevant for understanding communication challenges in individuals with TBI, who often struggle with discourse-level organization despite intact sentence-level language skills.

To investigate these issues, we sought a metric that could capture the cognitive cost associated with shifting from one topic to another—if such structure-building processes occur as the conversation unfolds. Our approach focuses on changes in language production following topic shifts as a behavioral window into the processing demands involved in managing discourse structure. If the occurrence of a topic shift is cognitively demanding, then the allocation of resources to accommodate topic shifting processes should reduce the availability of cognitive processes necessary for sentence-level planning until the topic shifting processes are complete. Here, we assess if responding to a shift in topic does indeed incur processing costs, and if so, to which aspects of language processing. We rely on natural conversational data from both healthy participants and individuals with TBI who typically have difficulty with conversational structure.

Hypotheses and Predictions

If recognizing a topic shift in conversation engages cognitive resources involved in both macro- and micro-level language processes, then those resources may be less available for subsequent word- and sentence-level production during the listener's next turn. Accordingly, we predict a cost to language production when cognitive processes are occupied with identifying and/or establishing a new topic in the discourse representation, relative to when the conversation continues on the same topic. In contrast, if tracking topic structures does not impose demands on cognitive resources that are shared with micro-linguistic processes, increased demands at the discourse level should not affect lower-level language production processes and no differences should emerge between responses to topic shifts and responses within the same topic. Similarly, if listeners do not actively construct topic-level representations, or if doing so is not particularly demanding, then language production should be unaffected by the presence or absence of a topic shift.

We hypothesize that the nature of topic shift costs can help pinpoint the cognitive processes involved in macro- and micro-linguistic production. We assessed several aspects of language output, including utterance length, semantic and syntactic complexity, self-monitoring (via repetitions and revisions), and planning (via filled pauses). These measures were chosen to reflect different stages of language production that may be affected by an interlocutor's topic shift.

From the perspective of the Structure Building Framework (Gernsbacher, 1997), shifting topics involves activating new semantic content and inhibiting prior content. While this process is ongoing, some no-longer-relevant content may remain active. This could lead to competition from residual information, making it harder to access relevant ideas immediately after a shift. As a result, semantic complexity may decrease, planning time may increase (reflected in filled pauses), and speakers may revise more frequently as they work toward coherent output. Similarly, the effort required to reorient to a new topic might reduce syntactic complexity—either directly or by diverting resources away from sentence planning—and may lead to more revision if syntactic errors occur. Alternatively, if semantic and

syntactic planning are delayed until discourse-level structure stabilizes, topic shifts may lead to general slowing in output without selective effects on linguistic form.

Difficulties with topic maintenance following TBI may stem from two possible sources: (1) decreased ability to track and represent topic structures or (2) intact tracking but increased difficulty coordinating the cognitive processes involved in handling topic shifts.

If individuals with TBI do not adequately represent topic structures, then topic shifts should impose little or no burden, and shift-related effects on subsequent language production should be minimal, resulting in smaller or no shift costs relative to healthy individuals. However, if individuals with TBI do track topic structures but struggle because of impaired cognitive processes, they should show larger shift costs than healthy individuals, reflecting increased processing demands.

Given the heterogeneity of TBI, it is also possible that individual differences in injury severity or cognitive-linguistic profile modulate these effects, potentially obscuring group-level patterns (Covington & Duff, 2021; S. MacDonald, 2017). Injury severity is one of many factors that can affect communication outcomes after TBI. Measures of injury severity are frequently clinically available, feasible to assess, and informative about neurological factors that may affect cognitive resource allocation. Here, we rely on a common operationalization of TBI severity, length of coma (LOC), which is associated with both the extent of neurological damage (Levin et al., 1988; Wilde et al., 2016) and likelihood of regaining complex functional abilities post-injury (Kowalski et al., 2021; Lucca et al., 2019; Malone et al., 2019). More extensive neurological damage (indexed by LOC) may be associated with a greater number of focal injuries or impaired neural signaling and network connectivity. These disruptions could interfere with the rapid coordination of cognitive abilities needed to track topic changes.

One possibility is that under one of the above hypotheses—impaired tracking or increased cognitive effort—there will be more pronounced effects for individuals with more severe brain injuries because of a greater extent of neurological damage. Alternatively, the relationship between topic shift costs and TBI severity may be negative or non-linear. For example, individuals with less severe TBIs may

be able to track the discourse structure but with increased cognitive effort, leading to increased topic shift effects relative to healthy individuals, whereas individuals with higher TBI severities may be less able to track the discourse structure and thus show diminished topic shift effects relative to healthy individuals.

Finally, it is possible that discourse deficits observed in TBI are not related to topic structure per se, but instead arise from other sequelae or comorbidities such as hearing loss, slower processing speed, reduced attention, impaired social inference, or mental/physical health challenges (e.g., depression, post-traumatic stress, pain) (S. MacDonald, 2017). In this case, we would expect general group differences in language production, but no interaction between topic status and group—i.e., no evidence of differential topic shift effects.

Method

Conversation samples were drawn from the Coelho database on TBI Bank, a publicly available corpus consisting of multiple discourse tasks performed by participants with TBI and participants with no history of brain injury (NBI) (Coelho et al., 2002; Elbourn et al., 2019) (Corpus DOI: 10.21415/T5Q01Z). A detailed description of the full data collection protocol can be found in Coelho et. al (2002), but relevant details for the current study are described below. We analyzed all samples in the corpus except for one sample from the NBI group (N24) due to poor audio quality and incomplete transcript and two samples from the TBI group due to having no conversation partner-initiated topics (TB02) and having no conversation portion in the sample (TB42).

Participants

Participants were native English-speaking adolescents and adults who passed vision and hearing screenings and reported no history of psychiatric illness or substance misuse. They were categorized into two groups. Participants in the NBI group (N = 49; 16 Female) had no history of brain injury or other neurological disease or injury. Participants in the TBI group (N = 52; 13 Female) sustained a moderate (loss of consciousness/length of coma duration less than six hours) or severe (loss of consciousness/length of coma duration greater than six hours) traumatic brain injury (Lezak & Lezak, 2004). Participants were

recruited from rehabilitation hospitals. NBI Group participants were individuals who worked in the hospitals where conversation samples were collected. TBI Group participants were patients in various programs in the same hospitals. Participants in the TBI group had no aphasia, dementia, or disorientation (Western Aphasia Battery Aphasia Quotient [AQ] greater than 93, Dementia Rating Scale Score 120 or above, Galveston Orientation and Amnesia Test score 75 or above), had no significant dysarthria as evaluated by a speech-language pathologist, and were rated as having a high level of brain injury recovery (Rancho Los Amigos Level VII (automatic-appropriate) or above). Participants in the TBI group sustained injuries that were primarily caused by motor vehicle accidents (N=42), with other causes including falls (N=6), blunt force injuries (N=3), and unknown (N=1). Some participants' radiological scans showed cerebral contusions, hemorrhages, subdural and subarachnoid hematomas, and swelling. Some patients presented with no neurological findings and information was not available for several participants (N=15). Demographic information for each group and injury information for the TBI group is shown in Table 1.

	NBI Group			**TBI Group**		
	Mean	*SD*	*Range*	*Mean*	*SD*	*Range*
Age (years)	32.2	13.6	16-63	29.1	12.6	16-69
Education (years)	14.1	3.0	11-22	13.1	2.4	9-21
Time Post Onset (months)	---	---	---	10.25	17.84	1-99
Length of Coma (days)	---	---	---	15.5	22.1	0-99

Table 1. Demographic and injury characteristics of participants in the non-brain injured (NBI) and traumatic brain injury (TBI) groups. Note: There are some discrepancies between participant demographic information posted on TBI Bank and demographic info information present in the transcript files. In cases of discrepancy, we used the information in the transcript file.

Participants signed informed consents for data collection and for release of the data through TBI Bank following local Institutional Review Board approved procedures and in accordance with the Declaration of Helsinki.

Conversation Sample Collection

Participants engaged in approximately 10-15 minutes of audio-recorded unstructured conversation. Samples were collected as part of a larger protocol also including a story retell task and story generation task, where task order varied across participants. Participants were told that the purpose of the conversation task was to learn more about conversational behavior. Conversations were unscripted and followed naturally from the responses of both parties. Their conversation partners were one of two speech-language pathologists, one male and one female, who had little to no previous interaction with the participants (Coelho et al., 2002).

Conversations were transcribed and formatted according to TalkBank Codes for Human Analysis of Transcripts (CHAT). Transcription and coding of relevant features (e.g., filled pauses, repetitions, retracings) were completed according to Computerized Language ANalysis (CLAN) conventions (MacWhinney, 2000). Transcription, utterance labeling, and feature coding were completed by Coelho and colleagues for distribution on TBI Bank. We made edits to the content of the CHAT files to correct errors that would affect analysis (e.g., incorrect speaker labels) but made minimal changes to the main content or coding decisions reflected in the TBI Bank files to facilitate reproducibility of results. A list of changes to the files is available in the Supplementary Information (available at https://osf.io/hvdpj/).

Coding Procedure

Coding procedures were completed by the author and a research assistant. For each conversation, the coder first listened to the conversation in full for a broad understanding of the general context. We followed conventions from Leaman and Edmonds (2020) to then identify broad topics and subtopics. Broad topics reflected major conceptual shifts, as determined from the general context. Transitions between subtopics were determined by identifying "topic-initiation mechanisms" (see Table 2), which reflect the various ways that interlocutors indicate a shift to a new topic within a conversation (Leaman & Edmonds, 2020). Coders then read through the transcript line by line. Audio files were consulted when spoken information was needed to disambiguate topic-initiating mechanisms (e.g., length of silences or

intonation). At each topic initiating mechanism, we inserted a comment line denoting the location of a subtopic shift and wrote a summary description of the subtopic content. For the purposes of this study, "shifts" refer to the point in the conversation where we identified a subtopic-initiating mechanism. While this definition does not allow us to comment on factors such as the global hierarchical structure of topics and subtopics, we used this method because it was both reliable (see below) and approximated the information available to participants in real-time as the conversation unfolded.

Topic-Initiation Mechanism	**Description**	**Example**
Topic-closing	A routine manner of ending a previous topic which may include a series of turns where no new information is added, wrap-up statements, or pauses greater than one second.	"Yeah, that does sound nice"
Discontinuity Device	Words or phrases used to explicitly indicate that an utterance will include a change of topic. Bids for information about a new topic or elicitation of "news".	"By the way…", "So…", "Now…"; "Tell me about…"
Cohesion	Responses that appeal to a lexical or referential relationship to the previous utterance and create a logical transition.	Partner A: "I went to the mall with my friends this weekend." Partner B: "I went to watch a movie at the theatre there Saturday." Partner A: "I love that movie theatre."
Non-coherent comment	A statement that has no logical relation to the previous utterance, but the conversational partner accepts it as a new topic.	Partner A: "And so I cooked spaghetti for dinner last night." Partner B: "I hate how it's so cold out" Partner A: "It's so freezing I wore three layers of clothes."

Table 2. How transitions between topics were determined for coding. Adapted from Leaman & Edmonds (2020).

As subtopic shifts were identified, each utterance in the following turns was coded according to the identity of the speaker (investigator/I–referring to the speech-language pathologist conversation partner—or participant/P) and the relationship to the subtopic status (initiating the subtopic shift, responding to a turn that initiated a subtopic shift, remaining on the same subtopic, or off topic; Table 3). The first turn after a subtopic-initiating mechanism was considered as a subtopic-initiating turn. The first subsequent turn from the other conversation partner was considered a response to a new subtopic. Every turn following the response to a new subtopic, until a new subtopic was initiated, was considered a turn on the same subtopic. Every utterance within the same turn received the same code, except for off-topic codes. Off-topic codes were assigned when the utterance was unrelated to the current subtopic, but subsequent utterances returned to the current subtopic. Utterances that were unrelated to the current subtopic, but which served as a non-coherent topic-initiating mechanism (because the conversation partner began talking about content related to the unrelated comment rather than returning to the current topic) were labelled as part of a topic-initiating turn. We also coded backchannels (e.g., yeah, uh-huh, right right), which served to prompt the speaker to continue but did not contain meaningful additions to the conversation. Backchannels did not constitute a turn, so utterances before and after a backchannel within the same turn would maintain the same code.

Code	Meaning
INEW	Investigator (partner) started a new subtopic
PNEW	Participant started a new subtopic
IPNEW	Investigator (partner) response to a participant-initiated new subtopic
PINEW	Participant response to an investigator (partner)-initiated new subtopic
IPSAME	Investigator (partner) response to a participant turn of the same subtopic
PISAME	Participant response to an investigator (partner) turn of the same subtopic
IOFF	Investigator (partner) is off-topic

POFF	Participant is off-topic
BACK	Backchannel

Table 3. Utterance codes used to label each line of the CHAT file. This study analyzed productions labeled PINEW, or Shift responses, and PISAME, or No Shift responses. Note that "investigator" refers to the speech-language pathologist conversation partner. "Investigator" is a frequently used speaker label in CLAN used to denote a conversation partner who is part of the research team (i.e., versus a familiar communication partner or family member of the participant).

Reliability

Coders trained on the above coding scheme by reviewing the written instructions and through verbal discussion before practicing on three randomly selected transcripts, until substantial reliability was achieved (accuracy 80% or greater and Cohen's Kappa 0.6 or greater; McHugh, 2012). We defined reliability at the level of individual line codes, as correctly identifying subtopic boundaries, speaker turns, off topic utterances, and backchannels should lead to the same line by line codes. Discrepancies were resolved by collaborative discussion between the two coders. Inter-rater reliability between the two coders was confirmed with an additional 10% of the files (5 TBI, 5 NBI). The average percent agreement was 88.4% and the average Cohen's Kappa was 0.7, indicating substantial reliability (McHugh, 2012).

Analysis

To test how a speaker's topic shifts affect a listener's subsequent language production, we analyzed participants' spoken responses under two conditions: continued discourse on the same subtopic versus discourse following a partner-initiated subtopic shift. CLAN includes tools to extract utterances by code. This allowed us to create separate files for utterances that were relevant to the research questions—participant responses to partner-initiated new subtopics ("Shift responses") and participant responses within the same subtopic ("No Shift responses"). Each participant's Shift and No Shift files were analyzed using the EVAL and FLUCALC commands in CLAN. These commands automatically compute a variety of language production measures.

Our goal was to capture how production is affected under these conditions by examining several distinct dimensions of language production, from general output to more fine-grained measures of linguistic complexity and speech performance monitoring. We focused on five dependent variables:

- General output was indexed by *words per utterance*
- Syntactic complexity was assessed using *verbs per utterance*, which reflects the number of clauses and therefore the structural complexity of each utterance.
- Semantic complexity was assessed using *idea density,* which corresponds to the number of propositions (verbs, adjectives, adverbs, conjunctions, and prepositions) per word. This measure reflects how much relational or conceptual information is involved in a speaker's message and has been linked to memory and comprehension (C. Brown et al., 2008; Kintsch, 1974; Snowdon, 1996).
- Error detection and speech monitoring and revision processes were measured using the *number of repetitions and retracings per syllable*[1] – instances where speakers repeated or corrected themselves mid-utterance (Levelt, 1983; Postma et al., 1990).
- Production planning difficulty was assessed using the *number of filled pauses per syllable*. Filled pauses (e.g., "uh", "um") have been associated both with speech planning difficulty and with shifts in event structure in other types of discourse (Clark & Fox Tree, 2002; Crible, 2018; Meyer, 2023; Swerts, 1998).

We used generalized linear mixed effects models (glmer functions in R; R version 4.1.2 (2021-11-01); R Core Team, 2021) to assess the effect of group (NBI, TBI) and subtopic status (Shift, No Shift) and their interaction on each dependent measure, with covariates of age and years of education, a random slope of subtopic status and random intercept of participant nested within groups. We used the maximal

[1] We used per syllable, rather than per utterance, for repetition/retracing and filled pauses consistent with general conventions for measuring disfluencies (Riley, 1972). Syllable-based measures should reflect disruptions to the flow of ongoing speech and be sensitive to the total amount of speech produced (whereas utterances can vary in length).

random effects structure that would converge for each model. If the maximal random effects structure would not converge, we removed the random slope of subtopic status. Because all dependent measures (except idea density, see below) are rates, we used Poisson distributions to model these variables. The dependent variable was the numerator of each rate, while the denominator was accounted for by including an offset term (the base-ten logarithm of the denominator) in the model (Dalgaard, 2008). For models using Poisson distributions, we applied quasilikelihood estimation to account for overdispersion (Bolker, 2024). For the idea density measure, CLAN outputs include the final calculation only (propositions/words) and do not separately provide information about the number of propositions. We therefore analyzed arcsine square root transformed idea density proportions in a linear mixed effects model with the same structure of covariates and fixed and random effects described above (Sokal & Rohlf, 2012).

To assess the effect of TBI severity on shift costs (for the TBI group only), we used the same model structures and procedures as described above, but replaced the fixed effect of group with length of coma (LOC). We operationalized TBI severity as LOC duration (i.e., rather than a combination of LOC duration, Glasgow Coma Scale Score, and Post-Traumatic Amnesia duration) because this was the only one of the three typical severity-related factors available in the corpus dataset. Six participants were excluded from this analysis because they did not have LOC data. Random intercepts were included by participant (i.e., no longer nested because only the TBI group was included). We analyzed TBI severity only for dependent variables for which there were significant shift effects in the group analysis.

For all models, we grand-mean centered continuous variables (Hox & Roberts, 2011) and used a significance level of alpha = 0.05.

Results

Coding Results

Table 4 shows descriptive statistics for each code applied by the raters. Although statistical analysis of all of the speech act classifications and their relative distributions is beyond the scope of this

work, these patterns demonstrate that participants in the healthy and TBI groups produced a similar proportion of the PINEW and PISAME utterances, which are critical to the primary analysis. The non-critical codes for backchannels, topic introductions, and off-topic utterances follow a similar numerical pattern as in the original analysis of Coelho et al. (2002). Research investigator partners added slightly more to ongoing topics and introduced new topics more when speaking to a participant with a TBI. Participants with TBI produced off-topic utterances more frequently than participants in the NBI group. This suggests that our coding scheme, while optimized for utterance-level analyses, aligns with the relevant discourse structure features previously explored in this corpus.

	TBI Group				**NBI Group**			
	INV		**PAR**		**INV**		**PAR**	
Code	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
INEW/PNEW	0.08	0.03	0.06	0.07	0.07	0.03	0.07	0.06
IPNEW/PINEW	0.02	0.01	0.16	0.08	0.02	0.02	0.16	0.06
IPSAME/PISAME	0.19	0.10	0.30	0.08	0.15	0.06	0.31	0.08
BACK	0.13	0.04	0.04	0.03	0.14	0.04	0.05	0.03
IOFF/POFF	0.00	0.00	0.03	0.04	0.00	0.00	0.01	0.02

Table 4. Descriptive statistics for each of the codes employed in this analysis. Values represent the average proportion of utterances given each code relative to the total number of utterances labeled in each file.

Primary Analysis

We observed differences between responses to subtopic shifts versus responses that addressed the same subtopic on four of the five dependent measures. Specifically, following a subtopic shift, participants produced shorter utterances (fewer words per utterance; Figure 1A; Tables 5A, 6), less complex syntax (fewer verbs per utterance; Figure 1B; Tables 5B, 6), and fewer revisions (fewer repetition and retracing events per syllable; Figure 1D; Tables 5D, 6). Participants produced more filled

pauses per syllable, indicating greater planning difficulty (Figure 1E; Tables 5E, 6). Subtopic status had no effect on idea density (Figure 1C; Table 5C).

Two dependent measures differed between the TBI and NBI groups: participants in the TBI group produced shorter utterances (fewer words per utterance; Figure 1A; Tables 5A, 7) and used less complex syntax (fewer verbs per utterance; Figure 1B; Tables 5B, 7). There were no group differences in idea density (Figure 1C; Tables 5C, 7), revisions (repetitions and retracings per syllable; Figure 1D; Tables 5D, 7), or filled pauses (per syllable; Figure 1E; Tables 5E, 7).

There were no significant interactions between subtopic status and group for the dependent measures (Figure 1A-1E; Table 5). That is, individuals with TBI did not differ from healthy adults in how their language production was affected by subtopic shifts. Both groups showed a similar sensitivity to subtopic status, with comparable downstream effects on their subsequent language use.

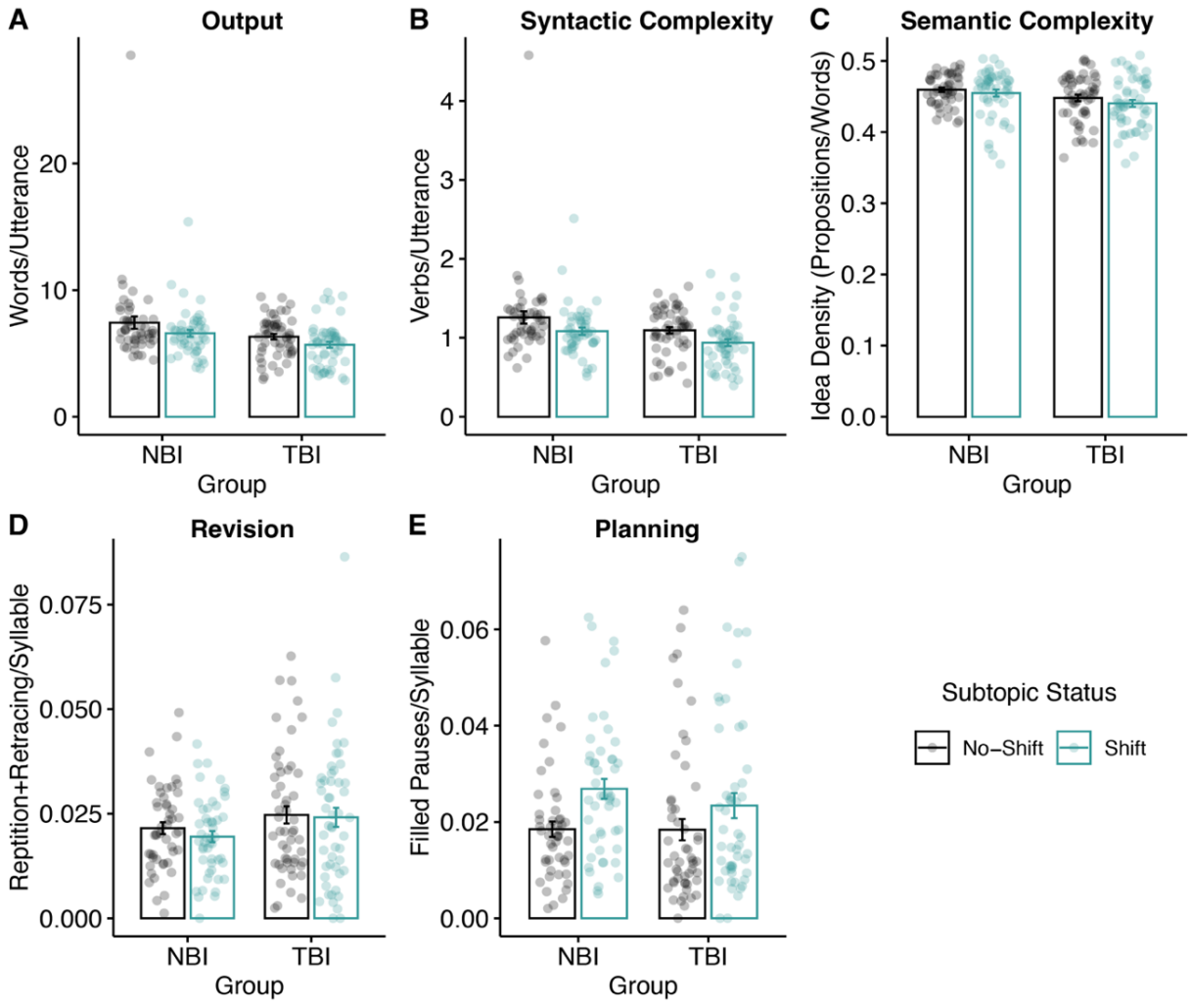


Figure 1. Linguistic outcome measures for turns remaining on the same subtopic (No Shift; black) and turns following a conversation partner-initiated subtopic shift (Shift; teal) for each group (NBI = non-brain injured healthy individuals; TBI = individuals with moderate-severe traumatic brain injury). Error bars show standard error. Points represent values for individual participants. A) Output measured by words per utterance. B) Syntactic complexity measured by verbs per utterance. C) Idea density measured by propositions per total words. D) Self-monitoring and revision measured by number of repetition events plus retracing events per syllable. E) Linguistic planning difficulty measured by number of filled pauses per syllable.

Dependent Measure	Term	Parameter Estimate	Standard Error	Z/t Statistic	P Value
A) Words per Utterance					
	Intercept	1.94	0.01	194.49	<0.001
	Group	-0.11	0.01	-7.84	<0.001
	New	-0.10	0.01	-10.91	<0.001
	Age	0.01	0.00	9.31	<0.001
	Education	0.02	0.00	6.37	<0.001
	Group*New	-0.02	0.01	-1.74	0.08
B) Verbs per Utterance					
	Intercept	0.16	0.02	7.12	<0.001
	Group	-0.09	0.03	-2.97	<0.01
	New	-0.13	0.02	-6.22	<0.001
	Age	0.01	0.00	4.33	<0.001
	Education	0.02	0.01	2.56	0.01
	Group*New	-0.04	0.03	-1.26	0.21
C) Idea Density					
	Intercept	0.74	0.01	56.29	<0.001
	Group	-0.01	0.02	-0.51	0.61
	New	-0.01	0.01	-0.86	0.39
	Age	0.00	0.00	0.83	0.41
	Education	0.00	0.00	1.55	0.12
	Group*New	-0.00	0.01	-0.40	0.69
D) Repetitions + Retracings per Syllable					
	Intercept	-3.92	0.07	-56.66	<0.001
	Group	0.08	0.10	0.78	0.43
	New	-0.11	0.05	-2.43	0.02
	Age	0.00	0.00	0.43	0.66

	Education	-0.01	0.02	-0.48	0.63
	Group*New	0.09	0.06	1.46	0.14
E) Filled Pauses per Syllable					
	Intercept	-4.15	0.08	-51.30	<0.001
	Group	-0.09	0.11	-0.76	0.45
	New	0.35	0.04	8.30	<0.001
	Age	0.00	0.00	0.51	0.61
	Education	0.04	0.02	1.71	0.09
	Group*New	-0.08	0.06	-1.23	0.22

Table 5. Model outputs for the effects of group and subtopic status on each dependent measure. Statistical models were of the form: glmer(Dependent Measure ~ Group*New + Age + Education + offset(log(Total_Utts)) + (1+New|Group/ID), family="poisson")where *Group* = factor representing TBI or NBI (reference level – NBI), *New* = factor representing subtopic status (1=Shift, 0=No shift, reference level = No shift), *Age* = age in years (centered), Education = years of education (centered), *ID* = participant, and *Total_Utts* = number of utterances. Specific models used for each dependent measure are available in Appendix 1.

	Subtopic Shift		**No Subtopic Shift**	
Measure	**Mean**	**SD**	**Mean**	**SD**
Words/Utterance	6.12	1.86	6.86	2.67
Verbs/Utterance	1.01	0.32	1.17	0.44
Idea Density	0.45	0.04	0.45	0.03
Repetitions+Retracings/Syllable	0.02	0.01	0.02	0.01
Filled Pauses/Syllable	0.03	0.02	0.02	0.01

Table 6. Means and standard deviations for each dependent measure by subtopic status, collapsed across groups.

	TBI		NBI	
Measure	**Mean**	**SD**	**Mean**	**SD**
Words/Utterance	6.01	1.67	7.01	2.78
Verbs/Utterance	1.02	0.31	1.17	0.45
Idea Density	0.44	0.03	0.46	0.03
Repetitions+Retracings/Syllable	0.02	0.02	0.02	0.01
Filled Pauses/Syllable	0.02	0.02	0.02	0.01

Table 7. Means and standard deviations for each dependent measure by group, collapsed across subtopic status.

Effect of Injury Severity Within the TBI Group

If conversational difficulties following TBI stem in part from impairments in managing the cognitive demands of topic shifts, then one might expect these difficulties—and the associated linguistic costs—to scale with injury severity. We therefore examined whether shift-related costs in language production varied as a function of LOC duration (only in the TBI group) for the four dependent measures that were sensitive to topic status in our main analysis: words per utterance, verbs per utterance, repetitions and retracings per syllable, and filled pauses.

Three of the four dependent measures were affected by TBI severity. Specifically, individuals with more severe brain injuries produced responses that had fewer words per utterance (Figure 2A; Table 8A), were less syntactically complex (fewer verbs per utterance; Figure 2B; Table 8B), and required less planning (fewer filled pauses per syllable; Figure 2D; Table 8D). There was no effect of severity on revision (repetitions and retracings per syllable; Figure 2C; Table 8C).

There was a significant interaction between TBI severity (LOC duration) and subtopic status on words per utterance (Figure 2A; Table 8A), verbs per utterance (Figure 2B; Table 8B), and filled pauses per syllable (Figure 2D; Table 8D). Across these measures, participants with more severe TBIs (longer

LOC durations) showed *smaller* shift effects. No such interaction was observed for repetitions and retracings per syllable (Figure 2C; Table 8C).

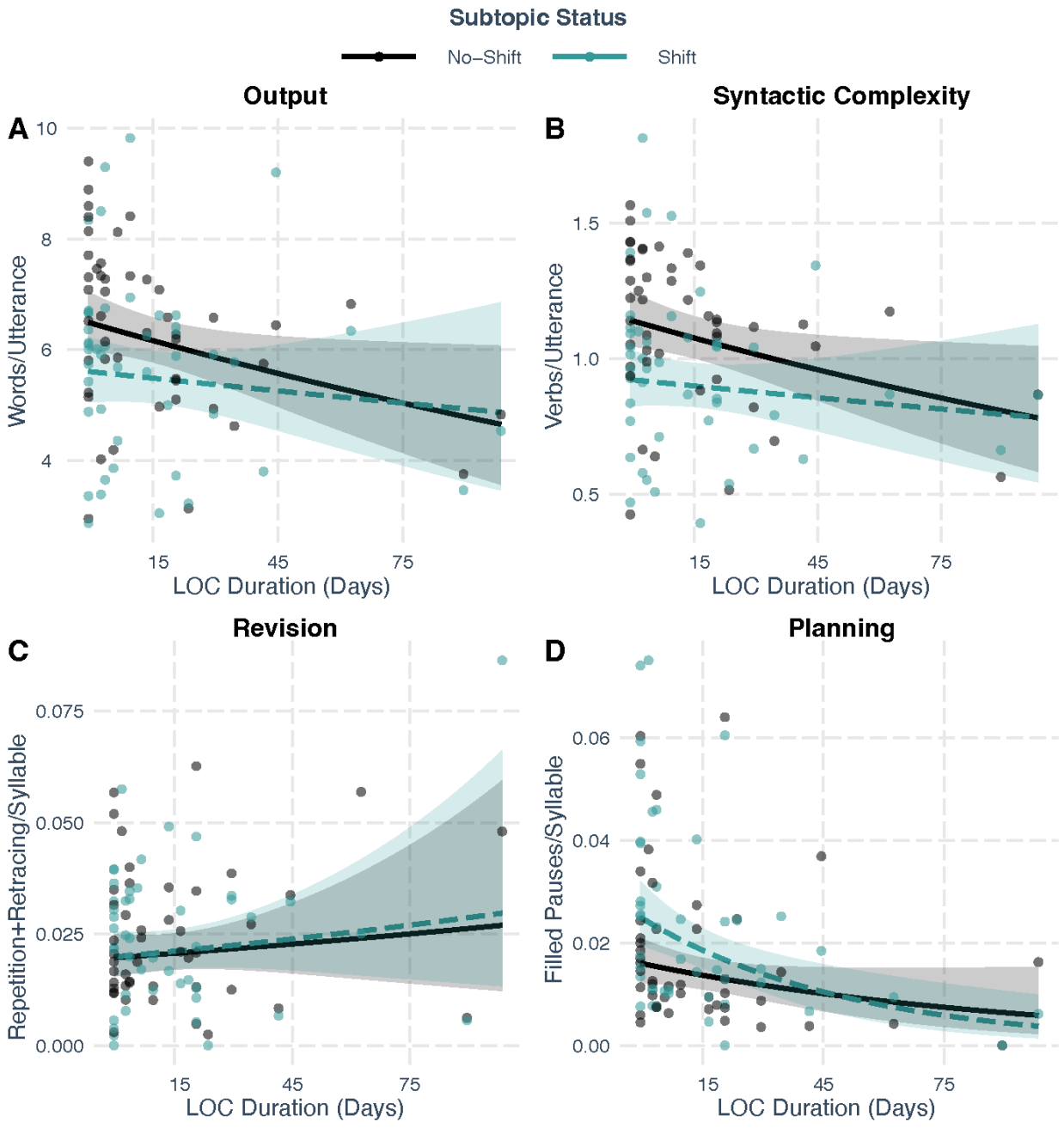


Figure 2. Interaction between the effect of brain injury severity (days of LOC) and subtopic status—turns remaining on the same subtopic (No Shift) versus turns following a partner-initiated subtopic shift (Shift) for the traumatic brain injury group. Lines show model estimates from models described in Table 8. Shaded regions represent 95% confidence intervals. For clarity, points represent raw data for individual participants. LOC values have been converted from grand-mean centered LOC measures for visualization. Interaction effects are shown for each linguistic outcome measure that was affected by subtopic status. A) Output, measured by words per utterance. B) Syntactic complexity, measured by verbs per utterance. C) Self-monitoring and revision, measured by number of repetition events plus retracing events per syllable. D) Linguistic planning difficulty, measured by number of filled pauses per syllable.

Dependent Measure	Term	Parameter Estimate	Standard Error	Z/t Statistic	P Value
A) Words per Utterance					
	Intercept	1.81	0.01	191.60	<0.001
	LOC_days	-0.00	0.00	-7.53	<0.001
	New	-0.12	0.01	-12.87	<0.001
	Age	0.00	0.00	2.66	<0.01
	Education	0.03	0.00	5.80	<0.001
	LOC_Days*New	0.00	0.00	4.70	<0.001
B) Verbs per Utterance					
	Intercept	0.07	0.02	3.21	<0.01
	LOC_days	-0.00	0.00	-3.59	<0.001
	New	-0.18	0.02	-8.20	<0.001
	Age	0.00	0.00	1.57	0.12
	Education	0.02	0.01	1.55	0.12
	LOC_Days*New	0.00	0.00	2.16	0.03
C) Repetitions + Retracings per Syllable					
	Intercept	-3.88	0.07	-57.89	<0.001
	LOC_days	0.00	0.00	1.01	0.31
	New	0.02	0.05	0.44	0.66
	Age	-0.00	0.01	-0.28	0.78
	Education	0.03	0.03	1.04	0.30
	LOC_Days*New	0.00	0.00	0.43	0.67
D) Filled Pauses per Syllable					
	Intercept	-4.29	0.08	-54.10	<0.001
	LOC_days	-0.01	0.00	-2.60	<0.01
	New	0.31	0.06	5.49	<0.001

Age	0.00	0.01	0.50	0.62
Education	0.04	0.03	1.03	0.30
LOC_Days*New	-0.01	0.00	-3.01	<0.01

Table 8. Model outputs for the effects of injury severity and subtopic status on each dependent measure, within the TBI group. Statistical models were of the form: glmer(Dependent Measure ~ LOC_days*New + Age + Education + offset(log(Total_Utts)) + (1+New|ID), family="poisson")where *LOC_days* = duration of length of coma in days (centered), *New* = factor representing subtopic status (1=Shift, 0=No shift, reference level = No shift), *Age* = age in years (centered), *Education* = years of education (centered), *ID* = participant, and *Total_Utts* = number of utterances. Specific models used for each dependent measure are available in Appendix 1.

Discussion

This study examined how changes in conversational subtopics influence language production in a speaker's next turn, following a partner's subtopic shift or continuation, in both healthy adults and individuals with TBI. We found that subtopic shifts impose measurable shift costs across both groups in multiple measures. Although individuals with TBI produced overall less elaborate language than their healthy counterparts, they exhibited similar patterns of subtopic-shift-related disruption. Interestingly, individuals with more severe injuries showed smaller shift costs, potentially reflecting reduced sensitivity to discourse structure. Together, these findings offer new insight into how topic dynamics shape language production and how these processes may be altered by brain injury.

Subtopic Shifts Affect Subsequent Language Production

Following a subtopic shift, participants produced utterances that were shorter, less syntactically complex, and contained fewer revision events compared to turns within the same subtopic. One might have expected greater production difficulty to lead to *more* rather than *less* revision; however, this pattern may be a consequence of the shorter and simpler utterances produced after subtopic shifts, where there are fewer opportunities to make errors and less demand for incremental planning. Finally, there was an increase in filled pauses following subtopic shifts, suggesting a corresponding increase in planning difficulty during language production. Taken together, the pattern of less complex production that requires additional planning is consistent with a processing cost to identifying and formulating an on-

topic response to a partner-initiated shift in subtopic. These "shift costs" suggest that listeners track discourse structure as changes occur, and that such tracking draws on cognitive resources also involved in generating fluent and grammatically complex replies. These results are consistent with the predictions of the Structure Building Framework and the idea that some shared cognitive resources underlie macro- and micro-linguistic planning.

Subtopic Shifts Affect Only Certain Language Production Measures

We also addressed which sub-processes involved in language production were affected by subtopic status. We found that overall output, semantic complexity, production monitoring and revision, and planning–but not semantic complexity–were sensitive to subtopic status. These results narrow the space of relevant behavioral measures to use to validate other methods or to use when manipulating elements of conversation in future research.

While the observed effects support the existence of subtopic shift costs, they do not unambiguously identify the mechanisms involved. The effects we observed for measures that showed shift costs are consistent with at least two possible explanations: (1) interference between different levels of structure building (e.g., discourse- and sentence-level), or (2) a general slowing of micro-linguistic processes until discourse-level planning is complete. Notably, we found that idea density was unaffected by subtopic status. This suggests that discourse-level processing does not primarily interfere with semantic or message-level planning. However, idea density may simply not be a sensitive enough measure at this level, as it is typically applied to entire narrative samples rather than subsamples of utterances. Idea density may also not be sensitive to the types of issues present in TBI, despite its utility in indexing cognitive impairment in dementia (C. Brown et al., 2008; Norman et al., 2022). In this case, other measures of semantic complexity could reveal differences, if they are indeed present.

These results also cannot speak to whether shift costs reflect demands on language-planning-specific processes or some domain-general processes such as attention, set shifting, or working memory. Understanding which specific processes are involved would both help to better specify current models

and provide insight into the appropriate targets for behavioral or neural interventions addressing discourse deficits. Future studies could better distinguish among the possible mechanisms underlying shift costs by employing alternative methods. For example, researchers might use electroencephalographic (EEG) indices of candidate processes like attention or set shifting, or incorporate attention probes near topic transitions (Boudewyn & Carter, 2018; Ji & Papafragou, 2022). If shift costs instead stem from language planning processes, other production analyses may help clarify which levels of planning are involved. For example, the location of filled pauses within an utterance (e.g., at the beginning of clauses versus before content words) can provide insight into sentence-level versus lexical-level planning difficulties (Beattie & Butterworth, 1979; Griffin & Bock, 1998; Hawkins, 1971).

Group Differences in the Extent of Subtopic Shift Effects Were Not Observed

Surprisingly, we did not find a difference in shift costs between the TBI and NBI groups. Sensitivity to subtopic status in both groups suggests that whatever processes are facilitating identification of and shifting to a new subtopic representation are intact following TBI. However, these processes might still be affected in TBI, but with the differences emerging not in topic processing but rather in the cognitive effort required (Winn & Teece, 2021, 2022). It could also be that understanding subtopic shifts is not so demanding that it exceeds the processing capacity for individuals with TBI, or that individuals in the chronic stages of TBI have (consciously or unconsciously) developed strategies that reduce the task demands of understanding subtopic shifts (Turkstra et al., 2025).

Alternatively, subtle differences in how the researcher communication partner interacted with each group could have masked group effects (Coelho et al., 2002). For example, the researchers may have subtly facilitated conversations with individuals with TBI, thus easing some cognitive burden. Conversely, if the researcher spoke to participants in the NBI group in an unnatural way, this could have induced greater shift costs than would normally exist.

Some evidence against these possibilities comes from an exploratory analysis of subtopic shift costs in the *researchers'* responses to *participant-initiated* shifts, given that participants have no obvious

motivation to systematically change their behavior when talking to the researchers. This analysis revealed the same pattern of shift costs in the frequency of filled pauses (Appendix 1; Supplementary Table 10), suggesting increased planning difficulty when the subtopic shifted for the researchers. These results are generally consistent with the primary analyses, suggesting that the presence of shift costs, and thus the lack of group effects, is not solely due to conversation partner characteristics. However, analyses that can account for factors about communication partners that are unavailable in the current corpus are needed to appropriately explore the effects of different dyad interactions on shift costs, and differences in these interactions across groups.

Note that the lack of group difference in shift costs is not because the groups were otherwise performing similarly across language measures. Participants in the TBI group generally said less (produced fewer words per utterance) and used syntactically simpler constructions (fewer verbs per utterance) than participants in the NBI group, regardless of the subtopic status of the preceding turn. Note, however, that we did not observe overall group differences in idea density, revisions per syllable, or filled pauses per syllable. Although participants in the TBI group performed within normal limits on tests of aphasia, these results reflect subclinical differences in word- and sentence-related processes relative to healthy adults. These results are consistent with other studies demonstrating differences in micro-linguistic measures within other types of discourse (e.g., narrative, procedural, persuasive) and extend these findings to conversational discourse (Ghayoumi et al., 2015; Norman et al., 2022; Stubbs et al., 2018).

While several possibilities could explain the similar shift costs across groups, these overall group differences provide some support to the hypothesis that individuals with TBI have adapted task strategies to facilitate responding to the demands of changing topics. As discussed by Peach and Hanna (2021), there may be a trade-off between discourse-level processes and those involved in word- or sentence-level production. Individuals in the TBI group may have reduced their overall contributions or the syntactic complexity of those contributions in order to allocate limited cognitive resources to higher-level discourse

processing. This trade-off may have enabled them to track subtopic shifts similarly to healthy adults. This suggests that understanding cognitive resource allocation strategies, in contrast to any particular cognitive resource per se, may be a fruitful area of future study. Specifically, examining cognitive effort, fatigue, and performance under different task demands could inform why individuals with TBI sometimes present with discourse difficulties and sometimes appear to compensate (Ettenhofer et al., 2018; Johansson et al., 2009; Müller et al., 2015; Taing et al., 2021).

However, because the language production costs were not specific to subtopic shifts (i.e., group differences were unaffected by subtopic status), these group differences may instead reflect general linguistic patterns that emerge under increased cognitive demands, such as during discourse-level tasks, compared to isolated word or sentence production. Alternatively, they may reflect features of conversation more broadly, but not topic-related ones, such as social-pragmatic demands or time pressure from turn-taking. Other factors associated with TBI, such as headache or fatigue, may have also contributed.

TBI Severity Affects Overall Language Output and Shift Costs

In the TBI group, language production across subtopic status was modulated by TBI severity: Individuals with more severe injuries produced shorter and simpler contributions to the conversation. In contrast, we found no effect of LOC duration (our estimate of TBI severity) on revision. This may reflect a floor effect, given the generally low frequency of repetitions and retracings for the TBI group. It may also indicate that revision abilities are similarly impaired across moderate and severe TBI.

Although we did not observe overall differences in shift costs between the TBI and NBI groups, we did find that TBI severity affected shift costs. Specifically, greater severity was associated with *smaller* shift costs for measures of output (words per utterance), syntactic complexity (verbs per utterance), and planning difficulty (filled pauses per syllable). This pattern suggests that individuals with more severe TBIs may be less sensitive to subtopic shifts, perhaps because they fail to detect them and therefore show no change in response. Although they may continue to respond to communication partners

to keep the conversation going, responses might be coherent only at the local level. For example, individuals might respond with something relevant to what was just said but perhaps unrelated to the higher-level topic structure, akin to answering a series of unrelated questions. Responses might also have limited information content but still allow for the conversation partner to continue in the same vein. In such cases where individuals could contribute without actually tracking a higher-level topic structure, cognitive resources available for language production would not differ between topic shift and no shift turns.

If individuals with more severe injuries are less likely to track and/or identify subtopic shifts, their representation of topic structure would likely differ, for example, containing fewer separate subtopics, relative to healthy adults. Clinically, even subtle reductions in sensitivity to topic shifts could contribute to communicative difficulties, particularly over extended interactions. This could affect discourse comprehension or memory for discourse content, and may contribute to reduced conversational engagement or off-topic comments. Future work could examine the effect of shift cost size on other aspects of conversation success, such as memory and pragmatic appropriateness, in order to better understand the relationship between topic tracking and communication outcomes. For example, it could be that less robust topic structure representations impair the ability to encode memories relative to discourse structure, or that changes in cognitive effort due to topic transitions alter the resources available for other discourse processes (Evans et al., 2024; Pichora-Fuller et al., 2016). Thus, in addition to contributing to theoretical considerations for models of discourse processing and cognitive flexibility following TBI, these findings could indicate some clinical relevance of supporting topic structure representation in cognitive-communication therapy, particularly for individuals with more severe TBI.

However, given the novel outcome measures used here, the practical and clinical significance of these effects remains uncertain. While the interaction of LOC and subtopic status are reliable, the fact that there is less available data at high LOC durations suggests that shift costs at these points should be interpreted with caution. In addition, when language production approaches a floor, it becomes

increasingly difficult to detect reliable differences between subtopic Shift and No Shift conditions. As a result, these language sample analysis measures may be less appropriate or informative for individuals with severe injuries, given insufficient speech output to assess discourse-level sensitivity.

Several considerations must also be addressed in order to determine the clinical application of measuring shift costs. First, the effect sizes of the interactions between subtopic status and severity were small, and so future work is needed to assess whether and at what level shift cost differences are reliably related to clinical outcomes (e.g., communication success), or are useful for predicting practical outcomes (e.g., quality of life, employment). Future work may also need to examine the effect of initial injury severity (e.g., LOC) versus current neurological status or cognitive-linguistic performance. Given the heterogeneity in deficits and recovery patterns from those deficits in the TBI population as they reach chronic phases of recovery, examining the relationship to present status may more clearly delineate which neurological factors or cognitive-linguistic processes are associated with topic shift comprehension in conversational discourse.

Considerations for Naturalistic Language Sampling Measures

One consideration when interpreting the results is that our study relied on observational coding of natural conversations. The analysis of natural conversations allows participants to experience all of the factors that may make conversation cognitively demanding (e.g., topic transitions, turn taking dynamics, social-pragmatic factors, combining visual/gestural and auditory/speech information, etc.), which may be important when examining complex communication abilities for which individuals with TBI report the most difficulty.

The trade-off is that many factors cannot be controlled. Without experimental control over variables such as content or speaker identity, we cannot rule out the possibility that the responses attributed to topic shifts are actually driven by correlated factors like the increased use of questions at topic transitions, the semantic content discussed, or characteristics of individual speakers. However, we did observe shift costs across both groups despite the factors that were not under experimental control.

In addition to individual characteristics of the interlocutors, the outcomes of conversations may vary considerably based on relationships between members of the dyad, such as power dynamics (e.g., clinician-patient), who is requesting versus giving information, or the participants' level of familiarity with each other (Hengst & Duff, 2007; Togher et al., 1997). Indeed, in this dataset, the research investigators had different relationships with TBI patients (clinicians) and controls (colleagues or strangers). As demonstrated in Table 4, the majority of the conversations consisted of participants contributing to ongoing topics, which might be less of a balance in contributions compared to what might be found in other kinds of conversational exchanges, like those between familiar communication partners. Experimenters were also not blind to participant group, as the TBI Bank protocol includes questions about injury history during other discourse tasks. Therefore, experimenters could have changed their conversational behavior to accommodate individuals with TBI (e.g., providing more cues to topic transitions, being prepared to take a more active role in questioning or allowing more time to talk on any turn in order to increase linguistic output for analysis). Future work can explore the same measures in dyads with different relationship parameters or may use neutral conversation prompts to avoid group-identifying information in order to examine possible effects of these parameters.

In addition, it is difficult to determine whether participants are responding to subtopic shifts because of increased demand on cognitive or linguistic structure-building mechanisms, or are simply following a natural conversational pattern—for example, starting with shorter or simpler utterances to gauge a partner's engagement before providing more substantive responses once a subtopic is established. Future research could establish whether any such consistent progression patterns exist by examining linguistic properties across turns within a subtopic, rather than only at topic transitions.

Furthermore, while we used objective markers of subtopic shifts and developed a coding scheme that accounts for utterance-by-utterance information, this method still requires rater decisions. The method employed in this study would benefit from validation from using other methods like eye tracking

or EEG, which could provide a way to measure cognitive processes or cognitive effort in the moment and could provide greater specificity about the particular cognitive processes involved.

General Conclusions

We identified a systematic pattern of costs to language production following subtopic shifts in naturalistic conversation. These results demonstrate that listeners are sensitive to changes in conversational subtopics. The occurrence of shift costs is consistent with theories of discourse processing (Gernsbacher, 1997; Peach & Hanna, 2021), but expands the scope of these theoretical models from monologic discourse, such as narrative and exposition, to conversational dialogues, despite the additional complexity of conversational discourse structure that emerges and changes across multiple parties. We also demonstrated that the extent of these shift costs is sensitive to brain injury severity in individuals with TBI.

While the specific pattern of shift costs observed in this study does not allow us to pinpoint the exact cognitive mechanisms underlying topic structure building, identifying behavioral metrics that may be sensitive to cognitive demands during naturalistic discourse offers a valuable starting point for future work. First, examining a listener's language production following a speaker's topic shift provides a complementary approach to traditional memory-based methods of probing cognitive processing at discourse or event boundaries (Davis & Campbell, 2023; Evans et al., 2024; Gernsbacher, 1997). These production-based measures offer a more immediate index of cognitive demand at the boundaries themselves, which can be correlated with subsequent memory performance. Second, behavioral metrics that occur close in time to topic shifts can inform the interpretation of results from time-sensitive measures like eye-tracking, pupillometry, EEG, or magnetoencephalography (MEG). Taken together, identifying shift costs from naturalistic language production can serve as a useful tool for testing hypotheses about specific underlying mechanisms, including attention, cognitive effort, and arousal.

We also found that neurologically healthy individuals and individuals with moderate-to-severe TBI were similarly sensitive to topic shifts in naturalistic conversation. This unexpected result narrows

the scope of possible hypotheses about conversational discourse performance in individuals with TBI. Specifically, this suggests that difficulties with conversation in TBI, such as producing off-topic comments, result from some other issue(s) besides a breakdown in topic structure tracking processes. For example, the individuals in the TBI group may have failed to inhibit inappropriate responses, thus producing topic-inappropriate responses, despite having intact topic structure representations, or may have over-relied on the conversation partner to continue the conversation due to issues with pragmatic knowledge or reduced attention to the conversation.

However, the limitations of this study also highlight the need for further investigation to determine whether topic structure tracking is a potential cause for discourse difficulty in individuals with TBI. First, the use of turn-based language production measures captured information several seconds following changes in subtopic structure, and so are insensitive to differences in language processes taking place over the course of milliseconds. Second, while the presence of shift costs in subsequent language production shows that some cognitively demanding process occurred in response to encountering a subtopic shift, it cannot delineate which sub-processes contribute to dynamically updating topic representations and to the cognitive demand. Experimental designs that can examine individual cognitive processes related to updating topic representations are needed to tease various possibilities apart. Given the heterogeneity in injuries and in TBI presentation, it could be that different participants have difficulty at different stages of the process, but this variability is averaged out in group-level analyses (Covington & Duff, 2021). Future work with this population may provide insight into discourse structure tracking processes and potential forms of breakdown in these processes if more time-sensitive measures and better accounting for heterogeneity can be incorporated into study designs.

In sum, our findings highlight the value of naturalistic language production as a window into discourse-level processing and lay the groundwork for future research to pinpoint the cognitive and neural mechanisms that support—or disrupt—successful communication in individuals with TBI.

Chapter 3: Examining Component Abilities Related to Tracking Topic Shifts

Rationale

The results of Chapter 2 suggest that individuals with TBI are sensitive to topic shifts (i.e., exhibit shift costs) to a similar extent as individuals with no history of neurological injury. These results are consistent with the hypothesis that individuals with TBI, like neurologically healthy adults, retain information about the previous discourse context in order to construct a topic representation, identify topic shifts, and engage in the cognitively-demanding processes required to update that representation. Difficulties with conversation, such as producing off-topic comments, may instead be due to some other mechanism besides a breakdown in topic structure building processes (Coelho et al., 2002). For example, the individuals in the TBI group may have failed to inhibit inappropriate responses, despite having intact topic structure representations to indicate that those responses were inappropriate, or may have over-relied on the conversation partner to continue the conversation due to issues with pragmatic knowledge or reduced attention to the task.

However, further investigation is needed to determine whether the conclusions of Chapter 2 reflect some limitations of the discourse analysis approach, rather than truly preserved topic structure in TBI. First, while the coding approach used in Chapter 2 examined language production processes that occurred relatively close in time to topic shifts, the use of turn-based language production measures captured information several seconds following changes in topic structure. These coarse-grain measures may not have been sensitive to differences in language processes taking place over the course of milliseconds. Although language processing in the NBI and TBI groups may have reached the same endpoint given enough time, leading to similar levels of performance in subsequent language production, there may have been differences over shorter time scales or differences in the amount of effort required to complete the same process.

Second, the presence of shift costs in subsequent language production can index that *some* cognitively demanding process occurred in response to encountering a topic shift, but cannot delineate which sub-process related to dynamically updating topic representations occurred or contributed to the cognitive demand. Experimental designs that can examine individual cognitive processes related to updating topic representations are needed to tease various possibilities apart. Given the heterogeneity in injuries and in TBI presentation, different participants could have difficulty at different stages of the process, but this variability is averaged out in group-level analyses (Covington & Duff, 2021). Relatedly, it is possible that participants varied in the extent to which discourse abilities were impaired, and some may not have had any difficulty with conversational discourse despite having sustained a TBI. Thus, the current experiments aim to (1) use measures with better temporal resolution; (2) investigate specific sub-processes hypothesized to support topic tracking abilities, and (3) obtain information about individual differences in participants' conversational difficulties (or lack thereof) as well as factors that may influence both cognitive abilities and communication following TBI, including post-traumatic stress symptoms (PTSS), physical symptoms (e.g., headache), and affective symptoms (e.g., depression, anxiety).

In the following sections, I will first address the cognitive processes hypothesized to be important for conversation structure comprehension. Then, I will describe how these processes may be affected by TBI. Because the cognitive processes implicated in conversational deficits in TBI are currently unknown, I will begin by examining what is known about these processes across various types of non-conversation task demands, raise relevant open questions about how the processes might be impaired in conversational contexts, and explain how deficits in these abilities could relate to conversational deficits characteristic of individuals with TBI. I will then examine evidence that addresses how healthy adults deploy these cognitive processes, specifically in the context of discourse comprehension. Finally, I will explain how comparing individuals with TBI to healthy adults within tasks that model conversational discourse processes can provide evidence about which, if any, of the relevant cognitive processes impact discourse

comprehension processes following TBI. Connecting what is known about general cognitive deficits following TBI with what is known about how these cognitive processes are deployed during discourse-relevant tasks in healthy adults will provide an experimental framework geared towards examining specific mechanisms that may lead to conversational issues following TBI.

Cognitive Processes Relevant to Discourse Structure Tracking

Given the need to examine specific components of topic structure building processes in real-time, the current work begins by focusing on processes known to be affected by TBI in non-conversation contexts. Based on evidence from discourse and event structure building frameworks (Gernsbacher, 1997; Zacks et al., 2007), I propose the relevant processes for determining when new topics arise depends on at least three key processes:

1. Memory: Remembering information that has already occurred in the conversation.
2. Prediction: Using information from the previous discourse and any available cues signaling a topic shift to formulate an expectation about whether incoming information will address the same topic (given information) or a new topic (new information).
3. Model updating: Comparing the top-down expectation about whether information will address the same topic or a new topic to the incoming language input to decide whether the input is consistent or inconsistent with the expectation, then updating the topic representation accordingly.

In the subsections below, I examine each of these three components in turn, highlighting what is known about these processes in individuals with TBI outside of conversational contexts, open questions about what individuals with TBI may do within conversational contexts, and describing the possible implications for conversation structure building processes.

Memory, Prediction, and Model Updating Following TBI

Memory

Individuals with TBI frequently experience memory deficits, including poor immediate and delayed recall of episodic information in visual or verbal modalities. While results are mixed regarding the extent of impairment for mild and moderate TBI, individuals with severe injuries have been more consistently shown to have immediate and delayed memory difficulties (Azouvi et al., 2017; Vakil et al., 2019). Meta-analyses and systematic reviews have also shown that individuals with moderate to severe TBI have deficits in working memory relative to healthy controls (Azouvi et al., 2017; Dunning et al., 2016). Models of discourse structure building suggest that discourse comprehension relies on both the ability to remember key facts, as in episodic memory tasks, and the ability to manipulate these facts to integrate them into the discourse structure, more akin to working memory tasks (Gernsbacher, 1997).

However, studies that have directly examined discourse suggest that non-discourse memory deficits may not translate to conversational contexts. Morrow et al. (2024) demonstrated that individuals with moderate to severe TBI do not appear to have a deficit in remembering details from conversations over short (20 minute) intervals, relative to healthy controls (Morrow et al., 2024). Individuals with TBI have also been shown not to differ from healthy adults in their ability to use discourse cues to boost memory for particular facts (Diachek et al., 2024). Thus, it appears that individuals with TBI and healthy controls have access to similar kinds of information from spoken discourse in memory. In turn, having memory for key facts available from the prior discourse may not be a primary issue driving conversational discourse deficits in TBI. Instead, it could be that information available in memory cannot be effectively utilized for prediction and model updating processes. Given that issues of memory, per se, within actual conversational contexts have already been addressed, the remainder of this chapter will focus on the prediction and model updating components, which have yet to be explored in relation to conversational abilities following TBI.

Prediction

Outside of conversational contexts, individuals with TBI have been shown to have difficulty using previous information to make predictions. Individuals with TBI have failed to make predictions in simple perceptual contexts, such as using motion patterns to predict where a moving object will appear (Diwakar et al., 2015). Although information about linguistic prediction, per se, is sparse, there is some evidence that individuals with TBI may not use information that is currently unfolding in the linguistic signal to narrow the scope of what information they expect to hear next. Individuals with moderate to severe TBI demonstrate reduced looks to items that are temporarily consistent with the current input, relative to healthy adults (Lord et al., 2026).

Whether a listener predicts upcoming information can be affected by whether the listener's predictions are likely to be accurate as well as the availability of reliable cues (Brothers et al., 2017; Sharer & Thothathiri, 2020; Thothathiri & Braiuca, 2021). Reduced tendency to predict in TBI could be due to either issue. Individuals with TBI may have access to cues but fail to use them to make predictions. If individuals with TBI have had experience making inaccurate predictions following their injury or if generating predictions is too metabolically costly within an injured brain, they may have adopted a general processing strategy of not predicting to avoid frequent prediction error or prediction generation costs (Kuperberg & Jaeger, 2016). Alternatively, individuals with TBI may not attend to or have fast enough access to apply useful cues to make predictions due to attention or processing speed deficits (Levin et al., 2014).

While studies have examined prediction in non-discourse contexts, there remain open questions about prediction in TBI as it specifically relates to the context of conversation. Prediction about topics in the context of conversation requires both prediction generation per se and the ability to use information from the previous discourse to generate that prediction (i.e., versus using something about the physical context, world knowledge, or linguistic knowledge). Because conversation partners also often explicitly signal shifts in topic, efficient prediction in this context may also require the ability to attend to and utilize

relevant cues (Gardner, 1987; Leaman & Edmonds, 2020). The relevant open questions about prediction within the context of conversational discourse are therefore:

1) Do individuals with TBI make predictions during online language comprehension when no previous discourse information is required?

2) If so, do individuals with TBI make predictions during online language comprehension within discourse tasks, when prior discourse information is required for accurate prediction?

3) Do individuals with TBI use reliable cues to discourse status (e.g., relating to the same topic or a new topic) to inform any predictions they do make?

Accurate prediction could prepare individuals to undergo a topic shift in advance, potentially increasing comprehension efficiency. Taking a wait-and-see approach rather than predicting whether upcoming information will be about the same topic or a new topic could mean less advance preparation for change and more difficulty with topic change at the point of difference. Conversely, it could also mean avoidance of prediction error and reduced cognitive effort if a topic shift is never initiated. However, in this case, there may be a failure to update the topic structure, which could mean difficulty integrating incoming information when it is inconsistent with the incorrect expectation (i.e. information relates to the new topic when the old topic is expected). Ultimately, reduced prediction about upcoming discourse information status could lead to increased cognitive effort, as noted in subjective report of conversations during TBI (Salas et al., 2018). Reduced prediction could also lead to differences in topic-related abilities noted in observational studies of TBI such as misinterpretation of topic structure, which could produce off topic comments, tangentiality, or reliance on conversation partners to guide the flow of conversation (Coelho et al., 2002).

Model Updating

Another possible source of difficulty with comprehending conversation may be related to deficits in model updating in response to conflicting evidence. Some individuals with TBI have difficulty comparing prior information to incoming information in order to determine when incoming information is

different enough to warrant updating a particular representation. Studies using electroencephalography (EEG) have consistently demonstrated that individuals with TBIs of varying severity show reduced amplitude in components of the P300 (P3a, P3b) event-related potential (ERP), compared with neurologically healthy adults (Duncan et al., 2011; Gilmore et al., 2018; Li et al., 2021; Solbakk et al., 2002). These ERPs are typically elicited using an 'oddball' paradigm, in which individuals hear or see an ongoing sequence of stimuli, most of which fall into a frequent category and some of which fall into an infrequent category. Participants must identify when the novel or infrequent stimulus occurs. The context-updating model of the P300 suggests that the amplitude of the P3a component of this ERP reflects cognitive resources required to compare the current stimulus to the previous stimulus held in working memory and to decide whether the stimuli belong to the same category or different categories. If a difference is detected, attentional resources are allocated to update the current context representation, and this attentional process is reflected in the amplitude of the P3b component (Polich, 2007). Thus, reduced amplitudes of the P3a and P3b components in individuals with TBI relative to healthy adults suggests that those in the TBI group have greater difficulty allocating, or that they less consistently allocate, attention to identify differences between previously occurring stimuli and novel stimuli. Reduced amplitudes could also indicate difficulty with or inconsistency in updating representations of the current context to reflect novel stimulus characteristics.

Individuals with TBI also frequently show deficits relative to healthy controls in Wisconsin Card Sort Tasks, in which participants are asked to sort cards with images that vary on several characteristics (e.g., color, shape, number) according to a rule (e.g., sort by color). Intermittently, the rule changes and participants must use experimenter feedback to determine a new rule. Individuals with TBI have been shown to have particular difficulty with detecting a new rule and shifting away from using the old rule, as demonstrated by a higher number of perseverative (i.e. using the old rule) errors versus healthy controls (King et al., 2002; Millis et al., 2001; Ord et al., 2010). This suggests that individuals with TBI have

difficulty updating their model of task relevant versus irrelevant information in response to information that the rule that was expected to continue has now changed.

According to conversation structure building models, listeners should need to compare incoming information to the expected model they have of the discourse structure. In the absence of a cue that the topic will change, the expectation should be that the current topic will continue (Gernsbacher, 1997). This is similar to the oddball task, in which frequent targets are likely to continue for multiple turns, or the WSCT, in which rules continue for some time before being changed. Thus, it is possible that individuals with TBI have the same kind of difficulty, or failure to engage in, comparing expected topic content with incoming utterances to decide if the topic structure representation needs to be updated. This suggests the following open research question:

4) Are individuals with TBI able to (consistently) update their conversation structure model when incoming information contradicts expected information?

Failure to identify the need to shift to a new topic structure could lead to similar issues as noted above regarding prediction issues. The ability to identify topic shifts and update the topic representation may be possible for individuals with TBI, but require more cognitive effort. A failure to appropriately update the topic representation could lead to difficulty comprehending incoming information in the context of the new topic (i.e., "feeling lost" as to why conversation partners are talking about what they are talking about), an inability to use the appropriate topic representation to constrain comprehension processes, or an inability to identify appropriate responses (i.e., produce information considered to be "off-topic" or fail to participate in the conversation). The potential similarity between the ultimate outcome of a breakdown in either process (prediction or comparison/model updating) highlights the need for understanding specific mechanisms, as encouraging accurate prediction and facilitating model updating could require two different treatment approaches.

Examining the Relevant Cognitive Processes in Discourse Contexts

The previous section outlined key processes that may be disrupted in individuals with TBI. These processes have been studied in non-discourse tasks, but it remains unclear whether similar impairments affect real-time language processing in natural conversation. While natural conversation involves many dynamically interacting processes, one way to address these open questions is to use tasks that are not completely unstructured, but still occur within the context of language comprehension or discourse, and can therefore serve as a simplified model of the specific prediction and updating processes outlined above.

The core aspect of the open questions raised above is how individuals understand topic structure as the conversation progresses, i.e., will incoming information continue to address the same topic or will it address a new topic and require an update to the topic representation. A similar concept in the broader discourse literature is the concept of given versus new information status–in other words, whether a concept or referent has been previously mentioned in the discourse or not. While simpler than topics (individual concepts versus sets of relevant information), given versus new information status can be manipulated in various ways to examine both prediction and model updating processes.

Healthy adults can reliably track which information has already been introduced in a conversation (i.e., given information) and use that representation to anticipate (i.e., generate predictions about) what might come next. As a result, they tend to expect upcoming information within the same discourse context to reference given information rather than new information. Healthy adults can also use reliable cues from the speaker to predict when new information is more likely than given information. When information is expected to be given but is actually new, or vice versa, healthy adults can override their initial expectation to attend to the interpretation supported by the actual input (Arnold et al., 2004; Haviland & Clark, 1974). Thus, examining listeners' understanding of given versus new information provides a means to assess prediction and updating processes in the context of discourse, without needing to control all of the other dynamic factors within unstructured conversation.

The following sections describe experimental paradigms used in previous work with healthy adults to examine prediction and updating processes. Each section then describes how applying similar experimental paradigms to individuals with TBI can answer the open research questions described above. Table 9 provides a summary for predicted task performance given the possible points of breakdown in the process of interpreting given versus new discourse information.

Prediction During Online Language Comprehension

An extensive amount of psycholinguistic research has established that healthy adult comprehenders use linguistic and world knowledge to actively generate predictions about what is likely to be heard before the information has actually arrived (although the level and specificity of the prediction can vary) (Altmann & Kamide, 1999; Brothers et al., 2017; Federmeier & Kutas, 1999; Kamide et al., 2003; Trueswell et al., 1994). However, no study has yet addressed whether individuals with TBI engage in this kind of online linguistic prediction. Other populations, such as individuals with hearing loss, may adopt wait-and-see versus predictive approaches. This suggests that linguistic prediction is not always applied, and prediction may be especially likely to be foregone under conditions of increased comprehension difficulty when predictions may be more likely to be wrong (McMurray et al., 2017). It is therefore important to examine this more basic linguistic prediction ability in order to be able to interpret results of studies where discourse information is used to inform linguistic predictions. Knowing whether individuals with TBI make predictions during online language processing can help determine whether they have difficulty generating predictions per se, difficulty applying previous discourse information to make a particular prediction, or both.

In order to gain a better understanding of linguistic prediction in individuals with TBI, Experiment 1 used a well-validated visual world eye tracking paradigm from Altmann and Kamide (1999) (referred to as “Simple Prediction Task” for the remainder of this chapter). This task was chosen both for its reliability across many replications as well as its similarity to other tasks in this study regarding the use of the visual world paradigm and prediction at the level of possible referents (versus prediction about

syntactic structures or particular lexical items, for example). In this task, participants hear a sentence containing a verb that either semantically constrains the possible later referents or does not (e.g., "The boy will [eat/move] the cake) while looking at a set of four possible referents in the visual world. When the verb provides semantic constraints (e.g., eat), healthy adult participants begin to look to the target (i.e., the cake) following the verb onset and prior to the target noun onset. When the verb does not provide constraints, looks to the target begin later, after some information about the target noun is available. Thus, differences in looking patterns to the same target for constrained versus unconstrained verb contexts demonstrate the use of semantic constraints of the verb to predict which of the items in the visual world is the intended referent (Altmann & Kamide, 1999). This prediction requires only the information from the current verb, rather than any information retained from previous discourse context. If individuals with TBI employ linguistic prediction, they should perform similarly to healthy adults. If individuals with TBI do not employ linguistic prediction, they should fail to use semantic information about the verb and wait until the onset of the referent to begin looking at the target object for both verb types.

Prediction Using Discourse Information

Experiment 2 utilized a task from Arnold et al. (2004) that provides a useful framework for examining abilities related to tracking given versus new information and using this to make predictions about upcoming discourse information (referred to as "Discourse Task" for the remainder of this chapter). When participants view a visual world containing some items that are phonological cohort competitors (e.g., **ca**ndle and **ca**mel start with the same consonant and first vowel) and hear an instruction about what item in the scene they should manipulate, like "Move the grapes above the candle", these competitors tend to draw an equal proportion of looks during the temporarily ambiguous period when both words are consistent with the input (i.e., /kae/) (Allopenna et al., 1998; Marslen-Wilson, 1987). In contrast, when one of the cohort competitors was previously given in the discourse (e.g., "Move the grapes above the candle. Now move the [candle/camel] below the salt shaker"), healthy adults had a greater proportion of looks to the given item versus the new item during the ambiguous period (Arnold et al., 2004). The

greater proportion of looks to the given item in this study demonstrates that healthy adults use information about the previous discourse to determine the discourse status of referents. Knowledge about the discourse status can subsequently guide predictions about upcoming language input. Specifically, this study demonstrated that participants expect given information to be referenced again.

Several factors may cause individuals with TBI to perform differently than healthy adults on this task. First, a failure to generate online linguistic predictions at any level (sentence or discourse) would preclude having an expectation about which cohort competitor is likely to be mentioned. In view of performance consistent with a failure to use top-down information to make predictions in visual and linguistic domains (Diwakar et al., 2015; Lord et al., 2026), individuals with TBI may not use prior discourse information to generate predictions about the status of upcoming information in the unfolding linguistic signal. If participants in the TBI group do not generate predictions, they should have an equal proportion of looks to cohort competitors rather than anticipating one or the other competitor given the previous context. If prediction is impaired, participants with TBI should also fail to make predictive looks to the target in the Simple Prediction Task, as noted in the previous section.

Second, individuals with TBI may not represent given/new discourse status, either because of an issue specific to discourse status representations or because of a failure to remember prior discourse information. This is because there would be no relevant information to guide a prediction, even if one could be made. If individuals with TBI can remember information from the prior discourse but do not use this information to classify the information relative to the discourse structure (i.e., as given or new), they should also have an equal proportion of looks to the cohort competitors. The pattern of performance across both the Simple Prediction Task in Experiment 1 and the Discourse Task in Experiment 2 can tease apart whether performance unlike healthy adults is due to a memory or discourse representation issue or a prediction issue. If there is in fact a memory or discourse status representation issue and not a prediction issue, participants should perform like healthy adults on the Simple Prediction Task but unlike healthy adults on the Discourse Task. Given previous evidence for comparable memory performance for key facts

about previous conversation between TBI and healthy groups (Morrow et al., 2024), individuals with TBI should be able to maintain information about the previous discourse referent similarly to healthy individuals. Therefore, it is more likely that intact prediction on the Simple Prediction Task and differences in performance on the Discourse Task would reflect an impairment in tracking given versus new discourse status.

Using Cues to Inform Predictions

Efficiently tracking given versus new discourse information may also be related to the ability to use cues that reliably signal upcoming new information. Experiment 2's Discourse Task similarly provides a useful task for examining this ability on a smaller, more controlled scale of discourse. This task also examined how disfluencies (e.g., "theee uh… camel" instead of "the camel") can change healthy adults' expectations about whether upcoming discourse information will refer to something given in the previous context or to something new. These types of disfluencies tend to occur when speakers are having difficulty generating a word for the next planned referent. Because it is easier to generate words for previously referenced items than it is for new items, a disfluency is more likely to occur before new information than it is before given information. Healthy adults are sensitive to this distinction. When a disfluency occurred in the same visual world paradigm as described above (i.e., Put the grapes above the candle. Now put **theee uh…** [candle/camel] below the salt shaker), participants had a greater proportion of anticipatory looks to the cohort competitor (e.g., camel, new information) than in the fluent condition. This suggests that healthy adult participants were able to use cues about speakers' production to change their prediction about whether upcoming information would be given or new (Arnold et al., 2004).

Individuals with TBI may have difficulty with using cues for several reasons. One possibility relates to the way in which healthy adults use disfluencies to make pragmatic inferences about speakers' intentions. In follow-up studies to their 2004 work, Arnold et al. demonstrated that the disfluency effect is not due to statistical knowledge about the co-occurrence of disfluencies and new referents, but instead to an inference about the speaker's intent and internal speech production processes. Namely, listeners infer

that disfluencies occur when speakers are likely to have difficulty generating words for the intended referent, such as when objects are unfamiliar (Arnold et al., 2007). An inability to use disfluency cues to change online predictions may relate to inferencing deficits in TBI. Previous work has demonstrated that individuals with TBI are less accurate at making elaborative inferences (i.e., inferences requiring activation of semantic or social knowledge not immediately available in the discourse context) during conversations (Johnson & Turkstra, 2012). While there is less evidence related to online language processing, individuals with TBI show smaller P600 ERP responses to syntactic errors (Key-DeLyria, 2016). This could be consistent with a failure to infer what a speaker meant to say in order to engage in noisy-channel error correction (Gibson et al., 2013; Ness et al., 2024; Ovans et al., 2022). However, this pattern could also be consistent with differences in task strategies or issues with morphosyntax. Besides inferencing difficulties, other possibilities include failure to identify cues as such when they occur, either because the information is missed due to inattention (Duncan et al., 2011; Gilmore et al., 2018; Li et al., 2021; Solbakk et al., 2002) or because reduced processing speed makes it difficult to apply linguistic knowledge or inferences quickly enough to generate the appropriate change to expectations (Levin et al., 2014). If any of these factors apply, then in the Discourse Task, individuals with TBI should fail to adjust their expectations for upcoming given versus new information. That is, they should show no disfluency effect—provided they are still able to track prior discourse and generate predictions.

Possible Influences of Processing Speed

The hypotheses above assume that individuals with TBI have a deficit in one or more of the specific processes outlined above. However, if there is a general slowing in neural processing speed rather than specific deficits, participants in the TBI group should show the same effects as the healthy adults, but within later time windows. If individuals with TBI can perform these prediction processes within simple tasks, but it takes longer to do so, when scaling up to naturalistic conversation it could be the case that some operations cannot be completed before discourse information continues to unfold. Then, some topic

structure information may not be able to be accurately updated, leading to increased cognitive effort over the course of the conversation or "feeling lost".

Model Updating

In addition to issues with predictive processes, individuals with TBI may also have difficulty with comparing predictions that are made to incoming information and subsequently adjusting their discourse structure model. In the Experiment 2 Discourse Task, participants' predictions about whether the upcoming referent will be given or new can either be confirmed or disconfirmed once the target is disambiguated (e.g., ca… **mel** or **ndle**). In order to complete the task of moving the target item, participants must decide if their prediction matches the actual resolution of the target word, and if not, they must update their representation of the target item. Fixation patterns after the ambiguous region reveal that healthy adult participants do begin to look away from the expected object and fixate on the actual target in order to complete the task (Arnold et al., 2004).

While not examined in the original study, pupil responses from the same Experiment 2 task may provide information about the difficulty of comparing predicted referents to the actual referents when the actual referent was expected versus unexpected. Pupillometry, which measures changes in pupil dilation, can be used as an index of the locus coeruleus (LC)'s modulation of arousal states as an index of cognitive effort (and may be related to the P300 ERP component) (Chang et al., 2024; Contier et al., 2024; Elman et al., 2017). Some work has demonstrated that arousal or cognitive effort, as measured by pupil dilation, may differ when behavioral performance is the same, potentially providing information that would not be obvious from examining behavioral responses to the task (Aston-Jones et al., 1994; Elman et al., 2017; Winn & Teece, 2021). For example, even though individuals with TBI may eventually look away from the predicted target and towards the actual target and may be able to complete the behavior of moving the target object, they may need to exert more cognitive effort to make the prediction versus input comparison and update the representation of what the target is.

Based on evidence of differences in the P3b ERP component in this population (Duncan et al., 2011; Gilmore et al., 2018; Li et al., 2021; Solbakk et al., 2002), when predictions are made, individuals with TBI should have greater difficulty comparing actual input with predicted input in order to identify the actual target referent. Individuals with TBI tend to show a reduced P3b response compared to healthy adults suggesting less comparison of anticipated and actual input to categorize given versus new information. Based on this pattern in previous work, individuals with TBI may fail to compare expected input to actual input in order to identify the appropriate target. On fluent trials of the Experiment 2 task, participants should anticipate that the given referent will be referred to again as the target item. On trials where the actual target is not given, participants must revise their expectation to identify the actual target and complete the task (i.e., move the item). The opposite applies in the disfluent condition. If this comparison and revision does not occur, I expect less cognitive effort to be required than when it does occur (i.e., in healthy adults). Therefore, I expect less of an increase in pupil size on "expected" relative to "unexpected" trials in the TBI group versus in the healthy adult group. If comparison between expectations and bottom-up input is not reduced, two other patterns may emerge. If this comparison and revision process requires a greater amount of cognitive effort for the TBI group, there should be more of an increase in pupil size on "unexpected target" relative "expected target" trials in the TBI group versus in the healthy adult group. If there are no differences in this process relative to healthy adults, there should be no differences in effect on pupil size.

Hypothesized Deficit	**Will the TBI group perform like healthy controls?**			
	Experiment 1, proportion looks	Experiment 2, fluent, proportion looks	Experiment 2, disfluent, proportion looks	Experiment 2, pupillometry
Generating linguistic predictions	No	No	No	No
Using previous discourse information in prediction	Yes	No	No	No
Using reliable cues to alter predictions	Yes	Yes	No	N/A
Comparing expected to actual discourse information	Yes	Yes	N/A	No

Table 9. Predicted performance for the TBI group versus healthy adult group on each experimental task, given the presence of a deficit in any of the listed component processes for understanding given versus new discourse information.

A Note on Eye Measures in TBI

Some vision-related changes can be associated with TBI. However, these differences are not expected to affect the measures of interest for this study. While there are some basic differences between saccade and smooth pursuit abilities in individuals with TBI versus healthy controls, there are no known differences in fixation behavior, which is the relevant measure for all proportion of fixation analyses (Cifu et al., 2015). There may be some basic differences in pupillary light response dynamics between patient and healthy groups (Truong & Ciuffreda, 2016), but pupillometry measures have been shown to index relative cognitive load in both healthy and TBI groups and be able to detect group differences (Hershaw & Ettenhofer, 2018). This suggests the visual world paradigm and fixation or pupillometry measures are appropriate for comparing individuals with and without TBI.

Accounting for Heterogeneity

Individuals with TBI may experience a variety of symptoms and comorbid conditions that influence performance on cognitive-linguistic tasks. There is a high prevalence of comorbid post-traumatic stress disorder (PTSD) and post-traumatic stress symptoms (PTSS; symptoms associated with traumatic stress, without regard to whether individuals meet diagnostic criteria for PTSD) in individuals with TBI. This is particularly true among U.S. military service members and veterans, who comprise the TBI group in this study (Tanev et al., 2014). Post traumatic stress symptoms include memory disturbances, mental and physical avoidance of reminders of the stressful event, intrusive memories, sleep disturbances, experiences of strong negative feelings, and hyper- or hypo-arousal. PTSS severity has a significant impact on behavioral neuropsychological test performance. Recent studies addressing military service members from the same population have demonstrated that PTSS may even better account for neuropsychological performance than TBI severity (Lange et al., 2020; Lippa et al., 2021a). These results highlight the need to also account for this factor in linguistic task performance in order to isolate the

effects of brain injury per se from comorbid PTSS. Given the high rate of comorbidity, it is not possible to recruit participants matched on the variety of comorbid symptom presentations or exclude participants with comorbid conditions. Rather, I used a well-validated self-report measure of these symptoms, the Post-traumatic Stress Checklist for DSM-5 (PCL-5), in order to control for these factors statistically (Blevins et al., 2015). Other factors like somatosensory symptoms, such as headache, nausea, and sensitivity to light, and sleepiness or arousal may affect task participation generally (Fostick et al., 2014; Hoddes et al., 2015; Martelli et al., 1999; Pilcher et al., 2007; Ponsford et al., 2023; Shahid et al., 2010; Vanderploeg et al., 2015; Wüst et al., 2024). These neurobehavioral symptoms were also assessed in order to control for these factors statistically.

General Method

Participants

Study participants were recruited from two sites. One group of healthy adult controls was recruited from the University of Maryland (UMD). A group of healthy adult controls and clinic patients with a history of TBI were recruited from Walter Reed National Military Medical Center (WRNMMC). Data collection for controls at WRNMMC is ongoing and there were not enough participants available to analyze this group. Therefore, for the purposes of this dissertation, the Healthy Control group refers to participants from the UMD data collection site and the TBI group refers to patients from WRNMMC. The current Healthy Control group differs from the TBI group in a variety of relevant ways (e.g., age, occupational and training experiences, interaction with the medical system, potential exposure to traumatic events) that are discussed in-depth in the Limitations section. While the WRNMMC military controls are needed account for these differences, the current Healthy Control group allows for validation of task replication and comparison against a group similar to those used in the original Altmann and Kamide (1999) and Arnold et al. (2004) studies.

Fifty-four undergraduate students from the UMD community were tested. Of this group, six were excluded. Three were excluded due to reporting one or more conditions noted in the exclusion criteria

(remote TBI/concussion, use of psychoactive medication). Three were excluded due to technical error with the eye tracker. The remaining group of Healthy Control participants includes N=48 undergraduate students from the UMD community. Participants self-reported that they had no history of speech, hearing, or language disorders, had no history of neurological disorders (e.g., stroke, TBI, concussion, epilepsy), had no history of severe mental illness (e.g., schizophrenia spectrum disorders, bipolar disorder), had normal or corrected-to-normal vision, and no double vision. Participants reported that they had not used illicit drugs, alcohol, or psychoactive medications within 24 hours of the study. Participants were native English speakers, meaning they reported they heard and spoke English before age 5, regardless of other language exposure. All participants completed informed consent as approved by the UMD Institutional Review Board and received course credit for participation.

The TBI group included N=24 United States service members or veterans recruited from the National Intrepid Center of Excellence Brain Fitness Center (NICoE BFC) at WRNMMC. No participants who were tested were excluded from the analysis. The BFC provides adjunct services (e.g., online brain training programs, biofeedback, patient education classes) to patients of any WRNMMC clinic who present with a subjective cognitive complaint (Sullivan et al., 2020). Note that these interventions are not specific to any of the linguistic processes examined in this study. Participants may have been receiving other medical or rehabilitative services at the time of participating in the study. Patients who attend the clinic are informed of multiple ongoing research studies and those who indicated interest and met the inclusion criteria were invited to participate.

Standardized objective cognitive or language testing was not feasible given the time and privacy constraints of this protocol. However, the participants who were recruited are broadly representative of the population of patients with TBI history at the BFC. A retrospective analysis of this clinic population demonstrated that individuals who present with a primary diagnosis of TBI show mildly impaired (approximately 1 to 1.25 standard deviations below the mean) scores versus age- and sex-matched service members without TBI on objective cognitive tests assessing cognitive efficiency across memory,

attention, and processing speed domains (Automated Neuropsychological Assessment Metrics; ANAM Composite Score) (Shub et al., *in prep*; Cognitive Science Research Center, 2012).

Participants had a non-penetrating TBI of any severity (mild, moderate, or severe). Participants' most recent TBI must have occurred 6 months ago or longer (i.e., they are in the chronic phase). Because not all patients in the BFC have completed a full TBI evaluation and may have inconsistent TBI history in the medical record, TBI history was documented and severity was determined by the Ohio State University TBI Identification (OSU TBI-ID) protocol (Corrigan & Bogner, 2007a). Participants who answered yes to at least one of the OSU screening questions and had at least one injury event that led to feeling dazed/having a gap in memory, led to some loss of consciousness (LOC), or both, were classified as belonging to the TBI group. TBI severity was classified based on LOC duration: LOC < 30 min was classified as mild, LOC 30 minutes – 24 hours was classified as moderate, LOC > 24 hours was classified as severe. This classification was specific to the current study based on the information available from the OSU screening. The LOC bins agree with the Department of Defense TBI classification definition (*Traumatic Brain Injury: Updated Definition and Reporting*, 2015) and other classification systems used in civilian medicine (Marino et al., 2025), but these other classification systems also take into account other injury factors not available in the current study such as structural imaging, Glasgow Coma Scale score, and post traumatic amnesia duration.

Participants self-reported no other confounding neurological diagnoses (e.g., stroke, epilepsy, brain tumor). Participants also self-reported that they are native English speakers (heard and spoke English before age 5, regardless of other language exposure), had no history of severe mental illness (e.g., schizophrenia spectrum disorders, bipolar disorder), had normal or corrected-to-normal vision and no double vision, and had no use of illicit drugs, alcohol, or psychoactive medications within 24 hours of the study. Participants were not excluded if they self-reported a history of hearing loss. If participants had hearing aids, they completed all tasks except the hearing screening with their hearing aids on. Two

participants wore hearing aids. All participants completed informed consent as approved by the UMD and WRNMMC Institutional Review Boards.

Demographic, hearing screening results, and self-reported symptom characteristics of each group are shown in Table 10.

	Healthy Controls (N = 48)	TBI (N = 24)
Demographics		
Age in Years – Mean (SD); Range	18.9 (0.9); 18-21	37.8 (11.2); 22-63
Sex – Number Male (%)	13 male (27%)	17 male (71%)
Hearing Screening		
Number Failed (%)	7 (15%)	7 (29%)
Symptom Self-Report		
SSS – Mean (SD); Range	3.1 (1.3); 1-6	3.5 (1.5); 1-6
TBI-QOL Communication – Mean (SD); Range	37.4 (4.6); 26-45	29.3 (6.7); 13-42
NSI Total – Mean (SD); Range	15.4 (12.3); 0-60	36.3 (19.3); 0-69
NSI Vestibular – Mean (SD); Range	1.6 (2.1); 0-10	3.1 (2.7); 0-9
NSI Somatosensory – Mean (SD); Range	3.1 (3.7); 0-19	9.2 (6.6); 0-21
NSI Cognitive – Mean (SD); Range	3.6 (3.0); 0-13	8.7 (4.4); 0-16
NSI Affective – Mean (SD); Range	6.1 (5.1); 0-19	12.8 (6.1) 0-23
PCL-5 – Mean (SD); Range	17.8 (13.7); 0-56	36.8 (19.7); 1-76
Injury Characteristics		
TBI Severity, Most Severe – Number (%)		
Mild	---	20 (83%)
Moderate	---	2 (8%)

Severe	---	2 (8%)
Years Since Most Recent Injury – Mean (SD); Range	---	3.9 (4.8); 0-20

Table 10. Demographic characteristics and descriptive statistics of hearing and symptom self-report measures for each group.

General Procedure

Participants at WRNMMC completed the OSU TBI-ID protocol with the experimenter. Participants at UMD did not complete the protocol because history of TBI was an exclusionary criterion.

Participants underwent a hearing screening using a portable audiometer (UMD) or Madsen Astera 2 audiometer (Natus Sensory; WRNMMC) and supra-aural headphones. Participants raised their hand if they heard a tone presented at 25 dB HL at irregular intervals. The order of presentation was as follows: Right ear – 1000 Hz, 2000 Hz, 4000 Hz; Left ear – 4000 Hz, 2000 Hz, 1000 Hz. Participants were not excluded from analysis based on hearing screening results.

Participants then completed three eye tracking tasks. Across UMD and WRNMMC testing sites, the same tasks were used. The sampling rate for eye position and pupil size (1000 Hz) and the spatial resolution of the display (1024x768 pixels) were the same across sites. The UMD Healthy Control group was tested with an Eyelink 1000 tower mounted system while the TBI group at WRNMMC was tested with an Eyelink 1000+ desk mounted system. Participants were seated in front of a 14.5 x 11 inch computer monitor (UMD) or 23 x 12 inch computer monitor (WRNMMC) with a mouse and keyboard. Participants' heads were stabilized using a table-mounted chinrest. The chinrest was positioned 14 inches from the monitor at UMD and 21 inches from the monitor at WRNMMC. Participants completed a practice eye tracking task (8 filler trials of the Experiment 2 Discourse Task (4 fluent, 4 disfluent) that were not used in the main task) to ensure understanding of the click-and-drag features of the task and to allow for adjustment of the presentation volume. Stimuli were presented at a comfortable listening level adjusted from 60 dB HL by participant over supra-aural headphones (UMD) or speakers (WRNMMC). No participant indicated a need to adjust the presentation volume, so all participants completed the

experiment with the same audio input level. Participants then completed the Experiment 2 Discourse Task and the Simple Prediction Task with the same sound presentation parameters. A new calibration and validation procedure was performed before each task. There was no practice task for the Experiment 1 Simple Prediction Task.

Participants then participated in a 10-minute unstructured conversation with the experimenter (KM or one of three female undergraduate research assistants at UMD; KM or one other female research investigator at WRNMMC). The experimenter instructed the participant that the goal of the conversation was to have a short, casual discussion like you would with a friend or an acquaintance. The experimenter began each conversation by describing an interesting event that happened to them recently (e.g., going on a vacation). From there, the conversation continued naturally between the participant and experimenter, without any other instructions or interventions. Unstructured conversation data are not included in the main analyses for this study. However, having unstructured conversation available can allow for the calculation of various language sample analysis metrics. Follow-up studies can examine the relationship between key eye tracking measures hypothesized to relate to topic information tracking and topic shift costs or between symptom self-report measures and various macro- and micro-linguistic measures.

Finally, participants completed self-report questionnaires in the following order: Stanford Sleepiness Scale (SSS), TBI-QOL Communication Scale, Neurobehavioral Symptom Inventory (NSI), and Post-traumatic Stress Checklist for DSM-5 (PCL-5). All questionnaires were completed on paper forms.

Symptom Self-Report Scales – Stimuli and Procedures

The Stanford Sleepiness Scale (SSS) consists of a single question. Participants were presented with a scale including a description of the degree of sleepiness and a corresponding numerical rating. Ratings range from least sleepy, 1 (feeling active, vital, alert, or wide awake), to most sleepy, 7 (no longer fighting sleep; sleep onset soon; having dream like thoughts). There is also an option for “asleep” with no numerical rating (Hoddes et al., 2015).

Participants also completed the TBI-QOL v1.0 Communication Short Form 9a (Tulsky et al., 2016). This measure includes two sections. The first section includes five questions related to the difficulty of completing certain communication tasks (e.g., How much difficulty do you currently have following what other people are saying?; How much difficulty do you currently have speaking?). Each question is followed by a 5-point likert scale (none, a little, somewhat, a lot, cannot do). Each scale item is associated with a check box corresponding to numerical ratings, where 5 represents the least difficulty and 1 the most difficulty/cannot do. The second section includes four questions that address the frequency of communication difficulty over the past seven days (e.g., In the past seven days… I had difficulty following the topic of conversation; I had trouble finding the right words to express myself). Each question is followed by a 5-point likert scale (never, rarely (once), sometimes (two or three times), often (about once a day), always (several times a day)). Each scale item is associated with a check box corresponding to numerical ratings, where 5 represents the lowest frequency and 1 represents the highest frequency. Scores may range from 9-45, with lower scores indicating greater communication difficulty.

The Neurobehavioral Symptom Inventory consists of a list of 22 symptoms commonly associated with TBI, such as “feeling dizzy”, “sensitivity to noise”, or “difficulty making decisions” (Cicerone & Kalmar, 1995). Symptoms can be categorized into four factors—vestibular, somatosensory, cognitive, and affective (Vanderploeg et al., 2015). Each symptom is followed by a 5-point likert scale ranging from 0 (no disturbance – rarely if ever present; not a problem at all) to 4 (very severe disturbance – almost always present and I have been unable to perform at work, school, or home due to this problem; I probably cannot function without help). Total scores can range from 0-88. Each factor score has a different range: Vestibular 0-12, Somatosensory 0-28, Cognitive 0-16, Affective 0-24. Higher total or factor scores correspond to worse symptom severity.

The Post-traumatic Stress Checklist for DSM-5 (PCL-5) consists of 20 questions that address the symptoms of post-traumatic stress disorder, as described in the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) (American Psychiatric Association, 2013; Blevins et al., 2015).

For example, questions may ask, “In the past month, how much were you bothered by…repeated, disturbing, and unwanted memories of the stressful experience; trouble remembering important parts of the stressful experience). Each symptom was associated with a 5-point likert scale from 0 (not at all) to 4 (extremely). Participants marked the box corresponding to their rating on the likert scales for symptom severity. Scores can range from 0-80, with higher scores representing higher PTSS severity.

Experiment 1 Method

Stimuli

An adapted version of Altmann & Kamide (1999)’s task was used, with two items changed to reflect cultural shifts since the original study (e.g., references to smoking) and to be more appropriate for the test population (e.g., inject versus vaccinate). Audio stimuli were recorded by a female Native-English speaker. All items were recorded in their entirety; no splicing was used. Each sentence included an agent, a verb, and a direct object such as “The boy will [eat/move] the cake.” On half of the trials, the verb was more semantically constraining regarding the kinds of plausible direct objects (e.g., one can only eat edible direct objects but can move edible or inedible objects). Whether the verb was constraining or non-constraining for a particular item was counterbalanced across two lists. Each list contained 16 critical items. Participants also heard 16 filler items. Filler items consisted of sentences with similar construction where none of the objects in the visual display were mentioned in the sentence. Filler items were the same across lists. All audio stimuli were normalized to the same loudness level (-18 dB) using Audacity’s RMS loudness normalization function (version 3.1.3) (Audacity Team, 2021).

On each trial, participants first viewed a fixation cross in the middle of the screen. Participants then saw a black background with a picture of the agent in the middle of the screen, surrounded by four objects arranged as shown in Figure 3. One of the objects, the target, was more consistent with the constraining verb than the other three objects, while all objects were consistent with the non-constraining verb. The visual display was the same on constrained and unconstrained trials. The location of the target occurred with equal frequency in each possible position (top left, top right, bottom left, bottom right)

across the critical trials.

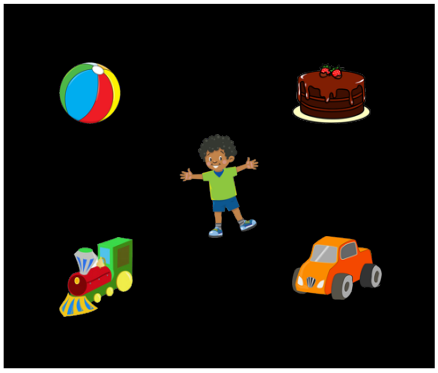


Figure 3. Example of the visual display in the Simple Prediction task.

Procedure

Each participant was randomly assigned to one of two counterbalanced lists. Before the task, participants were told that they would see objects on the screen and hear a sentence that describes the objects in the scene. They were instructed to, after each sentence, decide whether the action in the sentence could occur with the objects presented on the screen (e.g., if the sentence were "The man will light the fire" but there was no fire in the scene, the correct response is "no"). Participants then completed a 10-point calibration and validation sequence to set up the eye-tracker. There was no practice prior to this task. Participants completed 32 trials while their eye movements were tracked. Trial order was pseudo-randomly generated for each participant. The order of the trials was created using Experiment Builder Software's randomization tools such that the maximum number of filler or critical stimuli in a row was three but the order was otherwise random.

On each trial, participants saw a white fixation cross in the center of the screen on a black background for 500 milliseconds. Then, four pictures of objects and one picture of the agent of the sentence appeared in the positions shown in Figure 3. As soon as the pictures appeared, the sentence audio began to play. After the sentence was finished, participants used the "b" and "n" buttons on the keyboard to respond "yes" or "no", respectively. Because only filler trials contained no objects mentioned

in the sentence, the answer for all filler trials was “no”, while the answer for all critical trials was “yes”. Participants had five seconds to respond to the question before the trial timed out and moved on to the next trial. Pressing a keyboard button to respond automatically moved participants to the next trial.

Experiment 1 Analysis

Eye Tracking Data Processing

For each eye tracking task, data was initially exported using the Fixation Report and Sample Report functions in Eyelink’s DataViewer program (SR Research Ltd., 2024). Fixation reports aggregate data related to each detected fixation, or an eye movement that remains in one location until the gaze moves to a different location. Fixations are by definition not within blinks or saccades, so no data were excluded for these reasons. Sample reports provide data from each one millisecond sample. The section on pupillometry analyses details sample preprocessing.

Simple Prediction Task Analysis

The proportion of fixations (“looks”) on the target versus the average of the proportion of looks to the other objects per the total number of fixations was calculated within 50 millisecond time slices, across the whole length of a trial. Proportions were calculated based on fixations within the time slice, averaged across items of each trial type (constrained, unconstrained), for each participant[2]. A square 250x250 pixel interest area surrounding each object image determined whether a fixation was on a particular image.

To assess whether the results replicated Altmann & Kamide (1999), an initial model was run with only the healthy adult participants. Generalized additive mixed models (GAMMs) were used to assess the effect of verb type (constrained, not constrained) and object type (target, other) on the empirical logit transformed proportion of looks. The time window for analysis was from the onset of the sentence until the approximate end of the sentence across trials (0 to 4000 milliseconds). The emplogit() function was

[2] Data was averaged over individual items because there were not enough looks to items to effectively calculate the empirical logit value without additional averaging (i.e., there were a high number of values of 1 or 0). Models with item level data had poorer residual fits than models with participant level data. Empirical logit approaches rather than binomial models were used in order to be able to account for autocorrelation in the models.

used to perform an empirical logit transformation on the proportion of fixations to make the outcome measure unbounded in order to meet model assumptions (Fraundorf, 2022). Models were constructed using the bam() function from the mcvg package (Wood, 2003, 2011, 2017) and were assessed and visualized using the itsadug package (vanRij et al., 2022). Ordered factor coding was used to model the levels of each main effect and the difference in constrained versus unconstrained trials for each object type, with constrained verb type and target object type used as the reference level (i.e., factors IsUnconstrained, IsOther, IsUnconstrainedIsOther). Each factor was used as a parametric term and within an ordered smooth term with a continuous factor of time. Parametric terms represent differences in overall proportion of looks across time, or overall height differences. Each ordered smooth term represents a difference in the shape of the response over time between a condition and the reference smooth baseline (referring to the reference level specified in the ordered factor coding). For example, the ordered smooth term examining "IsUnconstrained" represents the difference in the shape of the proportion of looks to the target object when the verb is unconstrained versus when it is constrained. Each ordered smooth term was specified as a tensor product term. Factor smooth terms for participants and participants by each of the factor terms were used to model differences in responses to each factor type by participant, similar to including random slopes for object type, constraint, and their interaction by participants in linear mixed effects models (Sóskuthy, 2021). The AR1 value of the initial model was used to determine the autocorrelation parameter rho in the final model in order to account for autocorrelation between points in the time series, as participants' eye movements at any given time are related to where they were looking before. The rho value was then increased in magnitude manually until the AR1 value of the model was <0.2 and non-negative (Johns et al., 2024; Porretta et al., 2018). These and all subsequent model specifications can be found in Appendix 2.

To assess whether there was a difference in the proportion of looks to the target over time between the healthy control and TBI groups, a second model with the same specifications and model building process was used, but with an ordered factor of Group, rather than Object Type. Random terms

were included for participants and constraint by participants. Healthy controls and constrained verb type were used as the reference level.

An additional set of models were constructed to account for covariates of age, PTSS, and neurobehavioral symptoms associated with TBI in order to determine whether any of the observed effects were attributable to factors besides TBI status. Sleepiness score was not included as a covariate because the range of response options is too small to effectively model with GAMMs. Communication difficulty on the TBI-QOL was used as a descriptive factor for the population but not assessed as a covariate. Each model examined only the TBI group in order to assess whether patterns were consistent across these covariates (and therefore more likely to be attributable to TBI) or whether the observed group effects are more likely attributable to one of these factors rather than TBI per se. Given that these covariate factors primarily apply to the TBI group (e.g., TBI-related symptoms) and much of the range of the covariate values is not accounted for in the healthy groups (e.g., with mostly non-overlapping age groups) leading to missing data and an inability to interpret the model for those missing values, only the TBI group was assessed. The models included an ordered factor of disfluency or information status, without the group interaction. Non-linear interaction terms of Time and the covariate factor were used in the fixed tensor product smooth terms. An additional random factor smooth term of the covariate factor by participants was included (Johns et al., 2024).

Experiment 1 Results

Healthy Control Group Performance

The model addressing only the Healthy Control group confirmed that the updated task replicated the results of Altmann & Kamide (1999). As seen in Figure 4A, Healthy Control participants began to make predictive looks to the target object after verb information was available but before information was available from the actual target word. This is demonstrated by the significantly higher proportion of looks to the target on constrained versus unconstrained trials between 1697 and 3232 milliseconds into the trial, as the average onset of the verb and target noun occurred at 1813 and 2714 milliseconds, respectively

(Figure 5; Table 11). There was also a significantly higher proportion of looks to the target on unconstrained trials later in the trial, between 3394 and 4000 milliseconds. This difference reflects that participants waited to look to the target until after the onset of the target word on unconstrained trials, causing a peak in looks to the target later in the trial (Figure 5; Table 11). There was a brief period from 646-808 milliseconds in which looks to the target were greater for unconstrained than constrained trials, but this occurred prior to any verb or target noun information when there were no differences in the trial information in the trial and may therefore be spurious (Figure 5; Table 11). Figure 6 and Table 11 show a significant parametric and smooth effect of Other, demonstrating that looks to the target were different than looks to the other object throughout the trial. There was a significant interaction between object type and whether the verb provided constraining information, which occurred because looks to the other objects were not affected by the verb information in the same way that looks to the target were (Figure 7; Table 11).

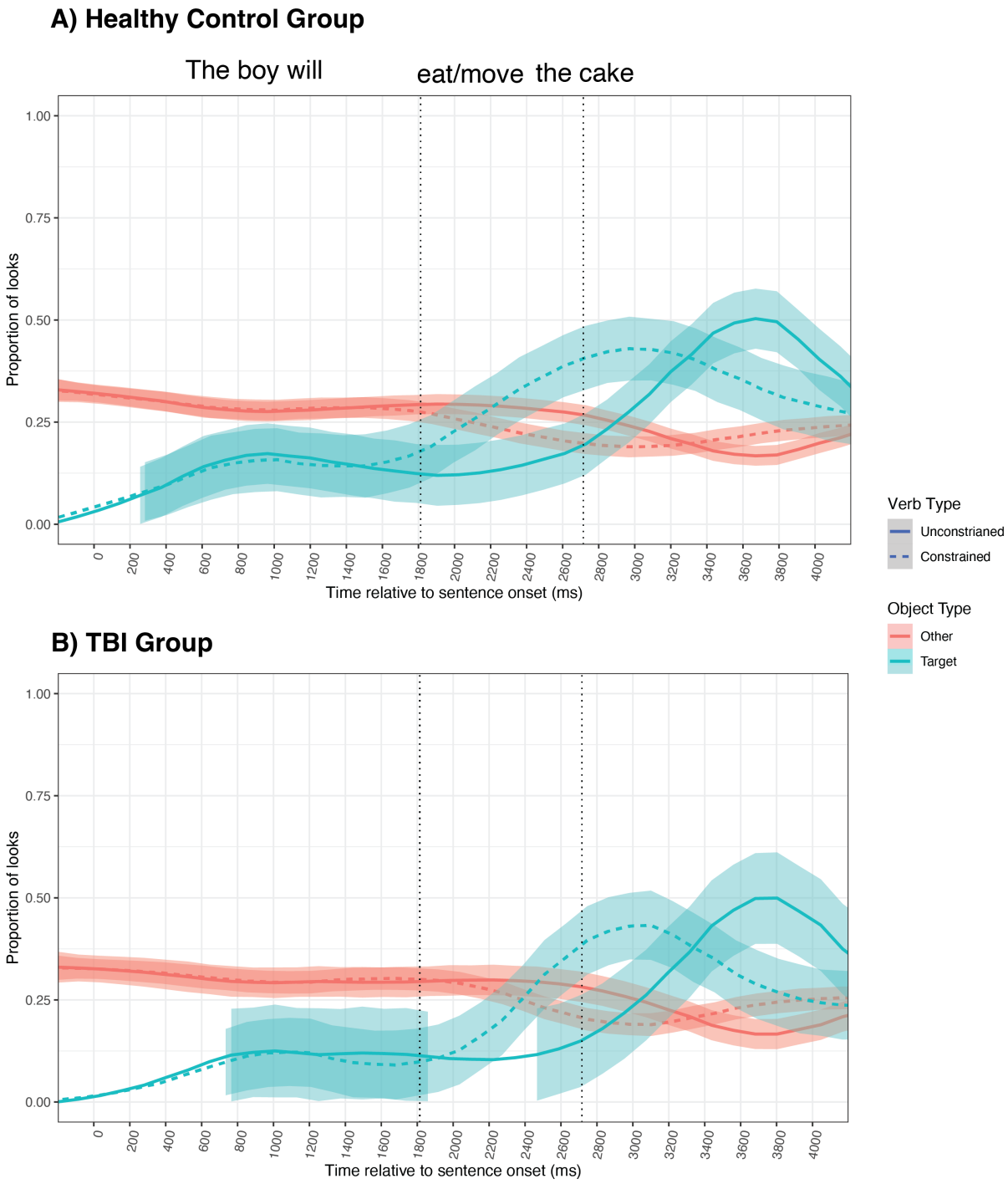


Figure 4. Average proportion of fixations (looks) to each object type per total fixations during each 50-millisecond time slice. Times represent the time in milliseconds since the onset of the sentence. Target = object named, Other = average of other 3 unnamed objects. Constrained = the verb used in the sentence is consistent with only the target object, Unconstrained = the verb used in the sentence could apply to any of the pictured items. Shaded regions represent standard error. Dotted vertical lines show the time window of the onset of the verb to the onset of the target noun, which is the time region where looks to the target are predictive (i.e., because the target has not yet been said). Example text shows the relative position of

each part of the sentence over time. Panel A shows the results for the healthy adult participants (N=48). Panel B shows the results for the TBI group participants (N=24).

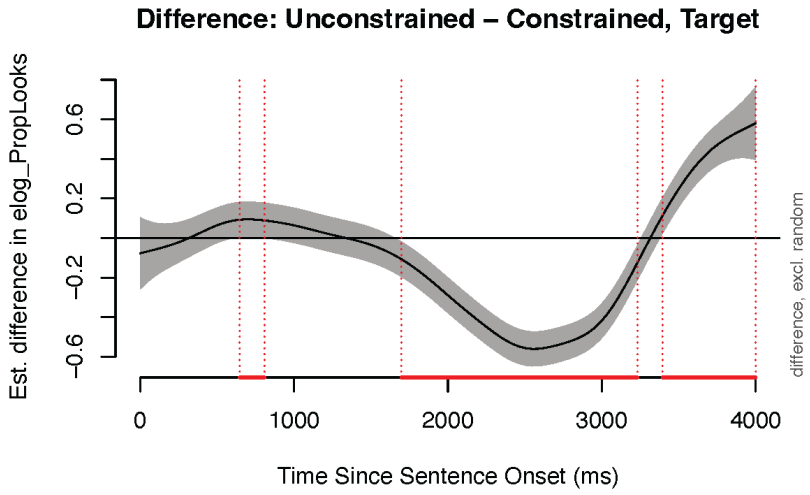


Figure 5. Difference curve showing model estimated empirical logit of the proportion of looks to the target on Unconstrained - Constrained trials for the Healthy Control group. Positive values indicate more looks to the target on unconstrained than constrained trials. Time represents the time in milliseconds relative to the onset of sentence. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

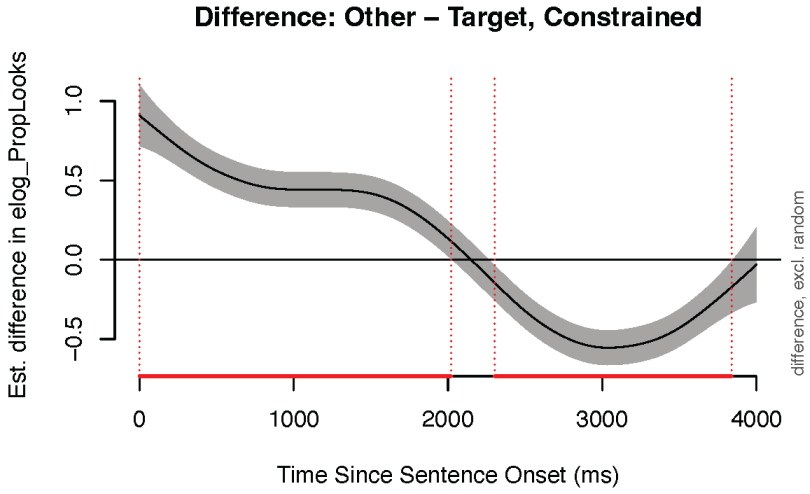


Figure 6. Difference curve showing model estimated empirical logit of the proportion of looks to the average of Other – Target objects on Constrained trials for the Healthy Control group. Positive values indicate more looks to the other than target object. Time represents the time in milliseconds relative to the onset of sentence. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

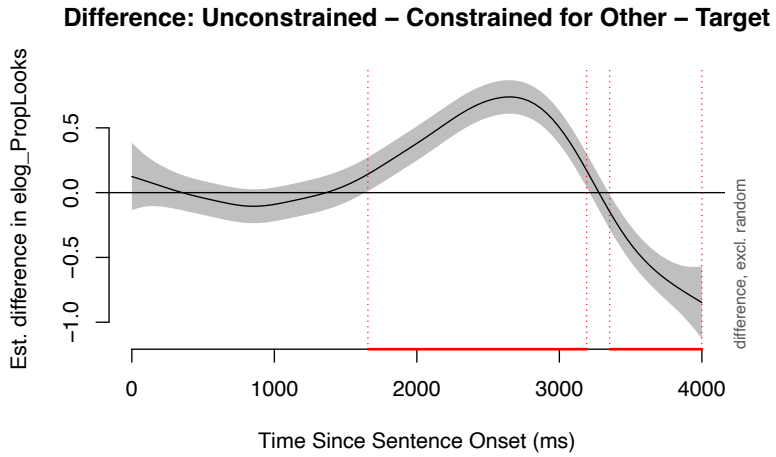


Figure 7. Difference curve showing model estimated empirical logit of the proportion of looks for the interaction of verb type and object type for the Healthy Control group. Positive values indicate a more negative difference between Unconstrained-Constrained trials for the Target than the Other object. Time represents the time in milliseconds relative to the onset of the sentence. Shaded areas show the 95% confidence interval for the difference between the two sets of conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.69	0.01	-46.23	<0.001
IsUnconstrained	-0.09	0.02	-5.30	<0.001
IsOther	0.08	0.02	3.17	0.002
IsUnconstrainedIsOther	0.11	0.03	4.20	<0.001

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value

Time	9.77	11.77	50.20	<0.001
Time:IsUnconstrained	9.55	11.56	25.04	<0.001
Time:IsOther	7.26	8.91	41.40	<0.001
Time:IsUnconstrainedIsOther	8.47	10.35	21.87	<0.001
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	53.20	959.00	0.09	<0.001
Time, Participant:IsUnconstrained	0.00	959.00	0.00	0.935
Time, Participant:IsOther	0.00	959.00	0.00	1.00
Time, Participant:IsUnconstrainedIsOther	0.00	959.00	0.00	1.00

Table 11. Model outputs for the model examining the effect of verb type (Constrained/Unconstrained) and object type (Target/Other) group on the empirical logit of the proportion of looks for the Healthy Control group

Group Comparison

Comparing the Healthy Control and TBI groups revealed no significant differences in the smooth term, or shape of the response curve for looks to the target over time between the groups. There was also no interaction between group and trial type. As shown in Figure 4 and confirmed by the model outputs (Figure 8; Table 12), across groups, participants made predictive looks to the target when information was available from the verb (i.e., in the constrained condition), but waited until information was available from the target word in the unconstrained condition. However, there was a significant parametric effect of group. Figure 9 shows that participants in the TBI group looked less to the target overall on constrained trials (Table 12). This effect occurred from 202-3313 milliseconds into the trial, which includes pre-verb portions of the sentence as well as the time between the verb and target noun onset. Overall, these results suggest that individuals in the TBI group engage in simple online linguistic prediction, despite making generally fewer looks around the screen before target information arrives.

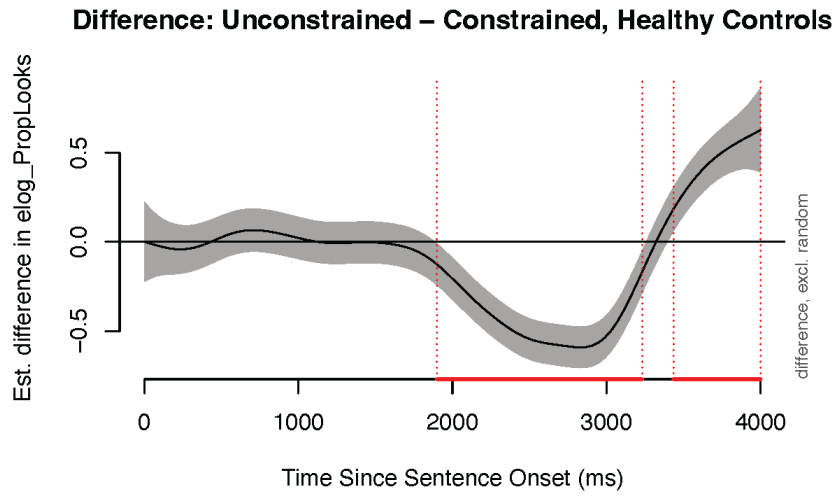


Figure 8. Difference curve showing model estimated empirical logit of the proportion of looks to the target on Unconstrained - Constrained trials for the Healthy Control group. Positive values indicate more looks to the target on unconstrained than constrained trials. Time represents the time in milliseconds relative to the onset of sentence. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

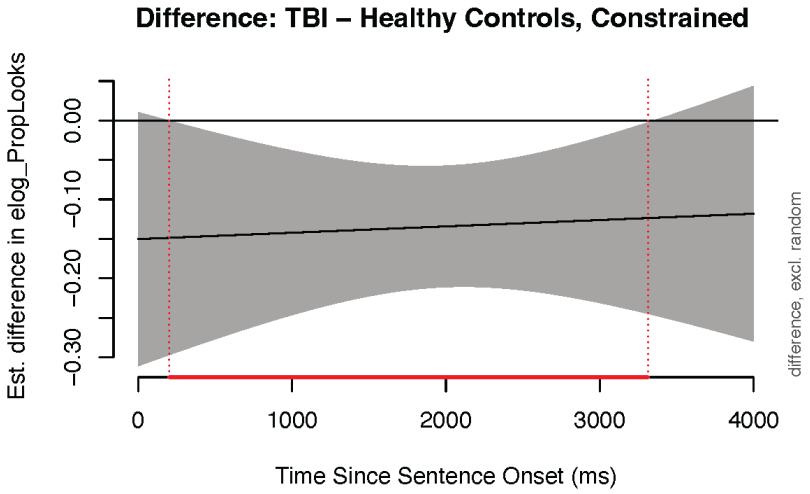


Figure 9. Difference curve showing model estimated empirical logit of the proportion of looks to the target on Constrained trials for the TBI group – the Healthy Control group. Positive values indicate more looks to the target for the Healthy Control group. Time represents the time in milliseconds relative to the onset of sentence. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms				
Term	**Parameter Estimate**	**Standard Error**	**t Statistic**	**P Value**
Intercept	-0.69	0.02	-30.31	<0.001
IsUnconstrained	-0.09	0.03	-3.68	<0.001
IsTBI	-0.13	0.04	-3.42	<0.001
IsUnconstrainedIsTBI	0.04	0.04	0.82	0.41
Smooth Terms				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time	8.40	10.29	29.67	<0.001
Time:IsUnconstrained	9.48	11.58	20.36	<0.001
Time:IsTBI	1.00	1.00	0.05	0.82
Time:IsUnconstrainedIsTBI	1.00	1.00	0.00	0.96
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	1408	1438.00	0.19	<0.001
Time, Participant:IsUnconstrained	0.00	1438.00	0.00	0.39

Table 12. Model outputs for the model examining the effect of verb type (Constrained/Unconstrained) and group on the empirical logit of the proportion of looks to the target.

Effect of Covariates Within the TBI Group

Within the TBI group, there were significant interactions between verb constraint condition and age, PTSS, and neurobehavioral symptoms (Table 13). As shown in Figure 10A, age primarily affected looking behavior in the constrained condition, with older participants' predictive looks to the target occurring later and at a lower proportion of total looks than for younger participants. The pattern of looks to the target on unconstrained trials was similar across ages. Figure 10B demonstrates that the interaction

of PTSS with verb constraint appears to be due to behavior on both constrained and unconstrained trials. On unconstrained trials, participants with higher PCL-5 scores began looking at the target earlier after the onset of target noun information but with a lower total proportion of looks to the target over time. Those with lower PCL-5 scores began looking at the target later but with higher peak proportion of looks. On constrained trials, higher PCL-5 scores were associated with a slightly later onset and peak of predictive looks to the target relative to the verb onset and slightly lower peak proportion of looks to the target. NSI score had a limited effect on looking behavior on unconstrained trials, but was associated with a later onset of predictive looks to the target on constrained trials (Figure 10C).

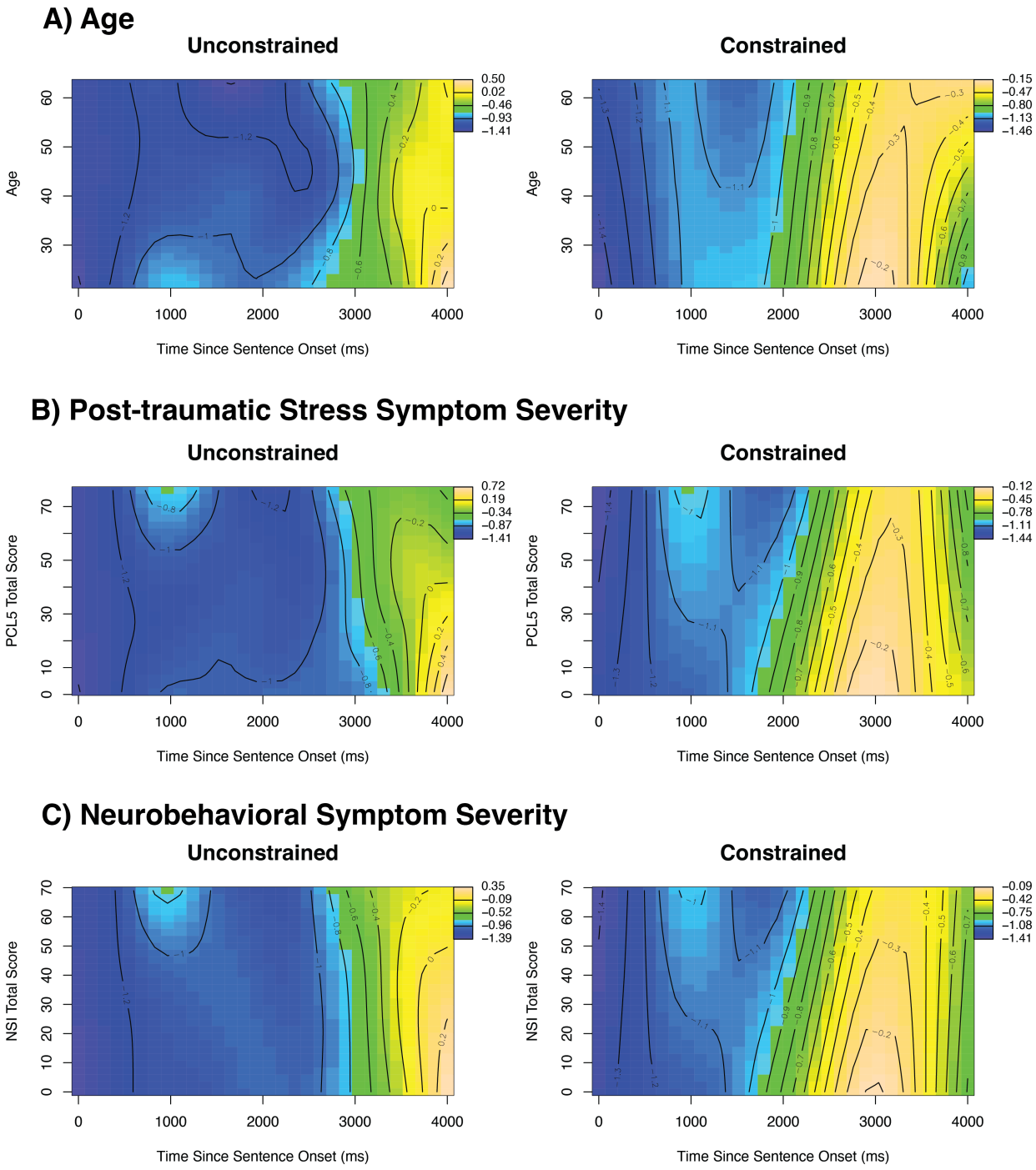


Figure 10. Heatmaps showing model estimated empirical logit of the proportion of looks to the target on Unconstrained or Constrained trials for the TBI group as a function of continuous covariate factors: A) Age in years, B) Post-traumatic stress symptom severity as measured by the PCL-5 total score, C) Neurobehavioral symptom severity as measured by the NSI total score. Warmer colors represent more looks to the target. Time represents the time in milliseconds relative to the onset of the sentence.

A) Age

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.81	0.03	-24.22	<0.001
IsUnconstrained	-0.05	0.03	-1.60	0.11

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, Age	11.57	13.57	10.16	<0.001
Time, Age:IsUnconstrained	17.92	23.38	4.61	<0.001

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Age, Participant	10.03	23.00	0.82	<0.001
Time, Participant	42.92	238.00	0.37	<0.001
Time, Participant:IsUnconstrained	0.00	238.00	0.00	0.30

B) Post-traumatic Stress Symptoms

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.81	0.03	-23.64	<0.001
IsUnconstrained	-0.05	0.03	-1.55	0.12

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, PCL5	12.15	14.21	10.19	<0.001
Time, PCL5:IsUnconstrained	17.14	22.59	4.58	<0.001

Random Effects

Term	edf	Ref.df	F Statistic	P Value

PCL5, Participant	9.54	23.00	0.74	<0.001
Time, Participant	41.70	238.00	0.35	<0.001
Time, Participant:IsUnconstrained	0.00	238.00	0.00	0.25

C) Neurobehavioral Symptoms

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.80	0.03	-24.00	<0.001
IsUnconstrained	-0.05	0.03	-1.56	0.12

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, NSI	11.92	13.96	9.81	<0.001
Time, NSI:IsUnconstrained	14.36	18.50	5.21	<0.001

Random Effects

Term	edf	Ref.df	F Statistic	P Value
NSI, Participant	13.50	22.00	1.80	<0.001
Time, Participant	39.80	238.00	0.35	<0.001
Time, Participant:IsUnconstrained	0.00	238.00	0.00	0.31

Table 13. Model outputs for the model examining the effect of information status (Given/New Target) and A) Age, B) Post-traumatic stress symptom severity, and C) Neurobehavioral symptom severity on the empirical logit of the proportion of looks to the competitor for fluent trials only and for the TBI group only.

Experiment 1 Discussion

Confirming Task Replication

First, analysis of only the Healthy Control group demonstrated that the results of the original Altmann & Kamide (1999) study could be replicated with our updated materials and within our

experimental procedures. These analyses also confirmed the time course of these effects matched those outlined in the original study using time-sensitive GAMMs. Thus, the meaning of each of the effects is interpreted in light of the interpretations supported by the previous literature on healthy adults.

Within-sentence Prediction in the TBI Group

Comparing the performance of the TBI and Healthy Control group on the Simple Prediction Task revealed no difference in the pattern of looks to the target object as a function of information available in the verb. In both groups, participants made predictive looks to the target during the time period between verb onset and target noun onset when the verb contained information that constrained the possible target noun. When the verb was uninformative about the possible referent, participants in both groups waited until the target noun onset before initiating looks to the target object. The absence of the group differences in the shape of the response suggests that there is not a general "wait-and-see" approach across participants in the TBI group. The parametric effect of group could be due to generally reduced looking to the target in the TBI group. While this effect is constrained in time, the time period does not specifically align to the pre-target period and there is no interaction between group and constraint condition, suggesting that individuals with TBI look to the target less regardless of when information about the target is available.

These results suggest that individuals with TBI are able to use information from the incoming language input to quickly generate (accurate) predictions about likely upcoming referents, and that the time course of this process is similar to processes in healthy adults. Thus, individuals with TBI may not have equivalent deficits in predictive processes across domains. Differences in impairment to predictive abilities in other domains, like simple motion tracking, may be due to domain-specific processes or differences across the task parameters. For example, motion tracking requires some proprioceptive and spatial comparison processes (e.g., knowing where one's eyes currently are and where they should go relative to another moving object) that are not necessary in the current task. Or, in regards to task constraints, there is a small, constrained number of possible options for the referent in the current task, but

other tasks may have a much higher number of outcome options, which makes it too difficult to generate many predictions quickly enough to be used during online processing or too difficult to maintain many predictions in working memory.

Effect of Covariates on Simple Prediction

Each of the covariates interacted with predictive looking behavior. Older age was associated with later and a lower proportion of looks to the target on constrained trials, but there were limited differences in looking behavior on unconstrained trials. General slowing or reduction of online prediction during sentence comprehension is consistent with performance on other language tasks in older adults (Federmeier et al., 2010; Wlotko et al., 2012). However, it should be noted that this sample included mostly young to middle aged adults, and effects occurred primarily at higher ages (>50), which are approaching ages typically addressed in studies of healthy aging. While some of the observed difference in the proportion of looks to the target during constrained trials between the TBI and Healthy control groups may be attributable to age, the small number of participants in the higher age range suggests that age likely does not completely explain the effect of group in the main analyses.

In regards to PTSS, the contradictory effects on constrained and unconstrained trials (i.e. looking later and earlier, respectively) could have cancelled out or flattened some of the constraint effects in the TBI group, which may have obscured some timing effects that may exist between the Healthy Control and TBI groups. However, the effect of PTSS was not strong enough to cause a group difference in the shape of the response curves in the primary analysis. Overall, these results demonstrate that there is not a categorical difference in predictive language processing for participants with TBI who also have higher levels of PTSS.

Neurobehavioral symptom severity had a similar effect to PTSS severity on constrained trials, but appeared to have less of an effect on unconstrained trials. While higher NSI scores were associated with later and slightly lower proportion of predictive looks to the target after the verb onset, this effect was also not strong enough to cause a timing difference in the group analysis.

Despite the limited effects on timing and the shape of the response curve, it is possible that the slightly lower proportion of looks to the target on constrained trials at higher levels of PTSS and neurobehavioral symptom severity contribute to the significant effect of fewer overall looks to the target in the TBI versus the Healthy control group. However, given current theoretical understanding of PTSD or symptom-self report, it is unclear why PTSS or neurobehavioral symptoms would primarily impact predictive processing, as in constrained trials, rather than generally slow all language comprehension across conditions (e.g., because of factors like hypoarousal or disengagement from the task) (American Psychiatric Association, 2013). It could be the case that making linguistic predictions is more difficult and requires more cognitive resources than waiting for linguistic information to arrive, and PTSS or high neurobehavioral symptom load are associated with lower cognitive resources or different resource allocation. However, a likely alternative is that individuals whose TBI history led to worse linguistic symptoms also have worse neurobehavioral and trauma-related symptoms all stemming from neurological damage. The current data cannot differentiate between alternatives of the cause of differences in predictive language processing between the Healthy and TBI groups. However, these data both provide context for the current results and provide useful information to guide future studies specifically designed to isolate effects of TBI from the covariate factors. Additionally, these results highlight the importance of examining a more demographically similar sample in order to confirm that factors like age, PTSS, and neurobehavioral symptoms are not driving the effects in the TBI group.

Experiment 2 Method

Stimuli

This task used stimulus items from Arnold et al. (2004), plus additional items created with the same method, added to increase the total number of trials. Audio stimuli for critical items consisted of two-sentence sets of directions, using the following format:

Sentence 1: “Put the grapes above the [candle/camel].”

Sentence 2: “Now put [the/thee uh…] [candle/camel] above the salt shaker”.

On each trial, the first sentence contained one member of a phonological cohort competitor pair (i.e., candle and camel have the same onset /kae/). This object would then be considered "given" in the discourse. The second sentence contained one member of the same cohort competitor pair. On "given" trials, the target (i.e., the object to be moved) in Sentence 2 was the same cohort competitor as in Sentence 1 (i.e., Put the grapes above the candle. Now put the candle above the salt shaker). On "new" trials, the target in Sentence 2 was the cohort competitor not mentioned in Sentence 1, meaning it was the discourse-new object (i.e. Put the grapes above the candle. Now put the camel above the salt shaker). Objects in cohort pairs were matched on Kucera and Francis (KF) frequency (Kučera & Francis, 1967), visual complexity, and familiarity (Rossion & Pourtois, 2001). T-tests demonstrated no significant differences between these properties between the words used as cohort competitors or between target and competitor objects across lists (*p*s>0.05). On half of the trials, Sentence 2 contained a disfluency (i.e., thee uh…).

A female Native-English speaker recorded the audio stimuli. For each trial, Sentence 1 used the same audio whether the following Sentence 2 had a given or new target. In Sentence 2, each trial contained the same recording of "Now put". This portion was taken from an original recording of a disfluent trial, in order to increase the naturalness of the pacing when splicing in disfluencies for disfluent trials. On fluent trials, the remainder of the sentence was recorded (i.e., "the candle above the salt shaker") and spliced together with "Now put". On disfluent trials, the same recording of "thee uh…" was spliced into each audio file, replacing "the" ("thee" rhymes with "tree", "the" rhymes with "duh"). There was no splicing for the filler trials. All audio stimuli were normalized to the same loudness level (-18 dB) using Audacity's RMS loudness normalization function (version 3.1.3) (Audacity Team, 2021).

A total of 24 critical trials were created according to the format above. Which object from the cohort pair was given in Sentence 1, whether the following target object was given or new in Sentence 2, and whether the trial was fluent or disfluent was counterbalanced across 8 lists. For each list, a quarter of the trials fell into each category: Fluent-Given target, Fluent-New target, Disfluent-Given target, or

Disfluent-New target. Each list also contained 36 fillers, half of which were fluent and half disfluent. Filler trials still contained sets of two directions and a cohort competitor pair, but the object given in Sentence 1 was not a cohort competitor member. For fillers, the given object was also always mentioned in Sentence 2, demonstrating to participants that in this environment discourse-given objects are likely to be referred to again. However, the discourse-given object was the target on half of the trials and the object of the preposition on the other half. On disfluent trials, the disfluency occurred only before the previously unmentioned object. Together, these filler properties remove any contingencies between cohort competitor status and role in the sentence, such that participants should not predict a particular object based on the cohort competitors (only based on given/new status). The filler trials were the same across all eight lists.

On each trial, objects appeared on a grey background. A fixation cross appeared for 500 milliseconds, followed by four objects displayed simultaneously in locations 1, 2, 3, and 4 in the positions shown below (Figure 11). Participants could use the mouse to click and drag objects according to each instruction. On critical trials, the instructions always included placing objects above or below the cohort competitor pair, such that there would be no intervening objects between the two competitors. The cohort pairs appeared horizontally to each other (i.e., positions 1 and 2 or 3 and 4, Figure 11). Half of the cohort pairs occurred on the top half of the grid (position 1/2), and half on the bottom (position 3/4). On filler trials, objects could be placed to the left or right of objects that would subsequently be the target. The visual display was the same for each item (sentence pair), regardless of given/new, fluent/disfluent status across lists.

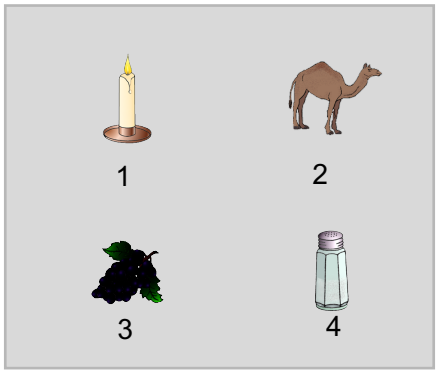


Figure 11. Schematic of the participant's view of the Discourse Task. Numbers were not visible to the participant; they are shown here for reference to object locations described in the text.

Procedure

Each participant was randomly assigned to one of eight counterbalanced lists. Before the task, participants were informed that they would see four objects and hear two instructions on each trial. They were instructed to listen to each instruction and complete the action by clicking on objects and dragging them to the appropriate location within 1.5 seconds. As in the original Arnold et al. (2004) study, participants were told that the instructions were recorded by another participant in order to increase the plausibility that disfluencies were interpreted as related to speaker difficulty generating a novel word. Before the main task, participants completed 8 trials of practice items while their eye movements were tracked. Practice trials were fillers that were not used in the main experiment task. Half contained a disfluency. After the practice, participants were given the opportunity to adjust the audio volume to a comfortable listening level. Two UMD participants completed more than one round of the practice, while the rest completed one round. All WRNMMC participants completed one round of practice. Prior to the main task, the 10-point calibration and validation sequence was repeated. During the main task, participants completed 60 trials while their eye movements were tracked. Trial order was pseudo-randomly generated for each participant. For each participant, the order of the trials was generated using Experiment Builder Software's randomization tools such that the maximum number of filler or critical stimuli in a row was 3 but the order was otherwise random.

On each trial, participants saw a black fixation cross on a grey background in the center of the

screen for 500 milliseconds. Then, a group of four pictures appeared on the screen at locations 1-4 (Figure 11). A mouse cursor also appeared in the center of the screen. As soon as the objects appeared, the first spoken instruction began. Participants had 1.5 seconds after the end of the Instruction 1 audio ended to click and drag the object to the appropriate location. After 1.5 seconds, the second spoken instruction began. Participants had 1.5 seconds after Instruction 2 ended to click and drag the target object to the appropriate location. If participants did not click and drag the objects within the allotted time, the trial events would continue automatically. After each trial ended, participants saw a break screen. Participants then pressed the space bar to move on to the next trial.

Experiment 2 Analysis

Fixation Analyses

The same eye tracking initial data processing described in Experiment 1 was used in Experiment 2. For the Discourse Task fixation analyses, the proportion of fixations on the target, competitor, or the non-target non-competitor object that was not moved ("other") per the total number of fixations anywhere on the screen was calculated within 50 millisecond time slices. Proportions were calculated based on fixations within the time slice, averaged across trial types (fluent-given, fluent-new, disfluent-given, disfluent-new), for each participant. Looks to an object were defined based on a rectangular interest area encompassing the image and surrounding area, which was 1/25 of the 1024x768 pixel screen area. A time window of -200 to 1000 milliseconds relative to critical target word onset was used to capture both a baseline level of looks to objects prior to the critical information and looks until approximately the end of Instruction 2.

Similar to Experiment 1, an initial model was created to examine performance in the Healthy Control group to assess whether the task replicates the results of the original Arnold et al. (2004) task. To perform statistical analyses on the proportion of looks, GAMMs were used with a similar construction as described for the Simple Prediction Task. The emplogit() function was used to perform an empirical logit transformation on the proportion of fixations to make the outcome measure unbounded in order to meet

model assumptions (Fraundorf, 2022). Ordered factor coding was used to model the effect of New information status (for fluent trials), Disfluent fluency status (for given trials), and the difference between the given versus new difference for disfluent versus fluent conditions on the proportion of looks to the competitor object, with a reference level of Given-Fluent trials. Looks to the competitor represent how much competition is generated by the cohort competitor across conditions and is the dependent variable used in previous analyses of this task (Arnold et al., 2004). Factors were entered as both parametric and ordered smooth terms. Factor smooth terms for participants and participants by each of the factor terms were used to model differences in responses to each factor type by participant, similar to including random slopes for information status and fluency, and their interaction by participants in linear mixed effects models. The same process for determining the rho value in the Simple Prediction Task was used.

To assess whether there was a difference in the proportion of looks to the competitor over time between the healthy control and TBI groups, a second set of models with the same specifications but with an ordered factor of Group was used. Random terms were included for participants and the either fluency or information status (see below) by participants. Healthy controls were used as the reference level. When there was an interaction with group, the model was re-leveled so that the information status or fluency contrast could also be examined for the TBI group. One model examined only Fluent trials and assessed the effect of information status (Given reference level), group, and their interaction. The other model examined only New trials and assessed the effect of fluency (Fluent reference level), group and their interaction. New trials were chosen to examine this effect because the proportion of looks to the competitor is expected to be higher overall, allowing for any effects that are present to be easier to detect. Separate models with two-way interactions were used because the meaning of the 3-way interaction between information status, fluency, and group is not relevant to the main research questions. Rather, the effects of interest are differences between groups in using prior discourse information (information status) and using relevant cues about the speaker to change predictions (fluency).

To address possible explanations for group differences in disfluency effects (see below),

additional exploratory models were constructed. Two models examined the remaining contrasts not addressed in the planned analyses: the disfluency contrast between the two groups on Given trials and the given/new contrast between the two groups on Disfluent trials, using the same model structure and parameters as the model described above. One model examined the disfluency contrast for new trials in regards to looks to the other object rather than looks to the competitor, using the same model structure described above. Finally, two models addressed whether there were group differences in the contrast of looks to the competitor on trials where the competitor was expected by Healthy Adults (Fluent-Target New, Disfluent – Target Given) versus trials where the competitor was unexpected (Fluent-Target Given, Disfluent – Target New). One of these models used expected trials as the reference level and the other used unexpected trials as the reference level. The model structure was the same as the other models used to examine group differences, but with an ordered factor of expected/unexpected, and used all trials rather than a subset.

An additional set of covariate models examining the TBI group was constructed for each of the primary/planned analysis models (information status effect, fluency effect) listed above, by adding interaction terms as described in the Simple Prediction Task. No covariate models were included for the pupillometry analysis because the main effect of interest was not significant even for the healthy control group (see below). Specific model specifications can be found in Appendix 2.

Pupillometry Analysis

To address differences in cognitive effort used when initial predictions must be changed (i.e., unpredicted target trials: Fluent condition, New target; Disfluent condition, Given target) versus when expectations are confirmed (i.e., predicted target trials: Fluent condition, Given target; Disfluent condition, New target), differences in the task evoked pupil response were examined during the time period from target onset until the average time of click and drag response onset. This time period was selected because it reflects when the participants can generate predictions based on the target onset, while avoiding possible changes in pupil size due to initiating a clicking response. The information available

during this time is equivalent between fluent and disfluent trials. Note that while participants are expected to need to change where the majority of their fixations occur on unexpected target trials, participants rarely fixate only on one object during a given time period, so the total amount of looking around the screen is not necessarily correlated with whether the target is expected or unexpected. Baseline pupil size was defined as the average pupil size from the start of Instruction 2 until the onset of the determiner ("Now move"), which is the same regardless of whether the trials were fluent or disfluent. Prior to modeling, the data were pre-processed by identifying samples that occurred during blinks or saccades and those that were outliers. Outliers were defined as samples where the change in pupil size between previous or subsequent samples was above the 95th percentile for change in pupil size or when the pupil size was smaller than the 5th percentile of all pupil sizes by participant. Data were downsampled into 50 millisecond bins. Bins including more than 50% of samples in a blink or saccade or with any outliers were excluded. Trials with more than 45% of the data excluded for these reasons in the baseline or the analysis window were excluded from analysis (Johns et al., 2024). GAMMs were used to assess the effect of trial type (Change or Confirm Prediction) on the pupil size over time. Ordered factor coding was used to model the effect of trial type (reference level Confirm), group (reference level Healthy Controls), and the interaction of trial type and group. These factors were entered as parametric terms and tensor product smooths. Additional factor smooth terms were included to account for baseline pupil size and gaze position, as participants were looking away from center, which affects pupil size measurement, and at different locations on the screen during the trial (Gagl et al., 2011). Random smooths were included for participant, items, and the random slope of trial type by participants and items. The same process as described above was used to determine the rho value to account for autocorrelation. Given the current modeling approach and use of explicit modeling of the baseline, the pupil response is represented by arbitrary units (a.u.).

Experiment 2 Results

Healthy Control Group Performance

Evaluating only the healthy adult participants demonstrated that the Discourse Task effectively replicated the results of Arnold et al. (2004) and that the effects of interest can be detected with the current modeling approach (Figure 12A). Healthy adult participants had an overall higher proportion of looks to the competitor on Fluent trials when the target was new/competitor was given (Table 14, Parametric Terms). There was also a difference in the shape of the looks to competitor response over time for the information status contrast (Table 14, Smooth Terms). Figure 13 shows that the time course of this effect from 224-733 milliseconds after the target word onset, which is consistent with Arnold et al.'s critical time window of 200-600 milliseconds. Figure 13 shows that healthy participants also had significantly fewer looks to the competitor on Fluent-Given Trials than Fluent-New trials later in the time window, from 939-1000 milliseconds post critical word onset. While there was no significant main effect related to fluency, there was a significant interaction (Table 14, Parametric Terms; Figure 14) between 42-648 milliseconds post critical word onset. This interaction reflects that the New-Given difference for Disfluent trials was overall smaller than the same difference for Fluent trials, which can be accounted for by the relative increase in looks to the competitor on Given trials and relative decrease in looks to the competitor on New trials in the Disfluent condition (Figure 12A). Overall, these results are consistent with the original finding that participants predict that they will hear given information again on fluent trials and that this prediction is reversed when disfluencies occur, allowing the inference that new information is upcoming.

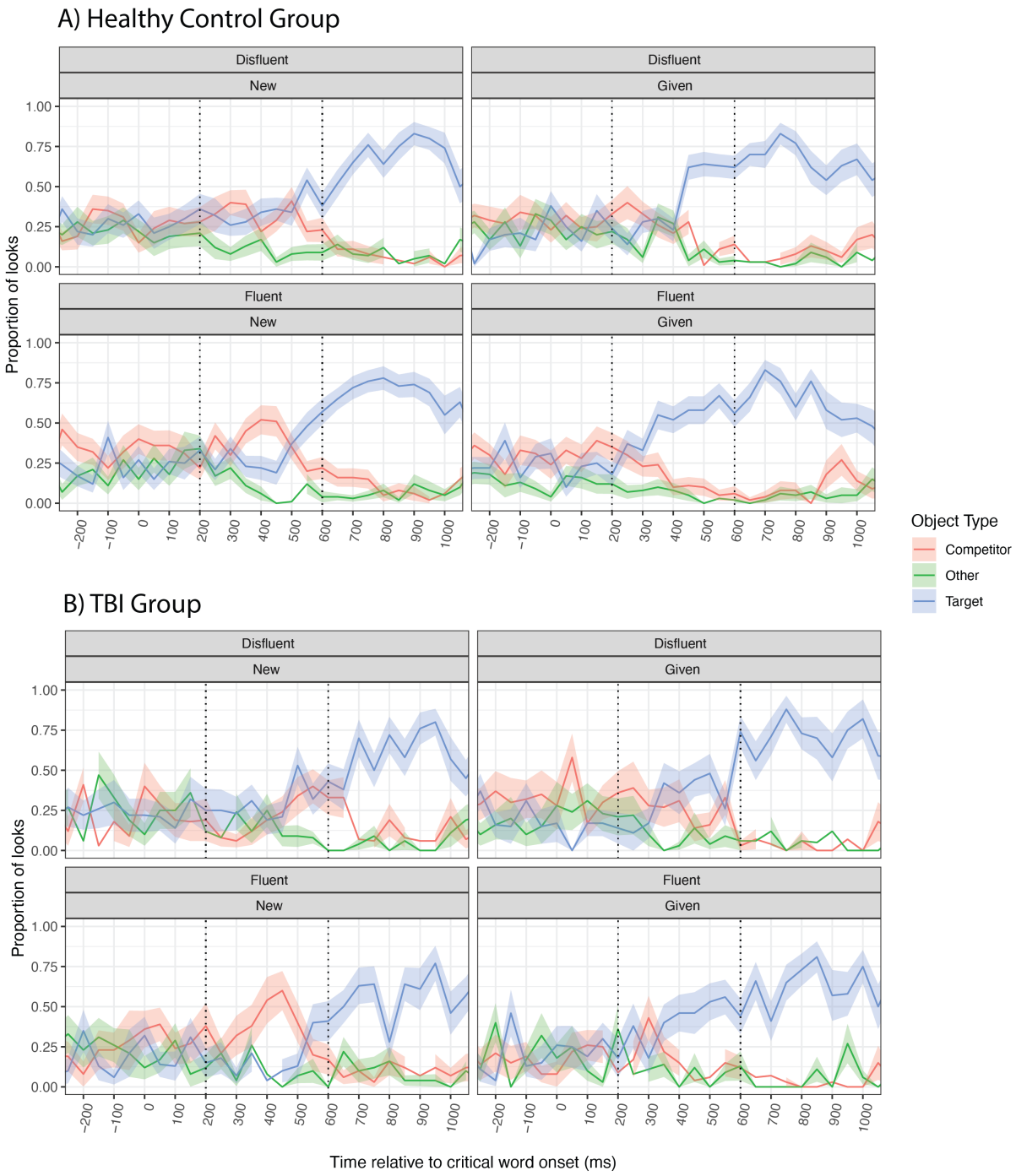


Figure 12. Average proportion of fixations (looks) to each object type per total fixations during each 50-millisecond time slice. Times represent the time in milliseconds since the onset of the critical word (target) within the second instruction. Target = object named, Competitor = unnamed cohort competitor, Other = non-cohort competitor item. Each panel shows responses for trials separated by fluency condition and the information status of the target. The information status of the competitor is the opposite of the target status (i.e., if the target is given the competitor is new and vice versa). Shaded regions represent standard error. Dotted lines show the time window of 200-600 millisecond post target onset, which is the time region of effects described in Arnold et al. (2004). Panel A shows the results for the healthy adult participants (N=48). Panel B shows the results for the TBI group participants (N=24).

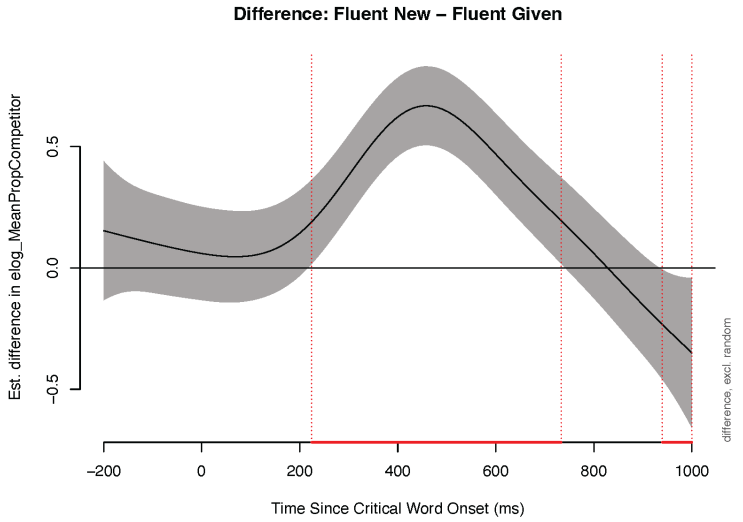


Figure 13. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Fluent New – Fluent Given trials for the Healthy Control Group. Given/New represent the information status of the target. Positive values indicate more looks to the competitor on Fluent New than Fluent given trials. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

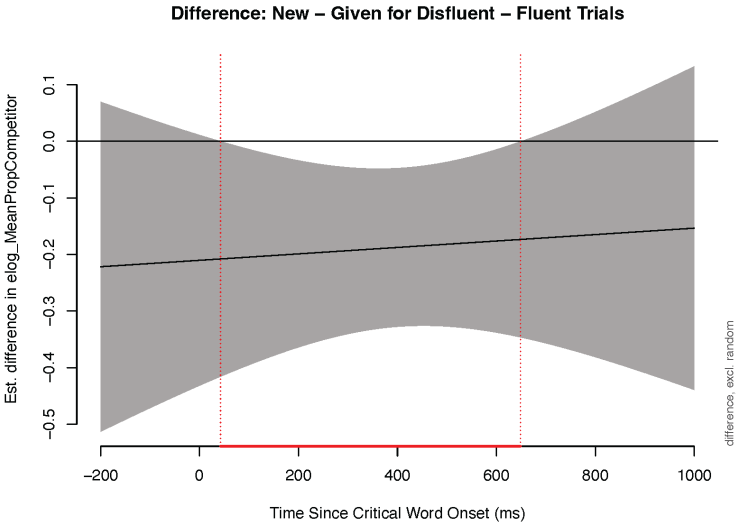


Figure 14. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor for the interaction of information status and fluency for the Healthy Control Group. Given/New represent the information status of the target. Positive values indicate a larger difference between New-Given trials in the Disfluent condition than the Fluent condition. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two sets of conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.83	0.04	-22.53	<0.001
IsDisfluent	0.07	0.05	1.39	0.17
IsNew	0.23	0.05	4.46	<0.001
IsDisfluentIsNew	-0.19	0.07	-2.62	0.009

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	5.25	6.26	12.53	<0.001
Time:IsDisfluent	1.00	1.00	0.009	0.92
Time:IsNew	5.55	6.58	9.41	<0.001
Time:IsDisfluentIsNew	1.00	1.00	0.07	0.79

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Time, Participant	44.75	479.00	0.14	<0.001
Time, Participant:IsDisfluent	40.30	479.00	0.12	<0.001
Time, Participant:IsNew	25.87	479.00	0.07	<0.001
Time, Participant:IsDisfluentIsNew	45.86	473.00	0.13	<0.001

Table 14. Model outputs for the model examining the effect of information status (Given/New Target) and fluency (Disfluent, Fluent) on the empirical logit of the proportion of looks to the competitor in the Healthy Control group.

Group Comparison – Use of Prior Discourse Information to Inform Predictions

The model addressing differences between the TBI group and Healthy Control group in their ability to use prior discourse information to make predictions revealed similar performance across the groups. The parametric (overall height) and smooth (response shape) effect of new information status followed the same pattern as the analysis of only the Healthy Controls, namely that there were more looks to the competitor on Fluent trials when the target was new/competitor was given than when the target was

given/competitor was new between 261-685 milliseconds after the critical word onset (Table 15; Figure 15). There was no significant group by information status interaction, suggesting that the pattern of performance on fluent trials was the same for the TBI group. As shown in Figure 16, there was a significant parametric effect of group, such that the TBI group had overall fewer looks to the competitor on Fluent-Given trials than the Healthy Control group, but not a difference in the shape of the response curve, between -200 to 91 milliseconds relative to critical word onset (Table 15).

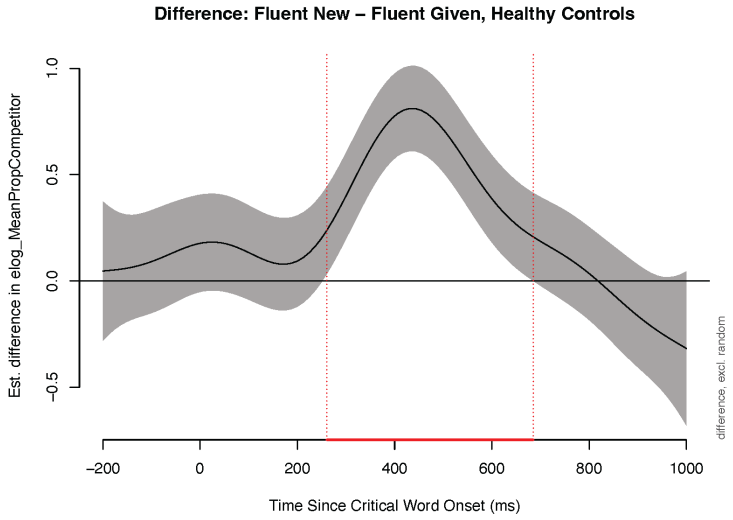


Figure 15. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Fluent New – Fluent Given trials for the Healthy Control group. Given/New represent the information status of the target. Positive values indicate more looks to the competitor on Fluent New than Fluent given trials. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

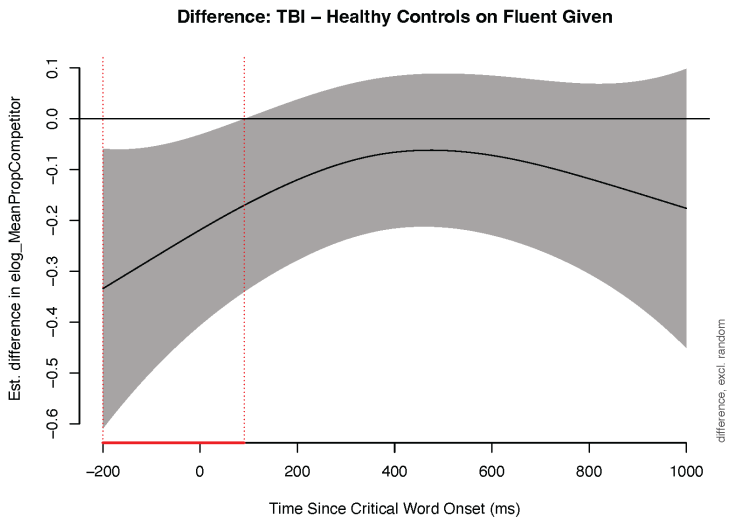


Figure 16. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Fluent Given trials for the TBI group - Healthy Control group. Given/New represent the information status of the target. Positive values indicate more looks to the competitor in the TBI group. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two groups. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.83	0.04	-22.25	<0.001
IsTBI	-0.14	0.06	-2.19	0.03
IsNew	0.23	0.06	3.63	<0.001
IsNewIsTBI	0.06	0.11	0.51	0.61

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	4.75	5.74	11.01	<0.001
Time:IsTBI	2.09	2.56	1.63	0.20
Time:IsNew	6.79	7.78	8.15	<0.001
Time:IsNewIsTBI	1.00	1.00	0.14	0.71

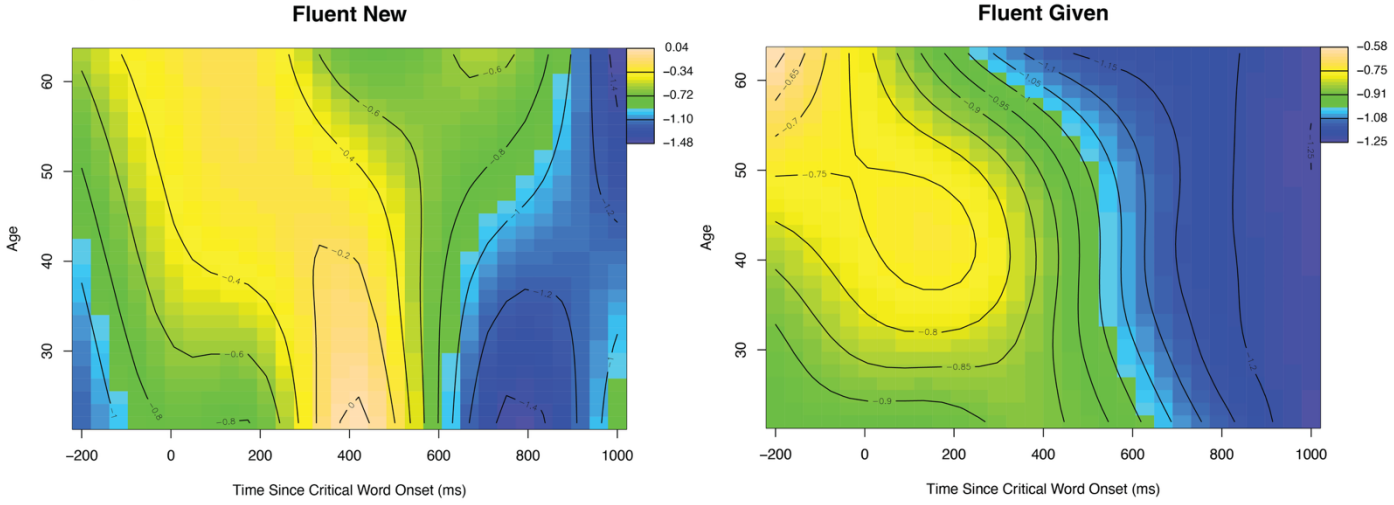
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	51.80	718.00	0.10	<0.001
Time, Participant:IsNew	79.27	715.00	0.19	<0.001

Table 15. Model outputs for the model examining the effect of information status (Given/New Target) and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor for fluent trials only.

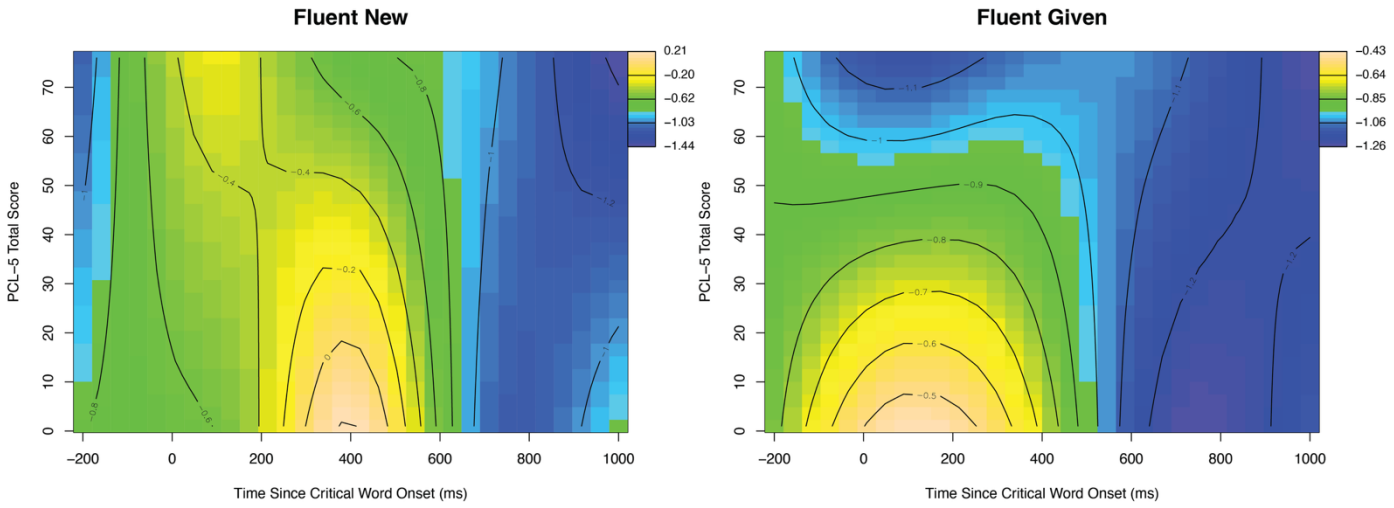
Effect of Covariates on Use of Prior Discourse Information Within the TBI Group

There were significant interactions between information status and age, PTSS, and neurobehavioral symptoms in the TBI group (Table 16). Figure 17A shows that participants across ages considered the competitor prior to critical word onset and approximately 200-600 milliseconds after on trials when the target was given/competitor was new (i.e., the expectation was for the competitor item not to be mentioned). On these trials, older participants looked to the competitor more than younger participants. On trials when the target was new/competitor was given, younger participants in the TBI group followed the pattern expected based on prior experiments with younger adults and the current Healthy Control group, namely having a higher proportion of looks to the competitor at approximately 200-600 milliseconds post critical word onset when the competitor was given in the prior instruction. Older adults showed fewer looks to the competitor during this time period. Figure 17B shows that as PTSS severity increased, participants had a decreasing proportion of looks to the competitor on both target new/competitor given and target given/competitor new trials. These effects occurred within the same time period as for participants with lower PTSS severity (Figure 17B). There were some earlier looks to the competitor on target/given competitor/new trials for participants with higher levels of PTSS between 0-200 milliseconds. A similar pattern of results for participants with higher levels of neurobehavioral symptoms occurred as for participants with higher levels of PTSS (Figure 17C).

A) Age



B) Post-traumatic Stress Symptom Severity



C) Neurobehavioral Symptom Severity

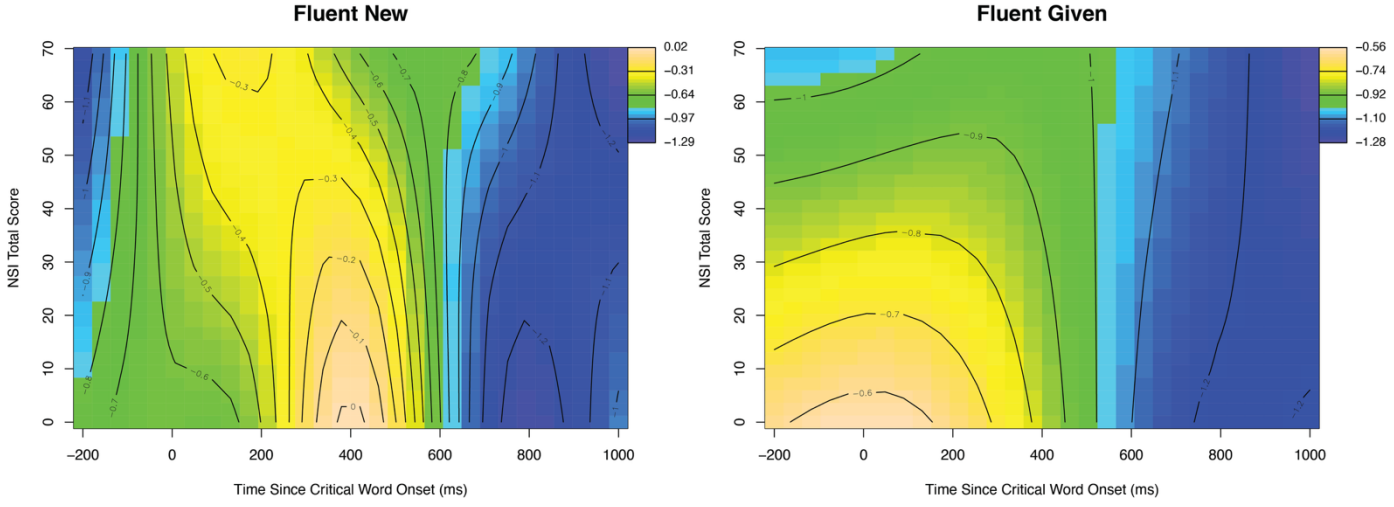


Figure 17. Heatmaps showing model estimated empirical logit of the proportion of looks to the competitor on Fluent Given or New trials for the TBI group as a function of continuous covariate factors: A) Age in years, B) Post-traumatic stress symptom severity as measured by the PCL-5 total score, C) Neurobehavioral symptom severity as measured by the NSI total score. Given/New represent the information status of the target. Warmer colors represent more looks to the competitor. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction.

A) Age

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.97	0.05	-21.53	<0.001
IsNew	0.28	0.11	2.60	0.009

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, Age	8.72	10.93	1.86	0.045
Time, Age:IsNew	10.58	12.55	2.59	0.002

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Age, Participant	4.70	22.00	0.30	0.03
Time, Participant	7.61	238.00	0.04	0.04
Time, Participant:IsNew	33.67	235.00	0.43	<0.001

B) Post-traumatic Stress Symptoms

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.97	0.05	-21.75	<0.001
IsNew	0.28	0.11	2.57	0.01

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, PCL5	7.14	8.56	2.97	0.002
Time, PCL5:IsNew	9.63	11.45	2.08	0.02

Random Effects

Term	edf	Ref.df	F Statistic	P Value
PCL5, Participant	2.44	22.00	0.13	0.05
Time, Participant	11.41	238.00	0.06	0.02
Time, Participant:IsNew	36.19	235.00	0.45	<0.001

C) Neurobehavioral Symptoms

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.97	0.05	-21.73	<0.001
IsNew	0.28	0.11	2.55	0.01

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, NSI	5.87	7.01	2.70	0.008
Time, NSI:IsNew	10.42	12.84	2.01	0.02

Random Effects

Term	edf	Ref.df	F Statistic	P Value
NSI, Participant	2.36	22.00	0.12	0.06
Time, Participant	10.93	238.00	0.05	0.02
Time, Participant:IsNew	36.35	235.00	0.45	<0.001

Table 16. Model outputs for the model examining the effect of information status (Given/New Target) and A) Age, B) Post-traumatic stress symptom severity, and C) Neurobehavioral symptom severity on the empirical logit of the proportion of looks to the competitor for fluent trials only and for the TBI group only.

Group Comparison – Use of Speaker Cues to Alter Predictions

The model addressing group differences in the ability to use disfluencies as a cue in order to change expectations about whether upcoming information will be given or new revealed differences in the pattern of responses between the TBI and Healthy Control groups. As shown in Figure 18, there was a significant parametric effect of fluency, demonstrating overall fewer looks to the competitor on New Target/Given Competitor trials when there was a disfluency versus when the trial was fluent (Table 17). This effect was consistent over the 103-527 millisecond post critical word onset time period for healthy adults. Releveling the model, there was not a significant parametric fluency effect for the TBI group (Table 18). This result is in line with the results in Arnold et al. (2004) and suggests that healthy adults interpreted disfluencies to mean that given information was less likely to occur. Although there was not an overall height difference between fluent and disfluent conditions for the TBI group, there was a difference in the pattern of looks to the competitor over time between groups. Figure 19 shows a significant smooth term for the interaction of fluency and group due to a greater difference between looks to the competitor on disfluent versus fluent trials in the TBI group than in the Healthy Control group within the time window of 224-394 milliseconds post critical word onset (Table 17)[3]. This unexpected effect appears to be driven primarily by differences between groups in the Disfluent-New condition, where individuals in the TBI group continued to look at the competitor at chance or less during this time, while individuals in the Healthy Control group had already begun to look at the competitor at greater than chance levels. Both the raw data and model suggest that the two groups' patterns of looks towards the competitor were similar 394 milliseconds post critical word onset and onward (Figure 12; Figure 19).

[3] The interaction term in the releveled model was not significant in the re-leveled model with the TBI group set as the reference level. This may be due to the differences in group size and variability difference in the TBI group versus the healthy group, and how the model addresses variability based on the reference smooth (Wood, 2017). The interaction from the healthy group reference level is interpreted. The primary purpose of the re-leveled model was to examine the value of the relevant contrasts for the TBI group.

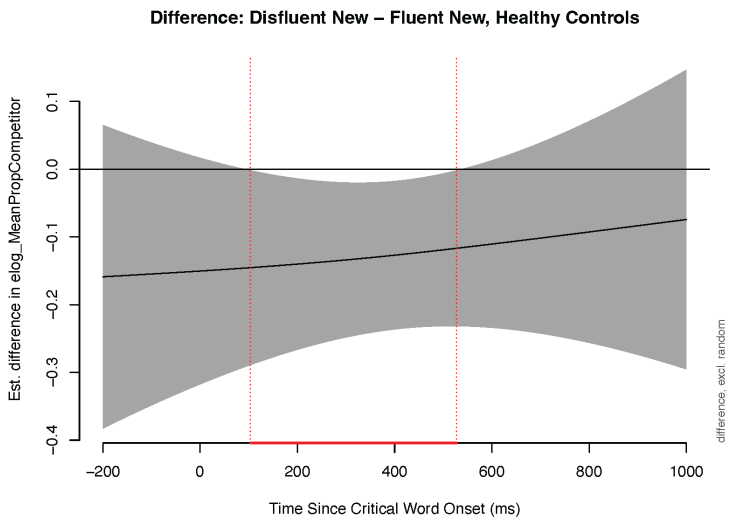


Figure 18. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Disfluent New – Fluent New trials for the Healthy Control group. Given/New represent the information status of the target. Positive values indicate more looks to the competitor on Disfluent New than Fluent New trials. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

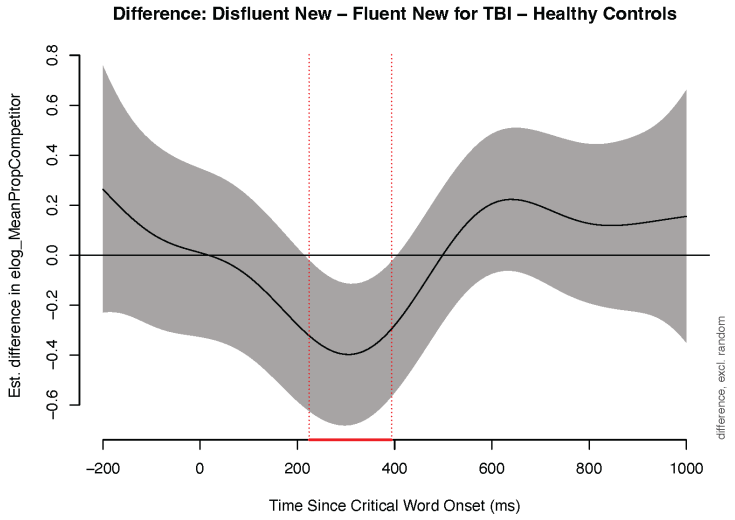


Figure 19. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor for the interaction of fluency and group for New trials. Given/New represent the information status of the target. Positive values indicate a more positive difference between Disfluent-Fluent trials in the TBI condition than the Healthy Control group (note: the Disfluent-Fluent contrast is expected to be negative, so more positive differences are smaller in magnitude). Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two sets of conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms				
Term	**Parameter Estimate**	**Standard Error**	**t Statistic**	**P Value**
Intercept	-0.61	0.05	-13.61	<0.001
IsTBI	-0.08	0.08	-1.04	0.30
IsDisfluent	-0.12	0.05	-2.26	0.02
IsDisfluentIsTBI	0.00	0.09	0.04	0.97
Smooth Terms				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time	6.54	7.60	19.78	<0.001
Time:IsTBI	1.03	1.06	1.15	0.27
Time:IsDisfluent	1.08	1.15	0.23	0.74
Time:IsDisfluentIsTBI	4.86	5.87	2.80	0.02
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	89.93	718.00	0.25	<0.001
Time, Participant:IsDisfluent	51.06	712.00	0.09	<0.001

Table 17. Model outputs for the model examining the effect of fluency (Disfluent, Fluent) and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor for New trials only.

Parametric Terms				
Term	**Parameter Estimate**	**Standard Error**	**t Statistic**	**P Value**
Intercept	-0.70	0.06	-10.98	<0.001
IsHealthy	0.08	0.08	1.05	0.29
IsDisfluent	-0.12	0.07	-1.56	0.12
IsDisfluentIsHealthy	0.01	0.09	-0.07	0.94
Smooth Terms				

Term	edf	Ref.df	F Statistic	P Value
Time	6.52	7.59	13.60	<0.001
Time:IsHealthy	1.00	1.00	1.38	0.24
Time:IsDisfluent	2.49	3.07	1.74	0.15
Time:IsDisfluentIsHeatlhy	1.00	1.00	0.20	0.66
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	88.83	718.00	0.25	<0.001
Time, Participant:IsDisfluent	53.94	712.00	0.10	<0.001

Table 18. Model outputs for the re-leveled model examining the effect of fluency (Disfluent, Fluent) and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor for New trials only.

Group Comparisons – Exploratory Analyses to Examine Performance on Disfluent-New

To further explore whether participants in the TBI group also showed differences in use of the disfluency cue in Given trials or whether the effect was specific to New trials, the same fluency by group interaction was examined for the Given trials only. There was no disfluency by group interaction for the Given trials and no main effect of disfluency in smooth or parametric effects. There was a significant parametric effect of group, as the TBI group had a smaller proportion of looks to the competitor from -200 to 103 milliseconds (Table 19). This parametric effect reflects the same effect detected the given-new comparison between groups for fluent trials (Figure 16; though with slight differences in the time of significant effects because of the different model structure), suggesting TBI group participants' looks to the competitor were generally lower at baseline and early in the trial on Fluent-Given trials.

Parametric Terms				
Term	**Parameter Estimate**	**Standard Error**	**t Statistic**	**P Value**
Intercept	-0.82	0.04	-22.14	<0.001

IsTBI	-0.13	0.06	-2.09	0.04
IsDisfluent	0.06	0.05	1.09	0.28
IsDisfluentIsTBI	0.10	0.10	1.16	0.25

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	5.61	6.64	12.47	<0.001
Time:IsTBI	2.66	3.21	2.06	0.09
Time:IsDisfluent	1.21	1.36	0.11	0.90
Time:IsDisfluentIsTBI	1.00	1.00	2.31	0.13

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Time, Participant	111.01	718.00	0.25	<0.001
Time, Participant:IsDisfluent	85.98	707.00	0.18	<0.001

Table 19. Model outputs for the model examining the effect of fluency (Disfluent, Fluent) and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor for Given trials only.

Similarly, to examine whether unexpected group differences in the disfluent/fluent contrast could be explained by some feature of performance on Disfluent trials, group differences in the effect of new information status was examined for Disfluent trials only. There were no significant parametric effects. There was a significant smooth term for the effect of new information status (Table 20). Figure 20 shows that healthy participants looked at the competitor a higher proportion of the time when the target was new/competitor was given than when the target was given/competitor was new between 346-636 milliseconds. This result is consistent with the expectation that listeners should infer that they will hear given information. Releveling the model to examine the information status effect for the TBI group demonstrated a similar effect slightly later, between 467 and 721 milliseconds, as shown in Figure 21 and Table 21 Participants in the TBI group also demonstrated an effect of information status earlier in the trial, where looks to the competitor were lower on Disfluent-New than Disfluent-Given trials from -115 to

212 milliseconds (Figure 21). The results of the disfluency effect in the other models demonstrate that this expectation is diminished relative to Fluent trials, as expected from previous work. There was also a significant smooth term reflecting the interaction of group and information status (Table 20). As shown in Figure 22, early in the trial, from 164-382 milliseconds, the New minus Given difference was more negative for the TBI group than the Healthy Control Group and was more positive for the TBI group later in the trial, from 563-661 milliseconds. These results suggest that the timing of looks to the competitor follows a different pattern in the TBI group than the Healthy Control Group. Visual inspection of the raw data (Figure 12) again suggests that this difference is primarily due to the early difference in performance in the Disfluent-New condition, where the proportion of looks to the competitor was lower for the TBI group than for Healthy Controls.

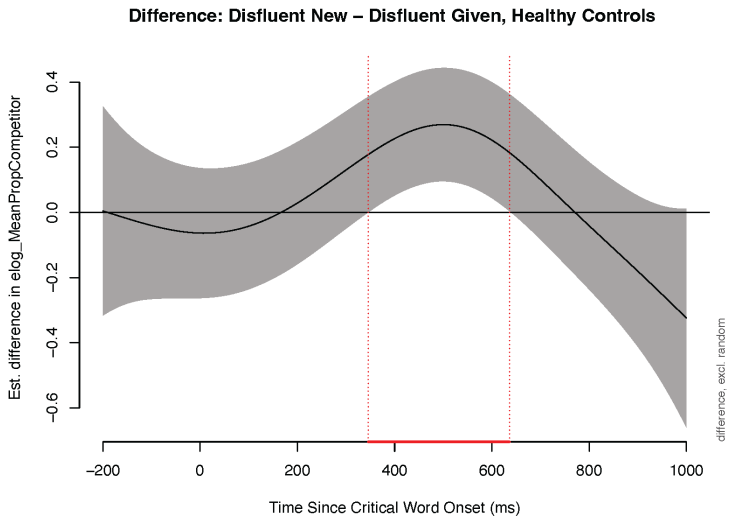


Figure 20. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Disfluent New – Disfluent Given trials for the Healthy Control group. Given/New represent the information status of the target. Positive values indicate more looks to the competitor on Disfluent New than Disfluent Given trials. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

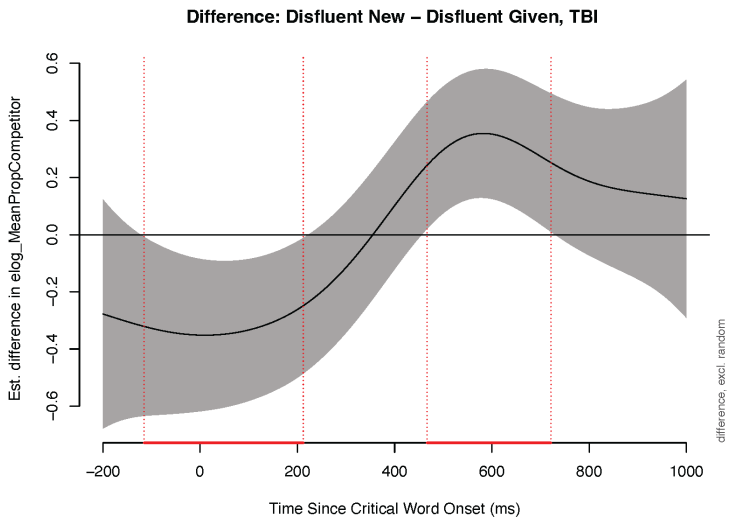


Figure 21. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Disfluent New – Disfluent Given trials for the TBI group. Given/New represent the information status of the target. Positive values indicate more looks to the competitor on Disfluent New than Disfluent Given trials. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

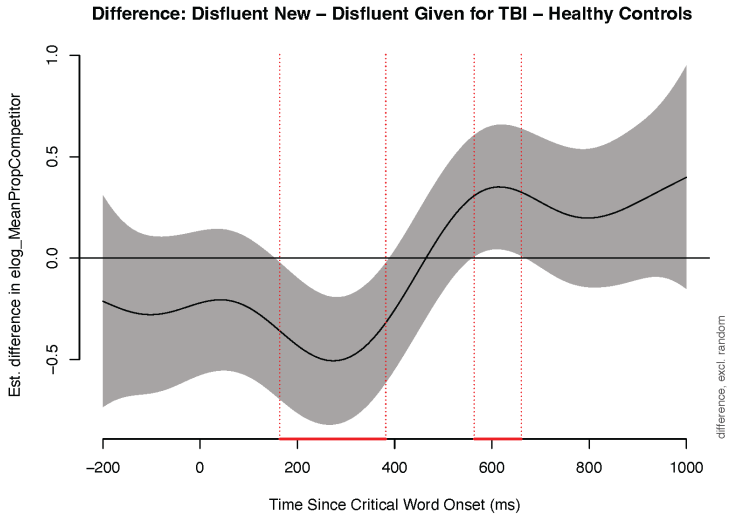


Figure 22. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor for the interaction of information status and group for Disfluent trials. Given/New represent the information status of the target. Positive values indicate a more positive difference between New-Given trials in the TBI condition than the Healthy Control group. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two sets of conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.77	0.04	-20.92	<0.001
IsTBI	-0.04	0.06	-0.66	0.51
IsNew	0.04	0.06	0.77	0.44
IsNewIsTBI	-0.03	0.10	-0.33	0.74

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	4.99	5.94	8.00	<0.001
Time:IsTBI	1.00	1.00	1.58	0.21
Time:IsNew	3.66	4.44	3.08	0.01
Time:IsNewIsTBI	5.51	6.52	3.13	0.01

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Time, Participant	129.32	718.00	0.31	<0.001
Time, Participant:IsNew	75.35	712.00	0.15	<0.001

Table 20. Model outputs for the model examining the effect of information status (Given/New Target) and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor for Disfluent trials only.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.81	0.05	-15.47	<0.001
IsHealthy	0.04	0.06	0.66	0.51
IsNew	0.01	0.08	0.12	0.91
IsNewIsHealthy	0.03	0.10	0.35	0.73

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	3.49	4.19	6.84	<0.001
Time:IsHealthy	4.20	5.02	1.88	0.09
Time:IsNew	4.60	5.52	3.78	0.001
Time:IsNewIsHealthy	1.77	2.14	3.12	0.04
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	128.71	718.00	0.31	<0.001
Time, Participant:IsNew	77.18	712.00	0.15	<0.001

Table 21. Model outputs for the re-leveled model examining the effect of information status (Given/New Target) and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor for Disfluent trials only.

To examine whether the disfluency effect in the new trials was specific to predictions about the target versus the competitor or whether participants in the TBI group adopted a general "wait-and-see" approach by considering multiple referent options even after the target word onset, the proportion of looks to one of the non-cohort objects on given trials was examined. There were no significant effects of disfluency, group, or their interaction, suggesting that hearing a disfluency specifically affected predictions about the relative likelihood of cohort items (Table 22).

Parametric Terms				
Term	**Parameter Estimate**	**Standard Error**	**t Statistic**	**P Value**
Intercept	-0.97	0.03	-32.29	<0.001
IsTBI	0.00	0.05	0.06	0.95
IsDisfluent	0.02	0.05	0.37	0.71
IsDisfluentIsTBI	-0.01	0.09	-0.13	0.90
Smooth Terms				

Term	edf	Ref.df	F Statistic	P Value
Time	4.15	5.02	7.46	<0.001
Time:IsTBI	1.00	1.00	0.00	0.97
Time:IsDisfluent	1.00	1.00	0.14	0.71
Time:IsDisfluentIsTBI	1.62	1.93	0.53	0.62
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	92.30	718.00	0.19	<0.001
Time, Participant:IsDisfluent	85.88	712.00	0.19	<0.001

Table 22. Model outputs for the model examining the effect of fluency (Disfluent, Fluent) and group (Healthy, TBI) on the empirical logit of the proportion of looks to the other (non-competitor, non-target) object for New trials only.

The previous exploratory analyses demonstrated that differences in looking behavior on Disfluent-New trials were specific to cohort competitors (cf. no group differences in looks to Other) but were not due to a general difference in use of disfluencies as a cue to adjust predictions about upcoming referents (cf. no group differences in disfluency effect, Given trials). Thus, the only remaining difference between Disfluent-New and Disfluent-Given trials is whether the competitor should be the initially expected item or not. The model examining the effect of whether the competitor item was expected or unexpected across all trial types, with a reference level of unexpected trials for Healthy Controls, showed parametric effects of expected status (Figure 23; Table 23) and of TBI (Figure 24; Table 23). Figure 23 demonstrates that the proportion of looks to the competitor was higher when the competitor was expected than when it was unexpected between 139-539 milliseconds for healthy adults, in line with all of the prior results. As shown in Figure 24, the parametric effect of TBI reflected that when the competitor was unexpected (reference level) individuals in the TBI group had a lower proportion of looks to the competitor from -200 to 382 milliseconds post critical word onset than Healthy Controls. This time period aligns with both parametric effects on Fluent-Given trials suggesting overall fewer looks to the competitor

during the baseline and the later smooth effect showing a lower proportion of looks to the competitor that accounted for the difference from Healthy Controls in the Disfluent-New condition. There was no interaction and there were no significant smooth effects. Re-leveling the model such that the reference level was expected trials demonstrated that there was no significant parametric effect of TBI on expected trials (Table 24). There was a significant parametric term reflecting the same difference between expected and unexpected trials as noted above. There were no other significant parametric terms and no significant smooth terms.

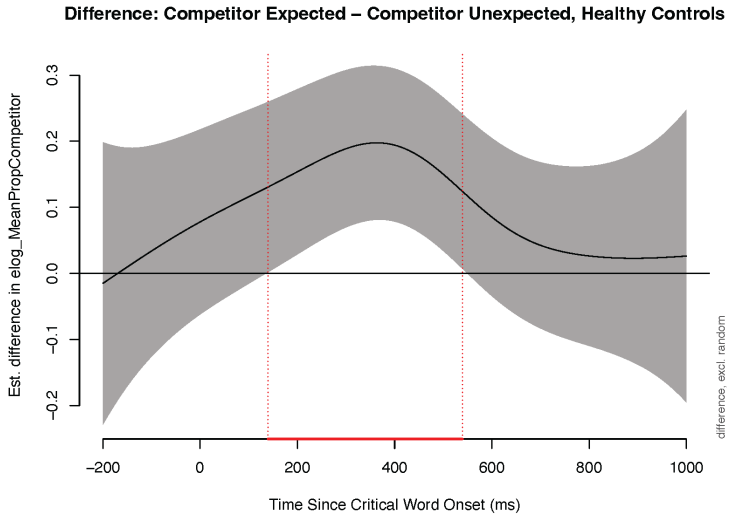


Figure 23. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Competitor Expected (Fluent-New and Disfluent-Given) – Competitor Unexpected (Fluent-Given and Disfluent-New) trials for the Healthy Control group. Given/New represent the information status of the target. Positive values indicate more looks to the competitor on Expected than Unexpected trials. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

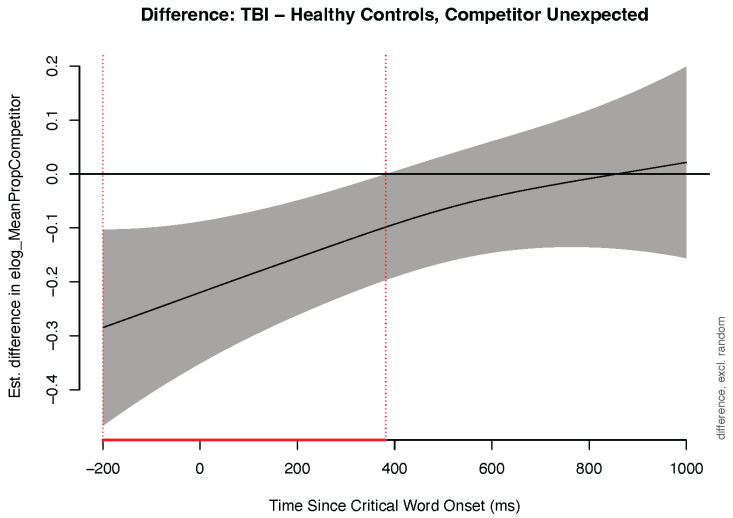


Figure 24. Difference curve showing model estimated empirical logit of the proportion looks to the competitor for the TBI – Healthy Control group for Competitor Unexpected (Fluent-Given and Disfluent-New). Given/New represent the information status of the target. Positive values indicate more looks to the competitor for the TBI group. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.78	0.03	-29.43	<0.001
IsTBI	-0.11	0.04	-2.35	0.02
IsExpected	0.01	0.04	2.11	0.04
IsExpectedIsTBI	0.05	0.07	0.70	0.48

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	5.25	6.32	17.05	<0.001
Time:IsTBI	1.49	1.79	2.47	0.06
Time:IsExpected	3.30	4.05	1.65	0.16
Time:IsExpectedIsTBI	1.00	1.00	1.99	0.16

Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	35.36	718.00	0.06	<0.001
Time, Participant:IsExpected	92.66	718.00	0.21	<0.001

Table 23. Model outputs for the model examining the effect of whether the competitor is expected and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor.

Parametric Terms				
Term	**Parameter Estimate**	**Standard Error**	**t Statistic**	**P Value**
Intercept	-0.69	0.03	-23.60	<0.001
IsTBI	-0.05	0.05	-1.07	0.29
IsUnexpected	-0.09	0.04	-2.58	0.01
IsUnexpectedIsTBI	-0.05	0.07	-0.81	0.42
Smooth Terms				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time	5.87	6.99	20.53	<0.001
Time:IsTBI	1.63	1.98	0.50	0.62
Time:IsUnexpected	1.96	2.43	0.98	0.39
Time:IsUnexpectedIsTBI	3.23	3.99	1.78	0.13
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	84.94	718.00	0.20	<0.001
Time, Participant:IsUnexpected	2.05	718.00	0.00	0.06

Table 24. Model outputs for the re-leveled model examining the effect of whether the competitor is expected and group (Healthy, TBI) on the empirical logit of the proportion of looks to the competitor.

Effect of Covariates on Use of Prior Discourse Information Within the TBI Group

In the covariate analyses, there were significant interactions between fluency (examining new trials only, when the competitor is given) and age and PTSS in the TBI group (Table 25). As shown in Figure 25A, looking patterns were relatively uniform across ages in the Disfluent-New condition, when the competitor should be unexpected. On Fluent-New trials, when the competitor was given, older participants in the TBI group made fewer predictive looks to the competitor, consistent with the previous analyses. In regards to PTSS, participants with higher PTSS severity showed earlier looks to the competitor in the Fluent condition, when the competitor was expected, as also demonstrated in the given/new model. In the Disfluent condition, participants in the TBI group with lower PTSS severity showed fewer looks to the competitor across the length of the trial, but particularly between approximately 0-300 and 700-1000 milliseconds post critical word onset. This demonstrates fewer looks to the competitor when it should be unexpected for participants who had low levels of PTSS, while those with higher levels of PTSS considered the competitor more (Figure 25B). There was no significant interaction between NSI score and the disfluency effect for participants in the TBI group (Table 25).

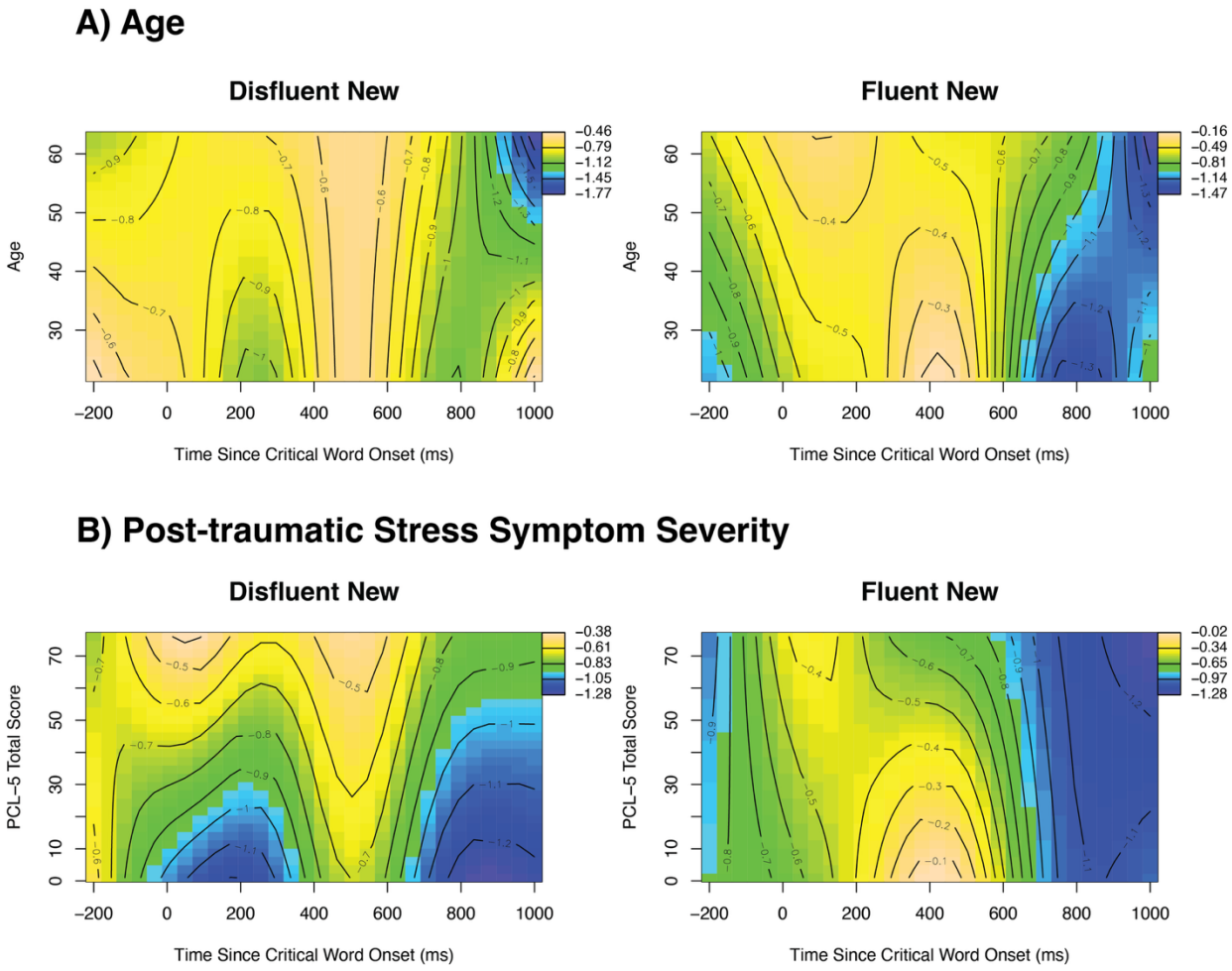


Figure 25. Heatmaps showing model estimated empirical logit of the proportion of looks to the competitor on Disfluent or Fluent New trials for the TBI group as a function of continuous covariate factors: A) Age in years, B) Post-traumatic stress symptom severity as measured by the PCL-5 total score. Given/New represent the information status of the target. Warmer colors represent more looks to the competitor. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction.

A) Age

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.71	0.08	-8.60	<0.001
IsDisfluent	-0.10	0.09	-1.18	0.24

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, Age	10.53	12.51	4.52	<0.001
Time, Age:IsDisfluent	7.46	8.97	2.04	0.03

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Age, Participant	9.26	22.00	0.73	<0.001
Time, Participant	22.97	238.00	0.14	<0.001
Time, Participant:IsDisfluent	10.29	236.00	0.08	<0.001

B) Post-traumatic Stress Symptoms

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.71	0.08	-8.86	<0.001
IsDisfluent	-0.11	0.08	-1.32	0.19

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time, PCL5	9.82	11.72	4.32	<0.001
Time, PCL5:IsDisfluent	7.61	9.27	2.04	0.04

Random Effects

Term	edf	Ref.df	F Statistic	P Value
PCL5, Participant	9.07	23.00	0.68	<0.001
Time, Participant	21.97	238.00	0.13	<0.001
Time, Participant:IsDisfluent	8.26	236.00	0.06	0.01

C) Neurobehavioral Symptoms

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.71	0.08	-8.94	<0.001
IsDisfluent	-0.11	0.08	-1.27	0.21
Smooth Terms				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, NSI	10.27	12.23	4.31	<0.001
Time, NSI:IsDisfluent	7.15	8.60	1.76	0.08
Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
NSI, Participant	8.39	22.00	0.62	<0.001
Time, Participant	20.91	238.00	0.13	<0.001
Time, Participant:IsDisfluent	8.75	236.00	0.06	0.005

Table 25. Model outputs for the model examining the effect of fluency (Disfluent/Fluent) and A) Age, B) Post-traumatic stress symptom severity, and C) Neurobehavioral symptom severity on the empirical logit of the proportion of looks to the competitor for New trials only and for the TBI group only.

Group Comparison – Pupillometry to Examine Effort Related to Model Updating

The model assessing whether there were group differences in pupil size, as a measure of cognitive effort, on trials when the prediction about the target item was confirmed or not revealed no significant effects of trial type and no interaction (Table 26). There was a significant smooth term for group. This effect reflects that pupil size was larger for the TBI group than for the Healthy Control group initially after the critical word onset, but later became smaller for the TBI group than the Healthy Control group on trials when the prediction was unconfirmed and needed to be changed (Figure 26; Table 26). While this effect likely reflects the change in direction of the effect over time, there was no time period during which the difference between TBI and Healthy Control groups was significantly different from zero.

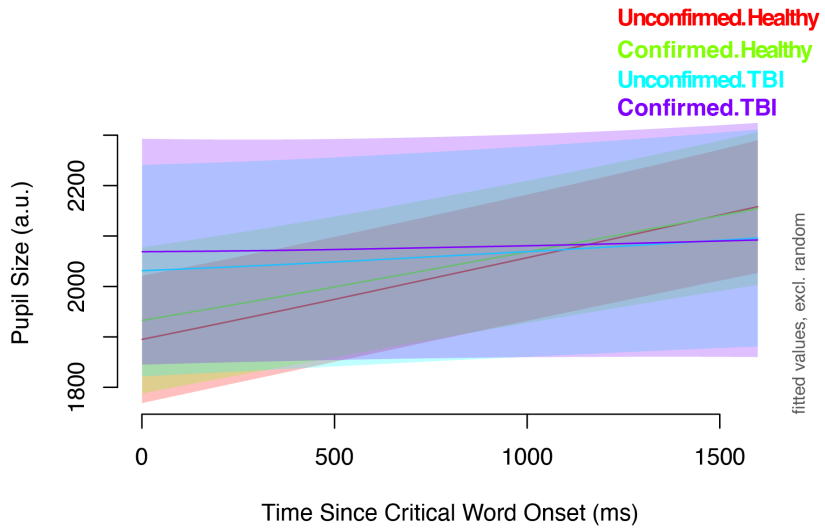


Figure 26. Model estimated pupil size for trials when the prediction that the target would be mentioned was confirmed (Fluent-Given, Disfluent-New) or unconfirmed (Fluent-New, Disfluent-Given), for each group. Time represents the time in milliseconds relative to the onset of the critical word (target) in the second instruction. Shaded areas show the 95% confidence interval for the looks to the competitor.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	2485.83	37.86	65.65	<0.001
IsTBI	54.45	90.63	0.60	0.55
IsConfirmed	20.49	35.65	0.56	0.57
IsConfirmedIsTBI	-29.38	51.93	-0.57	0.57

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Gaze X, Gaze Y	17.84	18.74	25.79	<0.001
Mean Baseline Pupil Size	15.86	17.62	591.75	<0.001
Time	1.31	1.38	40.73	<0.001
Time:IsTBI	1.00	1.00	16.76	<0.001
Time:IsConfirmed	1.00	1.00	0.90	0.34
Time:IsConfirmedIsTBI	1.00	1.00	1.23	0.27

Random Effects				
Term	**edf**	**Ref.df**	**F Statistic**	**P Value**
Time, Participant	318.67	1386.00	1.25	<0.001
Time, Item	138.38	479.00	1.23	<0.001
Time, Participant:IsConfirmed	235.24	1218.00	0.93	<0.001
Time, Item:IsConfirmed	107.78	479.00	0.83	<0.001

Table 26. Model outputs for the model examining the effect of whether the prediction for the target as the mentioned item is confirmed (Fluent-Given, Disfluent-New) or unconfirmed (Fluent-New, Disfluent-Given) and group (Healthy, TBI) on pupil size.

Experiment 2 Discussion

Confirming Task Replication

Analysis of only the Healthy Control group again demonstrated that the results of the original study could be replicated with the current sample, procedures, and analyses. Therefore, each of the effects is interpreted in line with the previous work on healthy adults.

Predicting Given/New Discourse Status

Participants in both the TBI and Healthy Control groups demonstrated an expectation for given information from the previous discourse to be repeated. Both groups had a greater proportion of looks to the competitor when it was given than when it was new, and the pattern of looks over time did not differ across groups. These results suggest that individuals with a history of TBI can use prior discourse information to guide predictions, meaning that they can both remember information over the short discourse exchange and use it to guide linguistic predictions during online language comprehension. This result suggests that it is unlikely that an ability to keep track of what information has already occurred in the conversation and to predict that current information will continue to be referenced is not a primary factor that leads individuals with TBI to experience difficulty in conversations. There was a parametric effect of TBI status, suggesting that individuals with TBI looked at the competitor less than Healthy

Controls on Fluent-Given trials. This effect is also reflected in the exploratory analysis of Fluent-Given versus Disfluent-Given trials. This effect occurred primarily during the baseline period. This could suggest general differences in looking behavior before critical information is available. Alternatively, because information about discourse status was already available before the second instruction, this could also be a reflection of reduced looks to the competitor when the competitor is unexpected (see below). However, the pattern that performance during the trial reached the same level as Healthy Controls suggests that individuals with TBI can use the relevant discourse cues to make predictions, and perhaps are even more committed to these predictions, given the lower baseline looks to the competitor.

Effect of Covariates on Predicting Given/New Discourse Status

There were significant interactions of age, PTSS symptom severity, and neurobehavioral symptom severity with information status across groups. Age and symptom severity measures affected looking behavior differently. Reduced predictive looks to the previously given item in older adults is consistent with both the prior literature (Federmeier et al., 2010; Wlotko et al., 2012) and the results of the Simple Prediction Task. However, it should be noted that this sample included mostly young to middle aged adults. While the large age range in this sample may introduce noise, the small number of participants in the age range where predictive looks to the competitor was reduced (approximately 45 and older) and the fact that group effects in the primary analyses were largely driven by trials in the disfluent condition (not addressed in this analysis), suggests that age difference from the Healthy Control group is not the sole reason for the observed group differences.

Participants with higher levels of PTSS demonstrated fewer looks to the competitor when the competitor was both given and new. However, participants with higher PCL-5 scores still demonstrated more looks to the competitor when it was given, suggesting that they still engaged in predictive processing. This overall reduction in looks to the competitor across trials may account for the parametric effect of TBI in the primary analysis. If so, this would further support the conclusion that history of TBI per se does not affect use of prior discourse information or generating predictions based on given/new

referent status during online language comprehension. There was also an increase in looks to the competitor on target new/competitor given trials for participants with higher levels of PTSS. This early effect, between approximately 0-200 milliseconds post critical word onset, may suggest that participants with higher levels of PTSS predicted that given information (in this case the competitor) would occur faster than individuals with lower levels of PTSS. Note that the presence of the disfluency allows listeners to have information about likely referents earlier than the onset of the target word, suggesting that looks within the time it usually takes to program eye movements after information is available are indeed predictive in nature. Faster use of incoming linguistic information to generate predictions and execute predictive eye movements with higher PTSS may be associated with hyperarousal features of PTSD (American Psychiatric Association, 2013). Alternatively, this could reflect other differences in life experiences that are correlated with being in situations that are likely to cause PTSS. For example, service members who are trained in quick target detection for certain combat situations may also experience trauma from those combat conditions, or exposure to repeated adverse conditions for any participant could lead to learned strategies or differences in approach to what is essentially a target detection task. While the influence of PTSS did not lead to overall group differences in the primary comparison, these results indicate that PTSS can affect online language processing and needs to be accounted for in future studies.

The similar pattern of results between PCL-5 Total scores and NSI Total scores may reflect the high correlation between these two measures (Hoover et al., 2023). The current data cannot differentiate whether the timing effects are due to PTSS per se, which changes how people are likely to respond to the NSI, TBI severity being a common cause of increased symptom reporting on both measures, or other common factors that affect how people behave when completing symptom self-report measures, regardless of the content.

Using Speaker Cues to Inform Predictions

There was a significant group difference in the use of disfluencies to influence whether participants

predicted that the speaker would reference given or new information. On New trials (when the competitor was given information), the difference in looks to the competitor on Disfluent versus Fluent trials was larger for the TBI group than for the Healthy Control group, particularly early in the time post critical word onset. This result is unexpected, as was initially hypothesized that individuals with TBI would have either a smaller difference because they were looking equally at the target and competitor (i.e., not making a prediction) or a difference in the opposite direction of the healthy adults (i.e., because they did not use the disfluency cue and expected to hear given information as in the fluent trials). However, model results and inspection of the raw data suggests that the observed effect was attributable to individuals with TBI having even fewer looks to the competitor early in the disfluent condition than the Healthy Controls (where Healthy Controls were following the expected pattern based on prior work).

One possible interpretation of this effect is that participants in the TBI group were using the disfluency cue *more* quickly than the Healthy Controls. However, the exploratory analysis of the fluency contrast on Given trials between groups revealed no significant effects of disfluency or group. If participants in the TBI group were using the disfluency cues to make an inference that upcoming information would be new earlier than Healthy Controls, they should also show more looks to the competitor earlier relative to critical word onset on Target-Given trials when there was a disfluency. However, there was no difference in looking patterns between groups on Given trials.

Another possible interpretation is that when a disfluency occurs, participants in the TBI group wait until much of the information about the target item is available before committing to any referent, even ones that are inconsistent with the onset of the target word. In this case, they should be more likely to look at the non-cohort items in the Disfluent condition relative to the Fluent condition and this would differ from the pattern focused on cohort competitors demonstrated by Healthy Controls. The exploratory analysis of looks to the Other (i.e., non-cohort) item revealed no significant effects of disfluency or group, suggesting that it is also not the case that individuals with TBI took the disfluency to indicate that a broad wait-and-see approach was warranted.

Ruling out differences related to using disfluency cues and a general wait-and-see approach suggests that the observed differences between the TBI and Healthy Control groups could be consistent with at least two possible alternative explanations. First, while participants in the TBI group did change their pattern of looks to the competitor in response to the disfluency cue, it is possible that the meaning of this cue is different between healthy adults and individuals with TBI. Other known effects of disfluencies in discourse include signaling a need for increased attention to upcoming input, which has been shown to be intact in individuals with TBI (Diachek et al., 2024). While in this case the TBI group's performance would likely reflect more extreme effects with the same pattern as for fluent trials, it is possible that some other interpretation occurred or a combination of attention and prediction related factors interacted in unexpected ways, particularly since the meaning of disfluency cues can change based on listeners' experiences (Bosker et al., 2019).

Given that there is not a clear mapping between alternative interpretations of the disfluency cues and the observed performance, a more likely alternative is that participants with TBI had a similar interpretation of the disfluency cues as the Healthy Controls, but that differences are related to when the competitor is expected versus unexpected. The only other group effects, outside of contrasts including Disfluent-New trials, occurred early in the trial on Fluent-Given trials. Here, participants in the TBI group also showed fewer looks to the competitor. The common feature across these trials is that the competitor should be the unexpected referent based on a combination of prior discourse status and disfluency information. The results of the exploratory analyses of all trials when the competitor is expected versus unexpected based on the prior discourse and the disfluency cues suggests that participants with TBI look at the competitor less when it is not expected. While there was not a significant interaction, the presence of the parametric effect of TBI for unexpected but not expected trials points to a difference in looking behavior. Future work designed to specifically address this difference would be needed to confirm whether there is or is not such an interaction. If there is indeed a difference between behavior on expected and unexpected trials, it appears that individuals with TBI, rather than take a maximally flexible

wait-and-see approach as might be expected when input quality is poor (e.g., in individuals with hearing loss), generated a prediction early and did not consider other options until later in the trial. This pattern of reduced consideration of unexpected referents may have been more apparent in the Disfluent-New condition than the Fluent-Given condition because in the Disfluent condition the occurrence of the disfluency before the critical word onset provides extra time to establish an expectation about whether given or new information is likely to occur. In contrast, on fluent trials, a prediction about which cohort competitor might be mentioned cannot be generated until the onset of the critical word confirms that the referent will be one of the competitors (especially because filler trials do not always mention cohort competitors as targets, so participants should not learn to anticipate them).

The potential consequences of this difference in online language processing approach for broader discourse processing could relate to either cumulative difficulty at the sentence level over the course of conversations or difficulty with discourse-level prediction. At the sentence level, if individuals with TBI initially make stronger predictions and are not considering other possible referents, when unexpected language input occurs, they may have more difficulty overcoming the initial interpretation to reach the correct one. If this occurs frequently over the course of conversation, this could account for some of the subjective difficulty experienced by this population. Strong initial predictions about upcoming given versus new information could have a similar detrimental effect in regards to prediction at the discourse structure level, impeding shifting to a new topic if a continuation of the current topic was strongly predicted.

This language processing approach could be related to processes that lead to perseverative errors noted across other types of cognitive tasks in individuals with TBI (e.g., reduced switching in verbal fluency, perseverative errors in Wisconsin Card Sort) (Kavé et al., 2011b; Ord et al., 2010). However, in the current study, it appears that participants were able to eventually look to the correct referent. The current result is somewhat difficult to interpret because predictions were expected to be altered due to the disfluency cue. Further studies would need to specifically examine the strength of initial predictions,

maintenance of multiple competing representations, and recovery from those predictions, for example by examining garden path processing in TBI.

Effect of Covariates on Using Speaker Cues to Inform Predictions

Similar to the previous analyses, the effect of age was primarily to reduce the proportion of predictive looks to the competitor, but it did not eliminate the disfluency effect.

In regards to the effect of PTSS, the same earlier onset of predictive looks to the competitor on Fluent-New trials, when the competitor should be expected, was observed in this model as described in the model examining this condition relative to the given/new effect. On Disfluent-New trials, participants in the TBI group with low levels of PTSS showed performance similar to the pattern observed in the raw data and primary analyses–namely a low proportion of looks to the competitor early in the trial. It is this effect that appears to be the primary source of differences between the Healthy Control and TBI groups. In contrast, participants with higher levels of PTSS showed a higher proportion of looks to the competitor throughout the duration of the trial, and in particular, while individuals with lower levels of PTSS were not considering the competitor. This result suggests that it is TBI per se and not high PTSS severity that is driving the observed differences between the TBI and Healthy Control groups. In fact, it appears that having high PTSS severity makes it more likely that individuals with TBI will perform more like Healthy Controls. This result highlights the need to account for post-traumatic stress in order to differentiate effects of TBI per se from effects that co-vary with TBI history and suggests that PTSS severity may not always make the effects of TBI "worse," or at least more different from controls.

In regards to neurobehavioral symptoms, there was no effect of symptom severity on the disfluency effect in the TBI group.

Cognitive Effort Required for Model Updating

The results of the pupillometry analysis revealed no effect of whether the prediction about the likely target was confirmed or not for Healthy Controls and no difference in this effect between groups. This finding runs counter to the hypothesis that more cognitive effort is required to complete the task

when initial predictions are contradicted. The failure to detect such an effect could be due to several reasons. First, there may be no difference in the cognitive effort required between the two types of task conditions, either because updating predictions generally does not increase cognitive effort (i.e., alter the cognitive resources required or resource allocation strategy) or because this task in particular does not demand a significant amount of cognitive effort. Another possibility is that the analysis window is not long enough to capture the full range of the slow task evoked pupillary response, as the task design is optimized for fixation analyses, not for pupillometry (Winn et al., 2018). While later time periods could be examined, any pupil response related to model updating would be confounded with pupil response related to initiating the motor response of clicking and dragging the target item.

However, there was a significant effect of group, such that participants in the TBI group had an initially larger pupil size and later a smaller pupil size than Healthy Controls on trials where the prediction would need to change. While there were no periods of significant difference between the groups, the direction of the effect could be informative if it could be confirmed with further studies. The pattern shown by the model estimates suggests that the change in effect over time is due to relatively stable pupil size in the TBI group, with increasing pupil size over time for the Healthy Control group. This effect could potentially indicate that individuals with TBI were exerting less cognitive effort on unexpected trials, and this effect may have continued to emerge if more of the pupil response could be examined. However, it could be that the expected task evoked pupil response does not emerge for the TBI group because there is too much variability and too few participants, while this was not the case for the larger and more uniform Healthy Control Group. Additionally, some of the later increase in pupil size may be related to preparation for and the onset of clicking responses, which emerges as an artifact of choosing the average click time across all participants as the end of the analysis period. For example, if Healthy Controls more frequently clicked sooner than the average across all participants, some additional pupil dilation may have occurred that had not yet occurred for the TBI group. Indeed, the average click onset for the Healthy Controls was 56 milliseconds faster than the onset for the TBI group (Healthy

Controls: 2256, TBI: 2312 milliseconds post target).

General Discussion

This study applied two visual world eye tracking tasks with well-understood results in healthy adult participants to examine listening performance in individuals with a history of TBI. The prediction and model updating abilities assessed with these tasks served as a simplified model of key abilities that are hypothesized to allow individuals to track when the content of conversations shift from talking about ongoing/given information to talk about new information. The Experiment 1 Simple Prediction task was used to establish whether and over what time course individuals with TBI used within-sentence information to make predictions about upcoming linguistic information, as making predictions about upcoming information is hypothesized to be required for efficiently updating discourse representations. The Experiment 2 Discourse Task was used to assess whether individuals with TBI used information from the prior discourse to make predictions about whether upcoming information will be given or new and to assess whether they can use reliable cues from the speaker to adjust these expectations. This task also allowed for examination of the cognitive effort required when predictions were contradicted by incoming linguistic information, as may happen when individuals encounter shifts in conversation topics.

General Summary

Comparing individuals with TBI to a group of neurologically healthy adults on two validated psycholinguistic tasks demonstrated that the TBI group's pattern of results did not differ in any of the primary comparisons. Individuals with TBI appear to make linguistic predictions about upcoming referents during online language processing, both using within-sentence information in the Simple Prediction Task, and using prior discourse information in the Discourse Task. Individuals with TBI also were able to use disfluency cues from a speaker in order to adjust their online predictions similarly to Healthy Controls. Neither group demonstrated differences in cognitive effort in response to language contexts where their initial predictions were disconfirmed by the actual input. Taken together, this suggests that individuals with TBI (at least, primarily with mild TBI) do not have deficits in basic

discourse memory, online prediction, or model updating abilities relative to healthy adults (cf. Table 10). Thus, differences in subjective experiences or objective performance during conversational discourse are more likely attributable to other cognitive factors (e.g., attention), higher-level discourse functions not addressed in the current experiment, difficulty integrating multiple sources of information that were simplified for the current experiment, or to other factors associated with TBI but not TBI per se (e.g., fatigue, physical symptoms, mental health).

However, the TBI group and Healthy Control group did differ in ways that were not initially hypothesized. Performance on the Discourse Task revealed that, once a prediction about the likely referent could be made, individuals with TBI tended to consider the unexpected referents less than the healthy adults. While further studies specifically designed to examine this difference are required, these results provide some preliminary evidence that individuals with TBI do have differences in predictive processes. However, these differences are related to commitment to a prediction once it is made rather than to generating a prediction at all. This pattern may be consistent with other perseveration-related behaviors that are characteristic of individuals with TBI. If this pattern is validated, it could suggest further investigation into difficulty over a conversation, as strongly-held predictions may lead to repeated prediction error, or difficulty with switching between topics once an ongoing topic has been established (though possibly only subjectively, cf. shift cost analyses in Chapter 2).

The results of covariate analyses demonstrate that some key factors need to be accounted for when examining effects of TBI on language comprehension, particularly in regards to examining military populations. Reduced linguistic prediction at higher ages could obscure group-level effects and should be addressed either in study designs or statistically. While the age effects have been addressed in other contexts in the previous literature, this study is one of the first to address PTSS as it relates to online language processing. While additional work is needed to differentiate effects of PTSS per se from factors that covary with PTSS (e.g., special military training or exposure to environments more likely to cause traumatic stress), differences in the timing of predictive language processing at higher levels of PTSS may

obscure timing effects when examining group effects. More generally, this result also suggests a need for further investigation into the effects of PTSS or the interaction of PTSS and brain injury on online language comprehension, which is currently under-represented in the PTSD literature relative to investigation of language production (Yu et al., 2025).

Limitations

One limitation of the current study is that the participants in the Healthy Control Group and TBI Group differed in regards to known factors such as age, PTSS severity, and physical symptom severity, and likely differed in unmeasured factors such as life experiences or training that could affect one's approach to the current tasks. This limitation can be addressed by comparing individuals in the TBI group to the service member and veteran control sample for which data collection is ongoing. While a control group better matched on unique life experiences that may occur for service members and veterans will account for some of these factors, the current study does provide an important comparison to the individuals who are more similar to the participants who have been studied in prior psycholinguistic work to derive the "classic" effects. Thus, this work provides an initial step towards comparing the performance of individuals with TBI history with what is generally expected from healthy adults, and provides groundwork for interpreting any differences or lack of differences that may emerge when comparing the TBI group to an age and occupation matched control group.

Other considerations that are related to the TBI group used in this study include injury characteristics such as severity and time since injury. The majority of participants (N=20; 83%) sustained mild TBIs and all participants were in the chronic period of recovery (>6 months post TBI). While all of the participants were seeking help for ongoing subjective cognitive difficulties at the Brain Fitness Center clinic, it is possible that some of their symptoms were attributable to other factors besides the TBI (e.g., PTSD, anxiety, depression) and therefore they did not have persistent neurological damage that would significantly differentiate them from healthy adults. Thus, some effects of neurological damage in more severe TBI cases could have been washed out by participants with less severe TBI history. While this

distribution of TBI severity is broadly representative of the prevalence of different severities in the military (Agimi et al., 2019; Defense Health Agency, 2025) and in civilians (Corrigan et al., 2018; Levin et al., 2014), the heterogeneous makeup of this group of participants could mean that some effects that do actually occur in more severe TBI were not detected in this study.

The nature and consequences of the injuries, even though most were mild, must also be taken into account. Of the 20 participants with mild TBI, 12 had a history of multiple mild TBIs, including repeated exposure to blast. Having multiple TBIs, even if they are mild, may lead to different symptom profiles than single injuries (Jannace et al., 2024; Sorg et al., 2021). Further, although participants either had normal hearing or used hearing aids, environmental exposures to blast (and also likely noise, given the occupational and combat environments where blast exposure occurs) could lead to subtle changes in hearing acuity or auditory processing that introduce additional variability into the TBI group (Brungart et al., 2019). Thus, there is considerable heterogeneity in the current sample, despite the overall prevalence of mild TBI. This heterogeneity may also contribute to difficulty determining group effects.

A final consideration regarding TBI characteristics is that many participants in the TBI group were several years post injury. Therefore, participants could have been in various stages of recovery and some participants may have differed in the development of strategies for cognitive-linguistic tasks. While some strategies may have been developed unconsciously over years of experience, other strategies may have been explicitly learned if participants had completed any rehabilitative programs following their injuries (e.g., speech therapy, occupational therapy, mental health counseling). The results of this study therefore represent cognitive-linguistic abilities across the range of the chronic phase of TBI recovery, and may not be applicable to patients who are no longer in the acute phase of their injuries but earlier in the process of recovery (e.g., 3-6 months post injury).

Limitations inherent to the tasks used for this study are that the linguistic content was shorter and less complex than unstructured conversation. While the benefit of using these tasks is the process-specificity and the ability to use tools with fine-grained temporal resolution, it is currently unknown whether there

are differences in how the processes that we hypothesize to be the same between the current tasks and tracking topics are actually deployed during unstructured conversations. The results of the current study suggest that reduced consideration of possible alternatives when language input is ambiguous may lead to difficulty if they build up over time, even if the processes per se are intact. Future work would need to explicitly test this by increasing the duration and/or complexity of the discourse content. Although participants were told that the short discourse samples were recorded by other participants to increase the social components to reflect some of the demands of conversation, future work also must account for dyad interactions, including turn taking considerations, social inferences, and non-linguistic communication (e.g., integrating linguistic content with gesture, facial expressions).

Conclusions

Taken together, these results have provided evidence that suggests memory, online prediction, and model updating are not likely to be the underlying mechanisms contributing to difficulty tracking discourse content in individuals with TBI (at least in short exchanges). While key covariates like age and PTSS do interact with language processing, they cannot account for the current results. However, differences in listeners' commitments to possible interpretations when language is ambiguous may be a promising future direction to provide a mechanistic explanation of discourse difficulty in TBI. Examining language comprehension in real-time is necessary to examine these processes because linguistic ambiguity is time-sensitive, and thus these novel methods remain a promising future direction for studying cognitive-communication abilities in individuals with TBI.

Chapter 4: Exploring the interaction of changing topics and online language processing

Rationale

The overarching goal of this dissertation is to examine dynamic conversational discourse processes in real-time and to examine how those processes may break down following brain injury. Chapters 2 and 3 have thus far described approaches on opposite ends of the spectrum, with Chapter 2 maximizing naturalness in unstructured conversation and Chapter 3 using simpler tasks to model individual sub-processes hypothesized to contribute to discourse structure in real-time. The study reported in this chapter aims to take a middle-ground approach by capturing more of the complexity of naturalistic conversation, while maintaining the benefits of using eye tracking measurements to obtain information about online linguistic processes.

In light of the need to examine online language processing in individuals with TBI within more complex conversational contexts, the goal of the current chapter is to develop a new testing paradigm and to first understand the expected performance in healthy adults. The same paradigm could then be used to compare participants with TBI to healthy adults in the future. Even though results of Chapter 3 suggest that there are no differences between individuals with TBI and healthy adults in regards to some of the processes that may support identifying topic transitions—namely memory, prediction generation, and model updating—there may be differences regarding how topic structure representations may interact dynamically with other, lower-level aspects of language comprehension.

In the previous discourse literature and across Chapters 2 and 3, an important underlying assumption is that individuals create, maintain, and update a representation of discourse structure as linguistic information continues to unfold. As previously discussed, while this assumption may be true for monologic discourse, it may not hold for dialogues, where conversation partners can abruptly alter the course of the conversation. In conversation, it may not be beneficial to try to maintain topic structure information and use it to guide any expectations about upcoming linguistic information in real-time

because the high variability within individual conversation partners and between pairs of interlocutors may preclude an individual's ability to reliably make accurate top-down predictions. Instead, it might be more beneficial to discuss information that is only locally relevant to a conversation partner's previous turn, rather than maintain a broader discourse representation to provide information about what is globally relevant to the topic. Or, it may be beneficial to wait for several utterances or several turns to pass before updating a topic structure representation, such that topic information is not always available to influence the interpretation of incoming linguistic information in real-time as it unfolds.

The results presented in Chapter 2, which demonstrate topic shift costs, are consistent with the hypothesis that individuals are sensitive to changes in global topic structure soon after they occur. However, this study used post-hoc topic structure identification by raters, and there was an extended time course between entire turns establishing a topic shift and subsequent production. Thus, open questions remain (1) about what participants themselves represent about topics as the discourse unfolds (versus what raters think they are representing), (2) about the time-course of when topics affect online language processing, and (3) about how topic representations affect language comprehension rather than language production. Answering these open questions can provide empirical evidence for the critical assumption that individuals represent and use topic information to guide language comprehension and production in real-time. It can also address questions that have yet to be explored in the psycholinguistic literature regarding the complexity of language comprehension in dynamic, naturalistic settings.

The sections that follow discuss how the field of psycholinguistics has addressed questions about the interactions between discourse context and online language comprehension. I will review factors that have been missing from previous approaches—namely, examination of when discourse context can change within the broader communicative task—and how these relate to open questions about language processing in naturalistic conversation. Then, I will discuss evidence that addresses some components of discourse context variability and describe a novel experimental paradigm that adapts these previous studies to examine the interaction of conversation topic structure and online language processing.

Approaches to Varying Relevance within the Broader Context

Addressing whether individuals represent topic information and deploy this information in real-time requires experimental paradigms that can capture the broader structure of discourse and the dynamics of topic changes in conversation. Much of the previous literature in psycholinguistics has addressed discourse context as a single, static piece of information. However, some work has addressed the changes in contextual information in more complex communication environments. This most frequently takes the form of examining different speakers as their own contextual constraints on moment-to-moment language comprehension or production decisions. For example, listeners can learn that different speakers consistently use different alternations of a syntactic structure or different kinds of errors and apply this information to make speaker-specific predictions about upcoming linguistic information (Kroczek & Gunter, 2017; Ryskin et al., 2018). Speakers can also adjust their production choices based on information about what communication partners experienced before or know about their physical context (Brown-Schmidt et al., 2015). Studies addressing variable constraints according to speaker identities demonstrate that different information sets can be applied within the same general communicative task. While topic structure is hypothesized to impose the same kind of demand, this previous work has not directly addressed questions about conversational structure during dialogue.

Brown-Schmidt and Tanenhaus (2008) addressed dynamic interaction during a dialogue more directly. In this study, participants worked in pairs to arrange game pieces in different physical locations on game boards that were concealed from each other's view. One participant had their eye movements tracked while they and a partner completed the task. Some game pieces had stickers with pictures on them and some did not, such that where one partner had a sticker, the same location on the other partner's board was missing a sticker. The goal of the activity was to take turns giving each other instructions about where to place the stickers so that their boards ultimately matched. Surrounding game pieces could be used as reference points to describe where to put the sticker. The boards were also divided into physically distinct sub-areas. Blocks with pictures depicted phonological cohort competitors (e.g., candle/camel;

cloud/clown). In some cases, the members of the cohort were placed within the same sub-area of the board and in other cases the cohort members were in separate sub-areas. Trials when the partner was speaking during the task were compared to trials outside of the task context, in which the experimenter asked the participant to look at the cohort items under the guise of calibrating the eye tracker. In traditional visual world studies, cohort competitors are typically both considered (i.e., have an equal proportion of looks) during the ambiguous onset period (e.g., /kae/, /kloʊ/). By comparing trials when the partner was giving instructions, where only competitors in the sub-area they were working in should be relevant, to trials when the experimenter was giving instruction, where all competitors should be relevant, the authors could examine whether the task-relevant physical space constrained the domain of reference. The authors found that when conversation partners referenced one cohort competitor block, there were fewer looks to the competitor than when the experimenter referenced cohort competitors (Brown-Schmidt & Tanenhaus, 2008). These results demonstrate that, during a dynamic, unstructured conversational task, participants constrained the possible domain of reference according to task-relevant delineations in physical space. This suggests that task-relevant divisions of the informational space are represented moment-to-moment and that these task-relevant representations influence online language processing decisions. While this study addressed the interaction of the physical space, task demands, and the role of communication partners, open questions remain about other ways in which contextual information can be dynamically altered over the course of an interaction and about different modes or purposes of interacting (i.e., besides completing physical tasks).

The Current Study

Given the evidence of discourse-relevant constraints within the physical "event space", this study examines whether the same kind of processes apply over "conversational space". While many interactions, like the ones in Brown-Schmidt and Tanenhaus' study, are constrained in terms of the physical space of the task, there are also many conversations where the task goals are related to informational space (even though they all take place in some kind of physical context). For example,

catching up with a friend over the telephone, asking about someone's opinions on a first date, discussing the themes of a novel in a book club, or planning for future events in a work meeting appear to have an internal organization that is not necessarily related to physical surroundings. These kinds of conversations also ostensibly have structural delineations in the form of topics. Therefore, the current study addresses questions about whether and how topic representations delineate what is or is not relevant by examining whether topics alter the domain of reference. This study used naturally generated conversations, with slight modification to achieve the parameters needed to make inferences about eye movement patterns within the visual world paradigm. Ultimately, this kind of approach can encapsulate some of the complexity of naturalistic conversation (as in Chapter 2), while also assessing an easily measurable, objective eye movement response that is sensitive to moment-to-moment processing decisions (as in Chapter 3).

However, several factors must be changed to translate the concepts of the previous work to examine conversation structure per se, rather than task structure. While Brown-Schmidt and Tanenhaus' work involved unstructured conversation between participants, the nature of the task was highly constraining. Even though there may have been some variability in speaking styles or task strategies, the content of the conversations and the physical environment was similar across participant groups. In contrast, for the types of conversations addressed in the study reported here, the content of the conversation and the topic structures will vary extensively depending on the participants' backgrounds, world-knowledge, and relationship with each other, among many other factors. Thus, as a starting point for developing experimental paradigms to examine topic structures, this study uses a more controlled approach by having participants listen to recorded conversations. This also allows for the experimenters to determine where topic transitions happen a priori, rather than having raters identify topic transitions post-hoc from unstructured conversations. However, because task-relevance and active participation in the communication event can modulate language comprehension and production processes (Brown-

Schmidt et al., 2015; Brown-Schmidt & Tanenhaus, 2008), the task also requires actively listening to the conversation in order for it to be completed.

In this task, participants were told that they are part of a team preparing to put on a stage show. They were informed that they are the prop manager and must listen to discussion between the writers who are planning the scenes and select the objects that are needed from a visual display. Thus, the participant's task requires information from the conversation, but the conversation is not necessarily about carrying out the task.

A conversation about stage scenes also provides a framework where objects are not necessarily semantically related or members of a particular semantic category, but are nonetheless relevant in the context of a particular topic (i.e., what happens in a scene). To ensure that the manipulation involves topic representations, which should depend on the context from the discourse rather than rely on general relationships outside of conversation information, study materials avoided simple semantic categories (e.g., animals, fruits) while ensuring that related items are clearly expected, given the topic. For example, without any prior information, a strawberry and an apple are likely to be considered members of the same *category;* however, if two distinct topics of conversation include making a strawberry short cake and making an apple pie, the same two objects should be considered relevant to different *topics*. While previous work has examined interactions of discourse context and world knowledge (e.g., overriding expectations about category associations due to world knowledge (Hald et al., 2007)), a key way in which this paradigm differs is that a conversation about multiple scenes means that different topics occur within the broader context and task. The multiple scenes are discussed in the broader context of the "show" and completing the object selection task requires keeping track of which topic is currently active at any point in the conversation.

One additional consideration is the type of temporary ambiguity that applies within versus across topics. While Brown-Schmidt and Tanenhaus (2008) used phonological cohort competitors, it is not possible to counterbalance the use of the same object within and across topics in the same way that the

same picture of a cloud could be used as a reference point whether it was in the same or different physical space as a picture of a clown. In the current study, the identity of the objects is important to signal topic relevance. To avoid this issue, a temporary referential ambiguity was created by including objects that can be described with the same adjective. Previous work has demonstrated that objects in a visual world display that can be described by the same adjective are all temporarily considered (fixated) until information about the referent is heard (Eberhard et al., 1995; Langlois et al., 2024; Lord et al., 2024). Brown-Schmidt and Tanenhaus (2008) also demonstrated consideration of multiple referents during ambiguous referring expressions containing adjective phrases within unscripted language games, though under slightly different analysis conditions. Thus, the adjective (and thus the ambiguity) can occur within or across topics, without changing the identity of the objects relevant to each topic.

Table 27 provides examples of this task design. For example, some objects may include black microphone, white curtain, black refrigerator, and white pan. Without context, when participants hear "black…" they would consider the microphone and the refrigerator (Eberhard et al., 1995). In this task, however, if the show contains scenes that are in a radio station and a commercial kitchen, when participants hear "black…" the two possible referents are now across topics. If participants are keeping track of topic in real-time, they should restrict their interpretation to exclude topic-irrelevant objects. So, when hearing the "writers" talking about the radio station scene, participants would look at the black microphone but not at the black refrigerator. On a counterbalanced list, the objects may include black microphone, black curtain, white refrigerator, white pan. Now, when participants hear "black"… the two possible referents, microphone and curtain, are within the same topic. If participants are tracking the current topic and the competitor is now relevant to that topic, it should be considered until disambiguating information about the actual target occurs. All the objects could be either color, so lists can be fully counterbalanced regarding which object is the target versus competitor and whether the competition is within or across topics. Filler items included the same objects with different features (e.g., a different color from all of the other critical objects) to increase the felicitousness of specifying objects with

modifiers. Because object pairs may be within or across topics, all of the objects that could be targets or competitors throughout the conversation needed to be visible. While this required a visual display with many more objects than the typical four-item visual world paradigm (Allopenna et al., 1998; Arnold et al., 2004; Hsu & Novick, 2016; Kukona et al., 2014; Langlois et al., 2024; Nozari et al., 2016), the results of Brown-Schmidt and Tanenhaus (2008) suggest it is possible to measure competition effects in a larger visual array. Similar to Brown-Schmidt and Tanenhaus (2008), participants were also asked to look at the objects in trials outside of the conversation context in order to ensure that competitors with the same modifier are both considered.

Hypotheses and Predictions

Outside of conversational contexts, incoming linguistic information should activate information about all possible items in the task environment that could be consistent with the input at the time. As additional information unfolds, the set of possible referents is narrowed to what is consistent with information about a particular referent. Therefore, I expect participants will consider multiple competitor referents that are consistent with a modifier until information about the target object is available in the linguistic signal.

Within the conversational context, I hypothesize that conversation topics—the discourse structure representations that define a scope of information that is relevant versus irrelevant—are maintained moment-to-moment and reflect the currently available information from the discourse. If topic information is available, it should be applied to decisions about parsing incoming language information in real-time. In this case, topic information should constrain the domain of reference to only topic-relevant items, eliminating consideration of any items that could be competitors but are not topic-relevant. Thus, participants should have a similar proportion of looks to a possible competitor object when it is within topic as the same competitor when there is no conversational context. Also, participants should have a lower proportion of looks to a competitor when it is across topics than the proportion of looks to the same competitor when there is no conversational context. If topic information is not used to constrain the

domain of reference, the proportion of looks to within- and across-topic competitors should be the same on trials within the conversation context as on trials outside of the conversation context.

Given the large body of evidence that known sources of linguistic and non-linguistic knowledge influence online language processing, I argue that the most likely interpretation of a null effect is that topic information is not available at all at the real-time time scale of sentence parsing. However, the difference between availability of topic information versus having topic information available, but this information not constraining referential ambiguity interpretation, is not distinguishable if a null effect occurs. Indeed, some previous work has shown that consideration of temporarily consistent competitors is not totally constrained by certain features, such as verb-semantic restrictions (Kukona et al., 2014; Langlois et al., 2024; Nozari et al., 2016). Therefore, additional follow-up studies would be required to adjudicate between these possibilities.

Method

Participants

Participants were N=41 (ages 18-24; Mean = 19.7 years, SD = 1.3 years; 25 female) healthy adults from the UMD community. Participants reported that they were native English speakers (heard and spoke English before age 5, regardless of other language exposure), that they had no history of speech, hearing, or language disorders, they had no history of neurological disorders (e.g., stroke, TBI, concussion, epilepsy), and they had normal or corrected-to-normal vision. This study was approved by the UMD Institutional Review Board. All participants provided informed consent prior to participation.

Stimuli – Conversation Contexts

The task instructions informed participants that they were part of a production team for a sketch comedy show that does multiple sketches each week. Participants were instructed that they were the "prop master" for the show, responsible for selecting the props that they will need to use by selecting them from a visual catalogue displayed on the screen. They were told that they would hear a discussion about a scene

and be tasked with selecting the appropriate props from the visual array which represents the contents of the "prop closet".

To generate the conversations, two research assistants were instructed to brainstorm a possible set of comedy sketches that included the critical items (see below), as if they were writers on a TV show. Critical items were provided in writing and speakers did not view the visual arrays while recording the stimuli. They were provided with three sets of two general topics for sketches (e.g., chef in a restaurant kitchen, hosts at a radio station) and instructed to switch between talking about the two sketches at natural points, but to avoid talking only about sketch until the idea was compete and then talking about the other in order to facilitate production of multiple naturalistic topic shifting markers. No further instructions were provided. Therefore, the content, linguistic forms (e.g., disfluencies, syntactic structures), and location and form of topic shifts was generated by the two individuals. This process was intended to generate important elements of conversation such as interruptions, disfluencies, and speaking style, which are difficult to replicate by writing dialogue, while allowing for control of critical topic-related elements.

The naturalistic conversation recordings were transcribed and minimally edited by the author to meet the desired form for the experiment stimuli. Each conversation was edited to transition between talking about scenes twice (e.g., scene A, B, A, B), with transitions between topics marked with an explicit discontinuity device (e.g., "For the kitchen sketch…) (Leaman & Edmonds, 2020). Edits included moving the position of some utterances so that there were only three topic transition points and each topic was discussed twice in alternating order, adding the adjectives for each of the items to create counterbalanced lists, adding additional sentences to refer to filler items, adding an explicit topic shifting device used elsewhere in the recordings to shift locations that were not explicitly marked, and making slight content adjustments to increase the generality of a statement such that multiple target items could be felicitously mentioned (e.g., "the instrument could get knocked over" changed to "something could get knocked over"). An example script is included in Appendix 3.

The scripts generated by this editing process were then provided to the same research assistants who generated the initial recording. They were instructed to read the scripts and try to use natural timing and intonation as much as possible despite reading, which was facilitated by each person reading the same content they originally said and by recording the reading of the scripts together in the same recording booth. Again, speakers did not have access to the visual scenes when recording the stimuli. The speakers recorded four versions of the same conversations from start to finish, where each version included all the same conversation content and filler items and differed only in the adjective and object pair used for each critical item. The speakers also re-recorded each of the critical sentences independent of the conversation context because some trials of the initial recording contained contrastive stress on the adjective, which might have influenced listeners' interpretations of intended referents. On the re-recorded critical trials, speakers were instructed to produce every trial with stress on the object word.

The four complete recordings and the re-recorded critical trials were manually divided into separate audio files for each utterance. The onset of each audio file was marked as the beginning of speech onset for that utterance, and the offset was marked at the end of the speech plus any unfilled pauses, thus maintaining the timing produced by the speakers talking together during the recording. For the conversation context utterances (non-critical, non-filler utterances), the majority of the utterances were taken from the third of four full recordings. This way, the timing would reflect a semi-naturalistic reading and the speakers had become familiar with the content and recording process, such that their reading was more fluent and closer to naturalistic speech. If there were reading mistakes or excessive background noises, the same utterance from one of the other full recordings was used. All critical trial utterances were taken from the re-recording session to ensure consistent intonation patterns. Audio files for critical items were not generated by splicing in order to approximate naturalistic conversation. While there may be coarticulatory or timing cues to the intended referent, participants hear the same audio across conditions so any differences in looking behavior should be attributable to only the surrounding context. To create the counterbalanced lists, the same context utterances were used on each list, while the

critical trial recording that corresponds to the adjective-object pair for that list was used. Within each vignette and each list, audio files were normalized to -25.0 RMS using Audacity's Loudness Normalization tool (Version 3.1.3). The same critical trial recording file used within the conversation context portion of the experiment was used as the No Context trial.

Stimuli – Critical Phrases, Context Trials

Each critical phrase was composed of an adjective-noun pair, where the adjective can also describe one other object present in the visual array. This shared feature described by the adjective produced a temporary ambiguity between the two possible referents that share that feature. To allow for counterbalancing, critical phrases were created in groups of four—two pairs of objects that can all be felicitously described by the same two adjectives and could all felicitously be interchanged within the same sentence given the surrounding conversation context. In each set of four, two possible combinations were objects relevant to the same topic and two possible combinations were objects relevant to opposite topics. In each version of the conversation vignette, one of the two objects in the pair was mentioned. The unmentioned object in the pair served as the competitor; whether it was the competitor to the Cross-topic or Within-topic object was counterbalanced across lists. Which object was mentioned and which adjective was associated with the objects were also counterbalanced across lists. The same beginning to the sentence (i.e., sentence frame) was used for each critical phrase across lists. The sentence frame did not contain any topic-specific information, such as people or actions taking place in the scene, and did not contain anaphoric references to the previous information (e.g., "We can get the…"; "We should use the…"; see Table 27). Sentence frames, if taken out of the preceding topic context, would not provide any constraining information. The same sentence frame could therefore be used in the No Context condition. A total of eight counterbalanced lists were created. An example of the counterbalancing scheme is shown in Table 27.

	List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8
Visual Array (subset)								
Critical Phrase 1, Context, Radio "We can get the …"	Target: Black microphone Competitor: Refrigerator (cross)	Target: Black microphone Competitor: Curtain (within)	Target: White microphone Competitor: Refrigerator (cross)	Target: White microphone Competitor: Curtain (within)	Target: Black curtain Competitor: Pan (cross)	Target: Black curtain Competitor: Microphone (within)	Target: White curtain Competitor: Pan (cross)	Target: White curtain Competitor: Microphone (within)
Critical Phrase 2, Context, Kitchen "We can use the …"	Target: White pan Competitor: Curtain (cross)	Target: White pan Competitor: Refrigerator (within)	Target: Black pan Competitor: Curtain (cross)	Target: Black pan Competitor: Refrigerator (within)	Target: White refrigerator Competitor: Microphone (cross)	Target: White refrigerator Competitor: Pan (within)	Target: Black refrigerator Competitor: Microphone (cross)	Target: Black refrigerator Competitor: Pan (within)
…								
Critical Phrase 1, No Context A "We can get the…"	Target: White curtain Competitor: Pan	Target: Black curtain Competitor: Microphone	Target: Black curtain Competitor: Pan	Target: White curtain Competitor: Microphone	Target: White microphone Competitor: Refrigerator	Target: Black microphone Competitor: Curtain	Target: Black microphone Competitor: Refrigerator	Target: White microphone Competitor: Curtain
Critical Phrase 2, No Context B "We can use the…"	Target: Black refrigerator Competitor: Microphone	Target: White refrigerator Competitor: Pan	Target: White refrigerator Competitor: Microphone	Target: Black refrigerator Competitor: Pan	Target: Black pan Competitor: Curtain	Target: White pan Competitor: Refrigerator	Target: White pan Competitor: Curtain	Target: Black pan Competitor: Refrigerator

Table 27. Example trials for one critical item, demonstrating the counterbalancing protocol across lists. Visual array example includes only the subset of items relevant to the trial. Each participant would see both conversation critical trials and either No Context A or B.

Several lexical and semantic factors were evaluated for the names of the objects to be used within the critical trials in order to avoid confounds in looking preferences due to features besides relatedness to the ongoing topic. To account for covarying lexical factors across competitors, objects were identified in pairs, where one object was in each of the two topics. LexOPS (Taylor et al., 2020), a tool for generating stimuli with specific linguistic parameters, was used select pairs matched within a tolerance of 0.35 for

frequency in zipf (base 10 logarithm of frequency per billion words) from the US Subtlex corpus, and for familiarity, concreteness, and imageability from the Glasgow norms (Brysbaert & New, 2009; Scott et al., 2019; Taylor et al., 2020). Then, the pairwise semantic similarity score (cosine similarity) was calculated using word2vec for all pairs of objects within a topic and for each object in one topic with the objects in the other topic (Mikolov et al., 2013). This evaluated whether there was a difference between the within-topic semantic similarity and across-topic semantic similarity. Therefore, any association between the objects should be based on the conversational context and topic rather than "decontextualized" semantic similarity. Paired t-tests confirmed that there were no differences between competitor items in each of the lexical factors described above as well as no differences in the degree of difference between the possible within-topic and across-topic competitor pairs (all $p > 0.05$). Paired t-tests also confirmed that there were no significant differences in the pairwise cosine similarities for items within a topic versus pairwise similarities for items across topics (all $p > 0.05$).

Nineteen UMD undergraduate students participated in a norming study to confirm that objects selected to belong to a particular topic were perceived as such by naïve observers. Norming study participants were provided with a description of each of the topics used in the experiment (e.g., "Two people are talking about hosts of a radio show and the studio they work in") and shown a picture and written label for half of the objects to be used in the conversation addressing that topic. Which objects were shown with which topic was counterbalanced across two lists. Participants were instructed to decide whether each object could be relevant to the given topic, such that it would make sense to hear people who are talking about that topic mention that object. For each item, participants could respond either "Yes the object is probably relevant" and "No the object is probably not relevant." Objects were considered to be perceived as topic-relevant if the proportion of participants responding "yes" when the object was expected to be relevant based on the stimuli construction process was greater than the proportion of participants responding "no". This criterion was met for all but one object. This object was replaced and

an additional eighteen participants responded to the norming study. The updated object met the inclusion criteria and the results were replicated for all other objects.

There were five critical phrases per topic, such that there were 30 critical trials per participant (five phrases/topic x two topics/conversation x three conversations). Half of the critical trials included Cross-topic competitors and half included Within-topic competitors, based on the format of the visual array. Participants also heard similar sentences mentioning five filler objects. Filler objects had the same object identity as one of the critical objects but were a color or texture that does not correspond to any of the adjectives in the critical phrases. Fillers should increase the felicitousness of using modifiers (i.e. because there are two microphones a color would be needed to differentiate them). The inclusion of fillers also meant that not every trial requiring the participant to click on an item contained a temporary ambiguity.

Stimuli - No Context Trials

Participants were told that the eye tracker needs to be calibrated before listening to the conversations, and that this involves looking at the object arrays and clicking on objects. This should dissociate the task demands of No Context from Context trials. Participants heard the same critical trials described above, but presented one at a time with no surrounding conversation context. The audio file was the same audio file used within the Context trials. The critical phrase in the No Context trial used the same visual array (so had the same competitor sets), but the target in the No Context trial was the unmentioned competitor object in from that list's conversation Context trial (Table 27). The visual array presented on the No Context trials was the same as the visual array and sentence set for the upcoming conversation Context trials. Thus, all No Context trials occurred before the topics related to the items had been introduced. Each set of No Context trials included a subset of five sentences from the corresponding Context lists. The utterances that were in the subset were counterbalanced across participants viewing a particular list. Each list had an A and B version that differed only on the No Context trials, such that all of the critical trials could occur in the No Context condition across lists, without any individual participant

hearing all of the competitor items that would occur subsequently in the Context condition. The counterbalancing also balanced any effects of prior exposure across participants, such that for any given Context item, half of participants previously heard the competitor in the No Context condition and half did not. Participants heard a total of fifteen No Context trials, corresponding to the same number of Within- and Cross-topic critical trials. Participants also heard nine filler trials (three per set of five critical trials) in the No Context condition. These fillers were also the same audio file that occurred in the Context list.

Stimuli - Visual Arrays

While listening to the No Context or Context trials, participants viewed a visual array of twenty-five objects arranged in a 5x5 grid (Figure 27). Ten objects corresponded to the target and ten to the competitor objects. Five of the objects were filler objects. Object pictures were primarily photographs of real objects, or when no photo was available, AI generated (via Adobe Stock or ChatGPT) photorealistic images. Object pictures were obtained from LinguaPix database (Krautz & Keuleers, 2022), Things Initiative database (Hebart et al., 2019), or Adobe Stock. Objects' colors or textures were altered using Photoshop so that the same picture was used with different coloring to correspond to each list. The same color parameters were used when overlaying a color on each object with that color. Objects remained in the same grid position for each list. Participants also saw a triangular mouse cursor which they used to click on items to complete the task.

Figure 27. Example visual display.

Procedure

Participants were seated in front of a computer monitor, keyboard, and mouse and positioned within the Eyelink 1000 tower-mounted eye tracker setup. Details of the setup, sampling rate, and spatial resolution are the same as those described in Chapter 3. Participants read instructions that informed them that their task was to be the "prop master" for a comedy show containing multiple scenes or sketches by listening to conversations between show writers and selecting all of the props that were mentioned as being needed for the scenes from a visual catalogue displayed on the screen. They were instructed to click on each mentioned item as soon as they heard it. Participants were instructed to attend to the conversations because they would be asked about them in a subsequent memory task (not reported here). Participants then completed two practice trials surrounded by three context sentences in order to become accustomed to the format of the visual display and the pacing of the task. The practice conversation and objects were not related to the three main task vignettes. Participants then viewed familiarization trials, in which the objects that would be shown during the task were presented one at a time with a text label (e.g., black microphone), in random order. This portion was intended to familiarize participants with the names of the items and ensure that they would be able to click on the correct target. Participants then completed calibration and validation procedures for the eye tracker. Prior to beginning the primary task, participants

read additional instructions that informed them there would be a short "calibration" task before each conversation where they would hear objects mentioned and they should click on the objects as they heard them. They then completed the eight No Context trials while viewing the visual array corresponding to the upcoming conversation vignette. Each critical phrase occurred with 1000 milliseconds of silence between trials. Participants then listened to the conversation vignette. The audio for the trials played continuously in sequence, with no additional time added between utterances. Eye movements were recorded during the No Context trials and the conversation vignette. While participants completed the No Context and conversation trials, the experimenter used a checklist to record accurate and inaccurate clicks[4]. The same procedure was repeated for the subsequent two conversations.

Analysis

To examine the time course of effects, the proportion of fixations (an eye movement to an object that remains on the object, lasting until an eye movement to a different location) per the total number of fixations on the target, on the competitor, and on a similarly spatially located non-target, non-competitor object ("Other") was calculated for each type of trial—No Context, Within (competition object is from within the same topic), and Cross (competitor object is across topics) . The Other object was the object one grid space above the competitor, or one grid space to the right of the competitor if the one above was the target or if the object was in the top row. This provided a control to differentiate looks specifically to the competitor driven by ambiguous information at the adjective from general likelihood to be looking in that area of the grid. Fifty millisecond time slices were used to determine the bins for the proportion of fixations. Proportions were calculated over participants. Trial-level information was not used because there were not enough observations per time bin in this format to use GAMMs in the format that would allow for accounting for autocorrelation (cf. Chapter 3 analysis). A rectangular interest area within the bounds of the grid location (1/25 of the total pixel area) surrounding each object determined whether a

[4] Recording clicks using the Eyelink software disrupted the timing of the audio files in the conversation portion, so accuracy was recorded manually.

fixation was on a particular object.

To assess whether topics constrained the domain of reference in real-time, generalized additive mixed models (GAMMs) were used to assess the effect of effect of condition (No Context, context Within, context Cross) and object type (competitor, other) on the empirical logit transformed proportion of looks. GAMMs are beneficial for this analysis because the time window for the expected effects is unknown, given the novel task and complex visual display. GAMMs can model the entire time series, avoiding the need for a-priori time windows or multiple analyses to identify differences in time, then assess differences in proportions. The time window for analysis included a 200 millisecond baseline prior to the adjective onset and the period from adjective onset until the (rounded) average end of the critical phrases across trials (-200 to approximately 2000 milliseconds relative to the adjective). The emplogit() function was used to perform an empirical logit transformation on the proportion of fixations to make the outcome measure unbounded in order to meet model assumptions (Fraundorf, 2022). Models were constructed using the bam() function from the mcvg package (Wood, 2003, 2011, 2017) and were assessed and visualized using the itsadug package (vanRij et al., 2022). Ordered factor coding was used to examine both parametric (overall height) and smooth (shape of the curve) effects, with a continuous factor of time. No Context was chosen as the initial condition reference level and other as the initial object type reference level. The model was re-leveled to examine both other and competitor objects as the reference level in order to evaluate whether looks to the competitor were influenced by condition but looks to other objects were not, and to examine Within and Cross trials as the reference level to examine pairwise differences between each of the trial types. Models contained terms for condition, object type, and the interaction between condition and object type. Each term was entered as a parametric term and a tensor product smooth term. Factor smooth terms for participants and participants by each of the ordered factor terms were used to model differences in responses to each factor type by participant, similar to including random slopes for condition, object type and their interaction by participants (Sóskuthy, 2021). The AR1 value of the initial model was used to determine the autocorrelation parameter rho in the final

model in order to account for autocorrelation between points in the time series. The rho value was then increased in magnitude manually until the AR1 value of the model was <0.2 and non-negative (Johns et al., 2024; Porretta et al., 2018). Model specifications can be found in Appendix 3.

Results

Accuracy results were available for N=35 participants. Accuracy data was missing from the first 6 participants due to adding the manual coding to the protocol at this point (but note that these early participants did not see a different version of the experiment). The very high accuracy (M = 95%, SD = 5%) reflects that participants were attending to the task and engaged with the task instructions.

Figure 28 shows the proportion of looks to each object type as the critical sentence unfolded by condition. Figure 29 shows just the proportion of looks to the competitor object. Visual inspection of the results demonstrated that participants did look to the target and the competitor items at greater than chance levels in response to hearing an item mentioned, and that this occurred across conditions. Thus, they behaved as expected based on previous visual world studies, despite the increased complexity of the conversation conditions and the visual displays. Notably, the peak in looks to the competitor was after the onset of the target word across conditions, meaning that participants' eye movements reflected continued consideration of multiple possible referents given the adjective, even after disambiguating information was available.

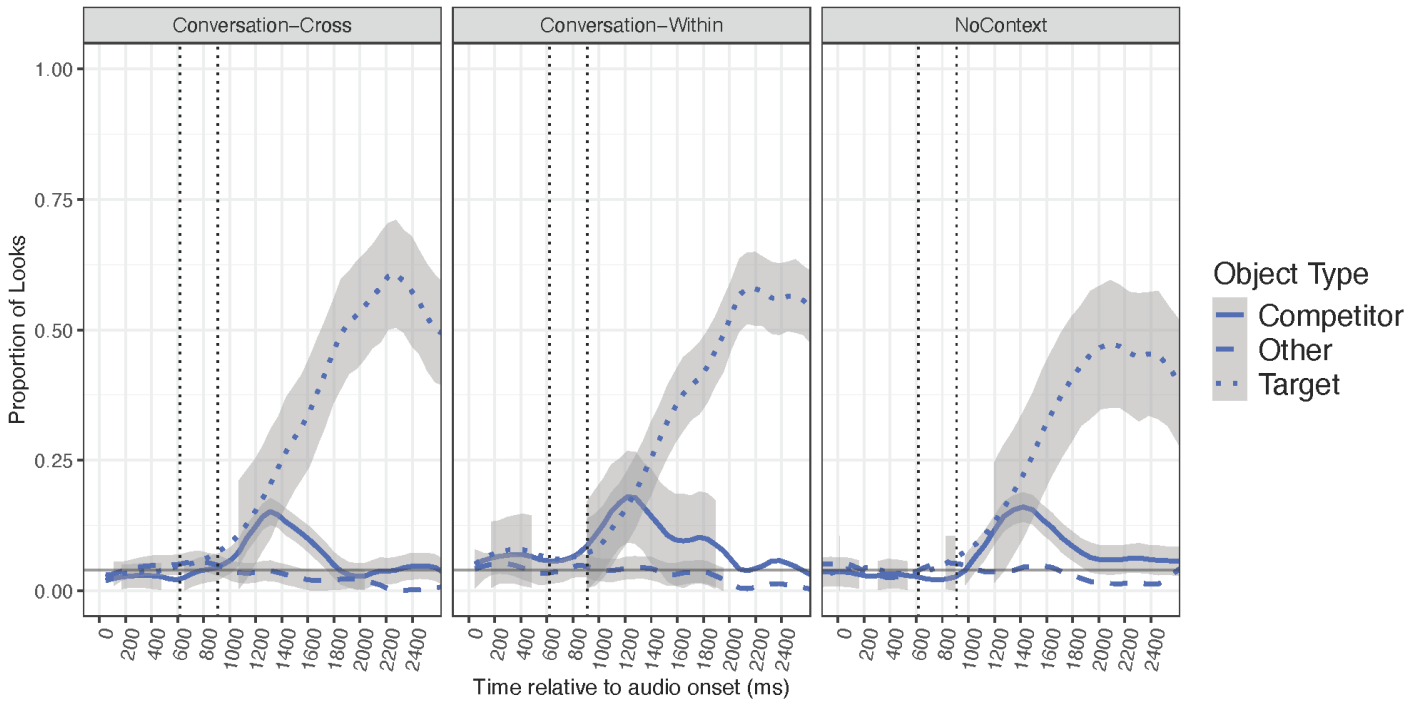


Figure 28. Average proportion of fixations (looks) to each object type per total fixations during each 50-millisecond time slice for each condition. Object Types: Target = object actually mentioned; Competitor = object that is the same color/can be described by the same adjective as the target; Other = object one space away from the competitor object. Conditions: Cross = competitor object is not relevant to the current topic, presented within the conversation context; Within = competitor object is relevant to the current topic, presented within the conversation context; No Context = critical sentences presented without conversation context. Times represent the time in milliseconds since the onset of the sentence. Shaded regions represent standard error. Dotted vertical lines show the time window of the onset of the adjective (e.g., black) to the onset of the target noun (e.g., microphone), which is the time region where looks to the target are predictive (i.e., because the target has not yet been said). The horizontal grey line shows the chance level of looking to any given item (1/25 = 0.04).

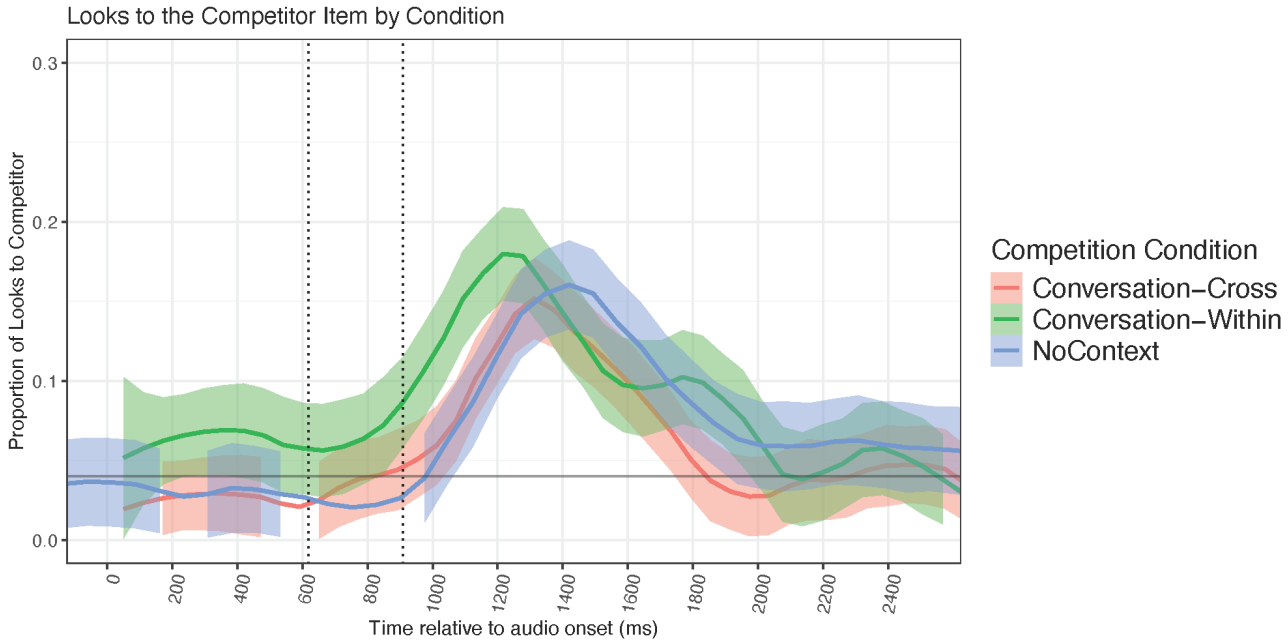


Figure 29. Average proportion of fixations (looks) to the competitor object per total fixations during each 50-millisecond time slice. Times represent the time in milliseconds since the onset of the sentence. Shaded regions represent standard error. Dotted vertical lines show the time window of the onset of the adjective to the onset of the target noun, which is the time region where looks to the target are predictive (i.e., because the target has not yet been said). The horizontal grey line shows the chance level of looking to any given item (1/25 = 0.04).

The modeling results demonstrated a significant difference in the proportion of looks and pattern of looking between competitor objects and other objects, from 360 (No Context) or 383 (Cross) milliseconds after the adjective onset until the end of the trial (No Context trials: Tables 28, 29, Figure 30; Cross trials Table 30, Figure 31). For Within trials, there was a higher proportion of looks to the competitor versus the other object prior to the adjective onset through almost the entirety of the trial (-200 to 1818 milliseconds post adjective onset; Table 31; Figure 32). As expected, there were no significant differences in looks to the other object by condition (Table 28). This confirms the pattern observed by visual inspection and shows that participants looked to objects that were temporarily consistent with the linguistic input more than items in the same general area that were unrelated to the linguistic input. Thus, further effects can be interpreted similarly to previous visual world studies.

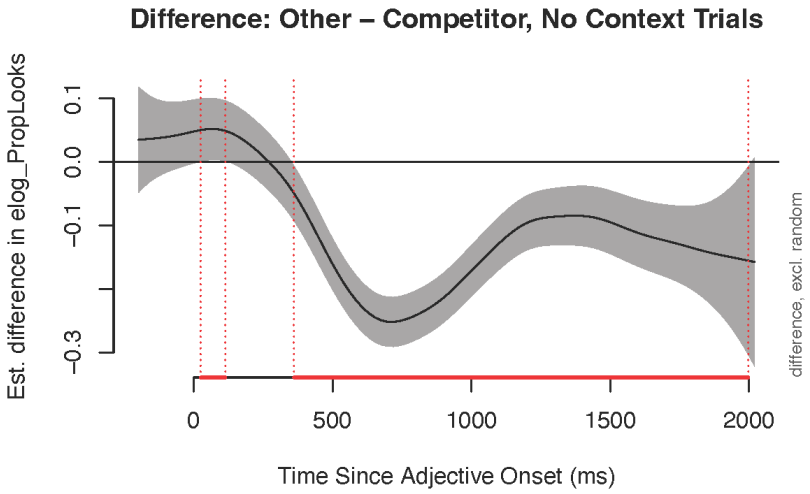


Figure 30. Difference curve showing model estimated empirical logit of the proportion of looks to the Other - Competitor on No Context trials. Positive values indicate more looks to the Other object. Time represents the time in milliseconds relative to the onset of the adjective. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

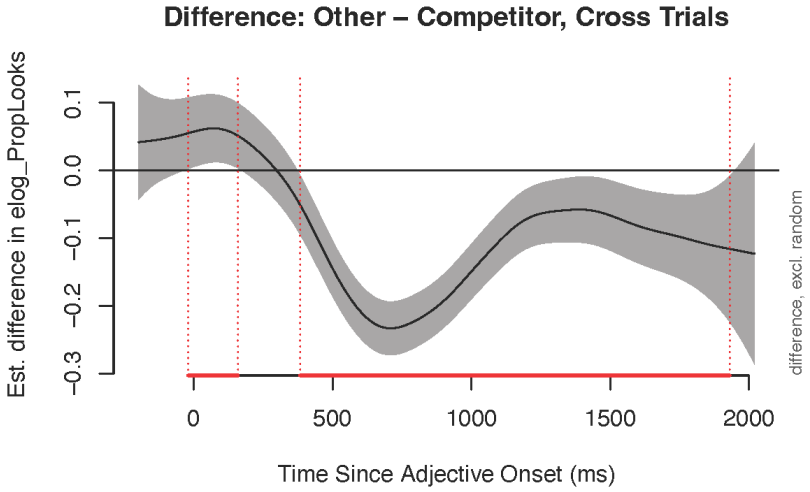


Figure 31. Difference curve showing model estimated empirical logit of the proportion of looks to the Other - Competitor on Cross trials. Positive values indicate more looks to the Other object. Time represents the time in milliseconds relative to the onset of the adjective. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

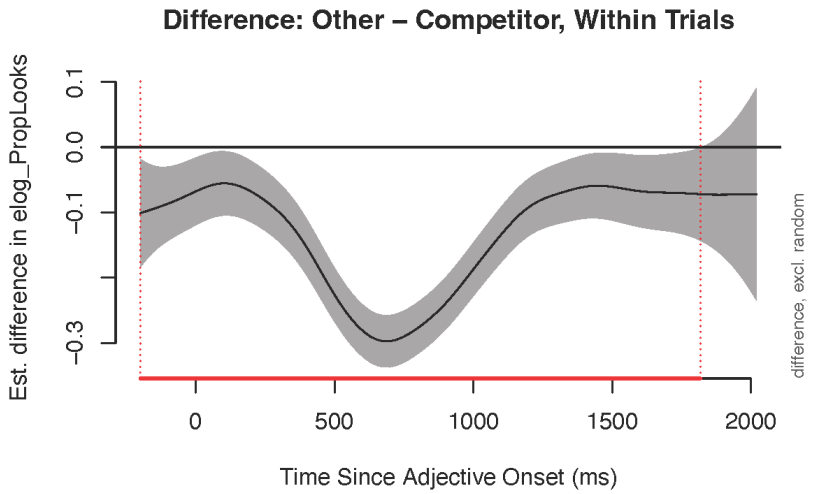


Figure 32. Difference curve showing model estimated empirical logit of the proportion of looks to the Other - Competitor on Within trials. Positive values indicate more looks to the Other object. Time represents the time in milliseconds relative to the onset of the adjective. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-1.05	0.01	-127.19	<0.001
IsCompetitor	0.11	0.01	9.12	<0.001
IsWithin	-0.01	0.01	-0.59	0.55
IsCross	-0.00	0.01	-0.27	0.78
IsWithinIsCompetitor	0.04	0.02	1.96	0.05
IsCrossIsCompetitor	-0.02	0.02	-1.15	0.25

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	1.22	1.40	1.84	0.12
Time:IsCompetitor	9.83	12.14	23.51	<0.001

Time:IsWithin	1.00	1.00	0.09	0.76
Time:IsCross	1.36	1.61	1.85	0.24
Time:IsWithinIsCompetitor	1.00	1.00	7.94	0.005
Time:IsCrossIsCompetitor	1.00	1.00	0.15	0.70

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Time, Participant	0.80	1024.00	0.00	0.26
Time, Participant:IsCompetitor	34.63	1024.00	0.04	0.001
Time, Participant:IsWithin	1.72	1023.00	0.00	0.12
Time, Participant:IsCross	13.79	1024.00	0.02	0.07
Time, Participant:IsWithinIsCompetitor	90.53	1024.00	0.15	<0.001
Time, Participant:IsCrossIsCompetitor	11.43	1024.00	0.01	0.06

Table 28. Model outputs for the model examining the effect of object type and condition on the empirical logit of the proportion of looks to the competitor. Reference levels = Other, No Context

Releveling the models to examine the competitor object as the reference condition and to examine each condition as the reference level revealed several significant effects. There was a significant difference in the pattern of looks to the competitor over time when comparing Within to No Context trials. As shown in Figure 33, when the competitor was relevant to the current topic during the conversation, participants looked to the competitor earlier and more than when the same sentence was presented with no conversation context (Table 29). Specifically, participants had a higher proportion of looks to the competitor on Within trials than on No Context trials prior to adjective onset until 832 milliseconds post adjective onset. Participants also looked less at the competitor in the Within condition than the No Context condition late in the trial, from 1437 milliseconds post adjective onset until the end of the trial. This pattern suggests that the overall response curve was shifted earlier in time on Within versus No Context trials (Figure 33).

While visual inspection of the raw data suggests fewer looks to the competitor on Cross versus No Context trials, there was no significant difference between the Cross and No Context conditions (Table 30). The Cross versus No Context difference is smaller in magnitude than the Within versus No Context comparison, so if there is a true effect it may be more difficult to detect with the current dataset.

However, there was a significant difference in looks to the competitor between Within and Cross trials (Table 31). Figure 34 shows that when comparing the two within-conversation conditions, there was a higher proportion of looks to the competitor on Within trials than on Cross trials and the looks to the competitor started earlier in time for Within trials.

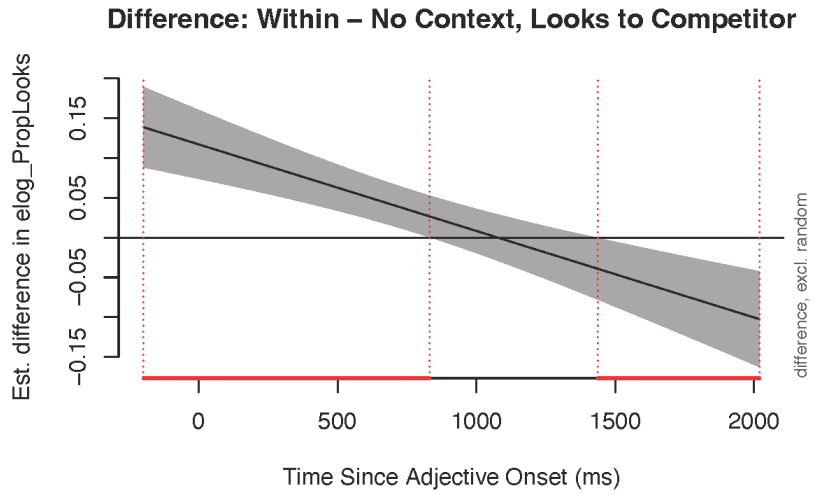


Figure 33. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Within - No Context trials. Positive values indicate more looks to the competitor on Within trials. Time represents the time in milliseconds relative to the onset of the adjective. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

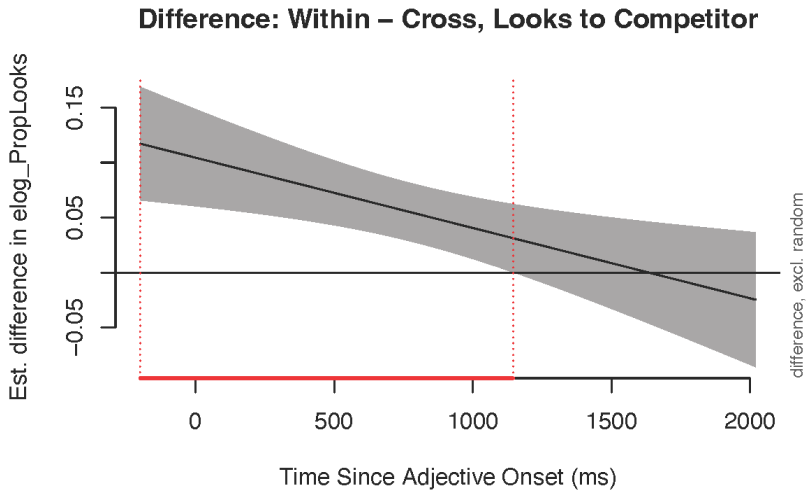


Figure 34. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor on Within - Cross trials. Positive values indicate more looks to the competitor on Within trials. Time represents the time in milliseconds relative to the onset of the adjective. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.94	0.01	-113.70	<0.001
IsOther	-0.11	0.01	-9.13	<0.001
IsWithin	0.03	0.01	2.31	0.02
IsCross	-0.02	0.01	-1.93	0.05
IsWithinIsOther	-0.04	0.02	-2.12	0.03
IsCrossIsOther	0.02	0.02	-1.16	0.25

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	10.12	12.50	28.21	<0.001
Time:IsOther	8.11	10.08	16.60	<0.001

Time:IsWithin	1.00	1.00	23.63	<0.001
Time:IsCross	1.76	2.18	2.57	0.08
Time:IsWithinIsOther	1.00	1.00	10.68	0.001
Time:IsCrossIsOther	1.00	1.00	0.17	0.68

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Time, Participant	0.37	1024.00	0.00	0.26
Time, Participant:IsOther	0.00	1024.00	0.00	0.99
Time, Participant:IsWithin	36.71	1024.00	0.05	<0.001
Time, Participant:IsCross	17.74	1024.00	0.02	0.08
Time, Participant:IsWithinIsOther	0.00	1024.00	0.00	0.73
Time, Participant:IsCrossIsOther	0.00	1024.00	0.00	0.96

Table 29. Model outputs for the model examining the effect of object type and condition on the empirical logit of the proportion of looks to the competitor. Reference levels = Competitor, No Context

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.96	0.01	-111.31	<0.001
IsOther	-0.09	0.01	-7.12	<0.001
IsWithin	0.05	0.01	3.98	<0.001
IsNoContext	0.02	0.01	1.82	0.07
IsWithinIsOther	-0.06	0.02	-3.20	0.001
IsNoContextIsOther	-0.02	0.02	-1.16	0.25

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value
Time	10.10	12.47	23.48	<0.001

Time:IsOther	8.11	10.08	16.60	<0.001
Time:IsWithin	1.00	1.00	7.83	0.005
Time:IsNoContext	2.56	3.20	2.27	0.07
Time:IsWithinIsOther	1.00	1.00	7.75	0.005
Time:IsNoContextIsOther	1.00	1.00	0.17	0.68

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Time, Participant	0.00	1024.00	0.00	0.43
Time, Participant:IsOther	0.00	1024.00	0.00	0.99
Time, Participant:IsWithin	36.01	1024.00	0.05	<0.001
Time, Participant:IsNoContext	13.51	1024.00	0.02	0.07
Time, Participant:IsWithinIsOther	0.00	1024.00	0.00	0.73
Time, Participant:IsNoContextIsOther	0.00	1024.00	0.00	0.99

Table 30. Model outputs for the model examining the effect of object type and condition on the empirical logit of the proportion of looks to the competitor. Reference levels = Competitor, Cross

Parametric Terms

Term	Parameter Estimate	Standard Error	t Statistic	P Value
Intercept	-0.91	0.01	-104.98	<0.001
IsOther	-0.14	0.01	-11.61	<0.001
IsCross	-0.05	0.01	-4.26	<0.001
IsNoContext	-0.03	0.01	-2.36	0.02
IsCrossIsOther	0.06	0.02	3.20	0.001
IsNoContextIsOther	0.04	0.02	2.12	0.03

Smooth Terms

Term	edf	Ref.df	F Statistic	P Value

Time	10.12	12.49	24.14	<0.001
Time:IsOther	8.10	10.07	15.18	<0.001
Time:IsCross	1.00	1.00	7.41	0.007
Time:IsNoContext	2.58	3.23	8.60	<0.001
Time:IsCrossIsOther	1.00	1.00	7.74	0.005
Time:IsNoContextIsOther	1.00	1.00	10.66	0.001

Random Effects

Term	edf	Ref.df	F Statistic	P Value
Time, Participant	0.00	1024.00	0.00	0.32
Time, Participant:IsOther	0.00	1024.00	0.00	0.99
Time, Participant:IsCross	17.28	1024.00	0.02	0.09
Time, Participant:IsNoContext	13.33	1024.00	0.02	0.07
Time, Participant:IsCrossIsOther	0.00	1024.00	0.00	0.97
Time, Participant:IsNoContextIsOther	0.00	1024.00	0.00	1.00

Table 31. Model outputs for the model examining the effect of object type and condition on the empirical logit of the proportion of looks to the competitor. Reference levels = Competitor, Within

There were also significant interactions of condition and object type (Tables 28, 29, 31). Figure 35 shows that the increase in looks to the competitor in the Within condition relative to the Cross condition was larger than the difference in looks to the other object in the Within versus the Cross condition from -200 to 1011 milliseconds post adjective onset. There was also a similar interaction by object type for Within versus No Context trials from -200 to 786 milliseconds post adjective onset, as shown in Figure 36. These results suggest, first, that the manipulation of relevance to the conversation topic was specific to looks to the competitor object that is related to the language input, and second, that contextual information from the conversation topic caused participants to consider possible referents earlier than when the possible referent was relevant to the current topic than when it was not relevant to

the current topic or when there was no discourse context. There were no significant interactions for contrasts relative to no context trials.

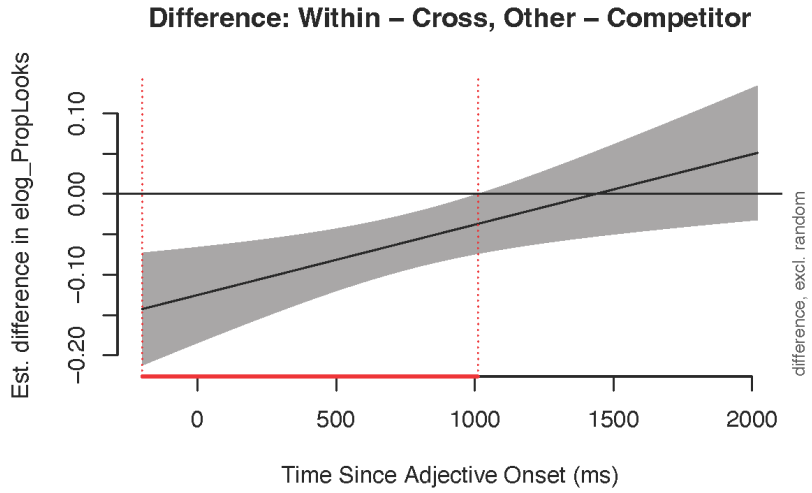


Figure 35. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor for the interaction of condition and object type, Within – Cross difference for Other – Competitor object. Positive values indicate more looks to the competitor on Cross than Within trials for the Competitor than for the Other object. Time represents the time in milliseconds relative to the onset of the adjective. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

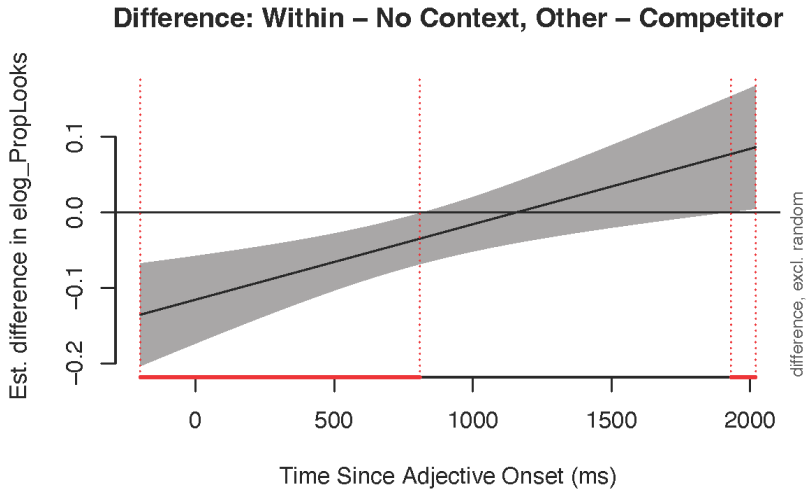


Figure 36. Difference curve showing model estimated empirical logit of the proportion of looks to the competitor for the interaction of condition and object type, Within - No Context difference for Other – Competitor object. Positive values indicate more looks to the competitor on No Context than Within trials for the Competitor than for the Other object. Time represents the time in milliseconds relative to the onset of the adjective. Shaded areas show the 95% confidence interval for the difference between the two conditions. Times highlighted in red show regions where the 95% confidence interval of the estimated difference does not include 0.

Discussion

The current study examined whether information about conversation topics is integrated with incoming language input in real-time in order to constrain the domain of reference to only topic-relevant information. Participants heard the same sentence containing a temporary referential ambiguity (e.g., "We should use the black…" when multiple black objects were in the scene) in three conditions while viewing the same visual scene items. In the No Context condition, sentences were presented one by one, without any surrounding discourse context. The target object in the scene matched the item mentioned in the sentence, and the competitor item could be described with the same adjective but was not the mentioned object. In the Within and Cross conditions, sentences occurred within the context of an ongoing conversation that alternated between two topics. On Within trials, the target and competitor items were both relevant to the current topic, while on Cross trials, the target was topic-relevant, but the competitor was not. Overall, the effects showed that relative to No Context trials, participants had a greater

proportion of looks to the competitor and this looking behavior occurred earlier when the competitor was topic-relevant on Within trials, but there was no difference in looks to the competitor on Cross trials.

Task Feasibility

On all trial types, participants looked at both the target and the competitor item at greater than chance levels, demonstrating that they considered multiple possible referents that were temporarily consistent with the incoming language input. This result is consistent with previous work using only sentence-level input and using less complex visual displays, demonstrating the feasibility of using more complex visual world setups than the typical four-item configuration (Allopenna et al., 1998; Arnold et al., 2004; Hsu & Novick, 2016; Kukona et al., 2014; Langlois et al., 2024; Nozari et al., 2016). Based on the onset of significant differences between looks to the competitor versus the other object on No Context trials, participants first began to look at the competitor item approximately 360 milliseconds after the onset of the adjective eliciting the temporary ambiguity. This timing is consistent with other studies using simpler visual displays with temporary referential ambiguity across a variety of linguistic and visual manipulations (Eberhard et al., 1995; Kukona et al., 2014; Langlois et al., 2024; Lord et al., 2026; Nozari et al., 2016). This suggests that incremental language comprehension processes occur similarly across simpler visual and discourse contexts and the more complex context of the current study.

Topics' Influence on Online Comprehension

The timing and degree to which participants considered the temporarily consistent competitor object differed according to whether the competitor object was likely (Within condition) or unlikely (Cross condition) to be mentioned in the current conversation topic. Crucially, the topics alternated twice over the course of each vignette. Because both topics were part of the overall context, all of the objects were broadly relevant to the ongoing conversation. However, each objects' relevance to the *current* topic varied as the topics alternated. Relative to the No Context condition, which represents the base level of consideration of temporarily consistent competitor items, participants looked at the competitor earlier and had a greater proportion of looks to the competitor when the object was likely to occur in the current topic

(i.e., both the target and competitor were within the same topic). There was no difference in participants' looks to the competitor when there was no context versus when it was unlikely to occur in the current topic (i.e., target and competitor were across topics). However, there was a significant difference between the within-conversation conditions, such that participants considered the competitor less when it was unlikely to occur than when it was likely to occur given the current topic. These results demonstrate that topic information influences participants' interpretations of the incoming language signal as comprehension processes are ongoing. Because the topics changed multiple times during the conversation, these results suggest that listeners were keeping track of the current topic in real-time.

The fact that topics shifted also highlights that it is not only prior discourse context per se that can influence parsing decisions. If prior context in general affected parsing decisions, all objects should be more likely to be considered within the conversation context than when there was no context because ideas and other objects related to the competitors were previously mentioned. Instead, consideration of the competitor depended on the competitor's relationship to the *currently active* topic. Participants were less likely to look at the competitor on Cross trials, when it was not relevant to the current topic than on Within trials, even though the topic the object was related to was previously discussed within the same overall task and by the same speakers. These results therefore highlight that, along with factors like speaker identity, knowledge about common ground, or physical context constraints, conversation topics are an additional factor that dynamically modulates the influence of certain components of the broader context on online language comprehension.

Differential Effects of Within Versus Cross Topic Status

It was initially hypothesized that competitor items on Within trials would be considered similarly to the same competitor on No Context trials, while the primary effect of topic would be to constrain the domain of reference such that competitors would not be considered on Cross trials. Instead, differences from the No Context trials were found only for Within trials. On Within trials, consideration of the competitor object occurred earlier than on No Context trials, including the period prior to the target word

onset and even prior to the adjective onset. This result is consistent with listeners anticipating that topic relevant objects are generally more likely to be mentioned while that topic is ongoing. This effect is, on one hand, not surprising because it is consistent with many studies demonstrating that contextual information can inform predictions about what items are likely to be referenced prior to the onset of any disambiguating information (Altmann & Kamide, 1999; Arnold et al., 2004, 2007, Chapter 3 and references therein). On the other hand, the results of Brown-Schmidt and Tanenhaus (2008), a study closer in scope and complexity to the current work, specifically highlighted the ability of task-relevant features to *constrain* the domain of reference to eliminate task-irrelevant information, rather than to facilitate task-relevant information. It may be that facilitating effects would emerge in the spatially constrained unscripted language game context, but the analyses were not designed to examine this difference. Alternatively, the facilitating effects of topic in the current study may reflect differences in how physical constraints and common ground between interlocutors interact with ongoing language input versus how topic structure may affect these processes. Further examination of physical versus discourse-level contextual constraints on language processing may examine both concurrently, which could benefit our understanding of conversation in naturalistic environments.

On Cross trials, looks to the competitor were not significantly different from No Context trials. Participants' consideration of the competitor on Cross trials suggests that they evaluated referents temporarily consistent with the bottom-up language input, even though the competitor item was unlikely to occur in the current topic. This result is similar to prior work demonstrating that verb-semantic constraints do not eliminate consideration of competitor items that are temporarily consistent with the adjective within similar constructions (e.g., for sentences like "She will eat the red pear" participants will look to both a red heart and red pear, even though heart is inconsistent with the verb) (Kukona et al., 2014; Langlois et al., 2024; Nozari et al., 2016). While there were no differences relative to No Context trials, there were later and significantly fewer looks to the competitor on Cross relative to Within trials. The failure to detect a significant difference between Cross and No Context trials, despite the appearance

of such an effect in the raw data, precludes strong conclusions about whether topic information can suppress (versus just not affect) consideration of topic-irrelevant referents. However, the effect may occur but be too small to detect at the current sample size. Alternatively, it could also be the case that if all referents receive an overall contextual "boost" within the conversation relative to when there is no context, then the difference from Within trials could reflect an active reduction in consideration of topic-irrelevant referents.

The failure to clearly detect a constraint on the domain of reference also contrasts with the constraints to the domain of reference in Brown-Schmidt and Tanenhaus (2008). A possible explanation for the difference between the spatial focus in the prior work and the informational focus elicited by topics may be that topics are less rigid than clearly defined spatial regions. For example, the "kitchen sketch" topic in the current experiment discusses a bell as a relevant object. Many restaurant kitchens do have a bell to indicate when orders are ready, but not all kitchens do and most participants likely have more experiences in kitchens without bells than with them. In contrast to this possible variability in whether an object is topic-relevant for any given listener, all listeners would likely agree where an object is located in space. Further, speakers in a conversation may bring up information that is topic-relevant but generally unexpected. In the case of the cohort competitors in a spatially constrained area in Brown-Schmidt and Tanenhaus' study, the candidate items are clearly defined and the competitors outside of the relevant area are also clear and consistent. Therefore, it may be a reliable strategy to pre-emptively ignore items that are known to be irrelevant in physical space, but it is a better strategy to initially remain open to multiple possible referents when the boundaries between relevant and irrelevant are fuzzier in regards to topics. While some of this effect may be attributable to factors specific to the experimental designs of these studies, the concept of being ready to contend with unexpected information in regards to topics is theoretically consistent with how topics frequently change over the course of a conversation and how many topic shifts are "smooth," "fuzzy," or "shaded" rather than abrupt and explicitly marked (Hobbs, 1990; Leaman & Edmonds, 2020; Riou, 2017). Future work could examine factors that influence how

open listeners might be to considering topic-irrelevant referents, such as the degree of unexpectedness for particular items, individual differences in background knowledge or knowledge about how much particular conversation partners engage in topic shading versus explicit transitions, or active engagement in guiding the conversation (i.e., versus "taking a back seat" listening to what others are talking about).

Taken together, the effect of a currently active topic seems to be to pre-emptively facilitate consideration of topic-relevant information and either have no effect on or possibly reduce consideration of topic-irrelevant information. When topics shift, it appears that listeners must change the relative level of facilitation for any particular item in the discourse context (visual and linguistic). This rapid updating of relative facilitation of some subsets of information is consistent with hypotheses put forth by Structure Building Framework and general event segmentation theories (Gernsbacher, 1997; Zacks et al., 2007). Although in this study participants only listened to conversations, future work could examine the same kinds of effects when participants are engaged in the conversation (e.g., by planning out the sketches) and assess whether they are related to shift costs in production (see Chapter 2). Understanding the specific effects of topic shifts in this way could constrain hypotheses about the representational format or neural instantiation of topic structure, beyond what can be inferred from discourse analysis or formal modeling approaches alone.

Relating Topic and Context in Online Language Comprehension

Questions regarding the effects of context, including discourse context, world knowledge, and physical context, on moment-to-moment sentence comprehension are not new to the psycholinguistic literature. For the last 30+ years, studies examining constraint-based models of sentence processing have demonstrated that multiple sources of information, including broader discourse context, are integrated to influence moment-to-moment parsing decisions as language input unfolds (M. C. MacDonald et al., 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1993, 1994; Trueswell & Tanenhaus, 1994) (for a review, also see McRae & Matsuki, 2013). However, what "context" means in light of these studies

requires unpacking in order to understand how topic-related effects in the current study relate to the broader literature on context effects.

Studies manipulating expectations about upcoming words or manipulating semantic interpretations have typically used short narrative or expository vignettes surrounding target words or sentences (Hagoort et al., 2009; Hald et al., 2007; Metusalem et al., 2012). For example, in the well-known "when peanuts fall in love" experiment, critical sentences are embedded within a six-sentence narrative that support or do not support an interpretation of a referent as animate and subsequent animacy violations are examined (Nieuwland & Van Berkum, 2006). Discourse context may also be instantiated as a suggestion of the types of things that might happen, such as being instructed that sentences are about a fictional world where unusual things can happen versus in the real world (Ness et al., 2024). In visual world experiments addressing the physical context, the physical setup is generally simple, all items are potentially applicable to the trial, and, usually, once the trial is over the physical context changes (Marslen-Wilson, 1987; Sedivy et al., 1999; Snedeker & Trueswell, 2004; Tanenhaus et al., 1995).

Across these studies, "context" is conceptualized as a whole, or a unit of information that can be combined with other units of information (e.g., a probability such as verb bias, a lexical feature like animacy). There is no need to update, change, or alter something about the discourse representation during any particular trial and once the set of trials is over there is no need to maintain information about the informational or physical context on the next set. However, in more naturalistic communication, like conversations, theories regarding what it means to be "on topic" would suggest that the value or relevance of any given aspect of the context can vary depending on the discourse structure. While the entire conversation constitutes a communicative context, topic changes occur within this broader context and previous topics may be returned to. Every unfolding sentence can be interpreted not only in the broad context of all the information (physical or linguistic) previously discussed or known, but also within a current topic, which should impose limits on what *subset* of information from the context is appropriate to consider (Hobbs, 1985, 1990; Van Kuppevelt, 1995; Roberts, 2012; van Dijk, 1979). Therefore, it may be

said that the previous studies had examined effects of information beyond the current linguistic input on moment-to-moment parsing decisions, but had not considered what happens when the availability, applicability, or relevance of that information changes as a function of the communication event structure.

A graphical representation of how the topic-related effects of the current study may relate to context in general is shown in Figure 37. Knowledge and physical context encompass all of the possible information that could exist within a particular conversation event. Knowledge may include world- or event-knowledge like facts, schemas, or episodic memories. The physical context may include any objects or interactions between objects (events) that could occur in the physical environment where the communication takes place. The black box denoting the discourse context represents any knowledge or physical context that has actually occurred during the communication event (e.g., a communication partner reports a fact; a communication partner picks up an object). This is a subset of all the possible information that could be part of the communication event. The yellow topic circles represent the subset of information that has already been discussed or has happened as well as related information from communication partners' knowledge or from the physical context that has not been discussed yet or happened yet. Topics may draw from the prior discourse context, but could also draw from information that has not been discussed yet at all. Although knowledge, the physical environment, and the prior discourse knowledge continue to exist, topics constrain the information that is relevant to consider at any given point in the conversation. The filled yellow circle represents the currently active topic representation. After new information within the topic is mentioned, the prior discourse context expands to encompass this information.

In the context of this schematic, prior psycholinguistic studies have focused on the effects of knowledge, physical context (green circles), and prior discourse context (black rectangle) on real-time language comprehension. The current work demonstrated that topics (yellow circles), which can co-exist within the same broader knowledge, physical, and prior context space but vary based on conversational structure and goals, also exert influence on the interpretation of incoming language input. In particular,

the results of the current study suggest that the current topic can dynamically modulate what is relevant to consider from the incoming input, even within the subset of information in the current discourse context (black rectangle). Participants did consider competitors more when they were topic-relevant than when they were irrelevant or presented without a conversation context, and this pattern depended on the topic that was actively being discussed.

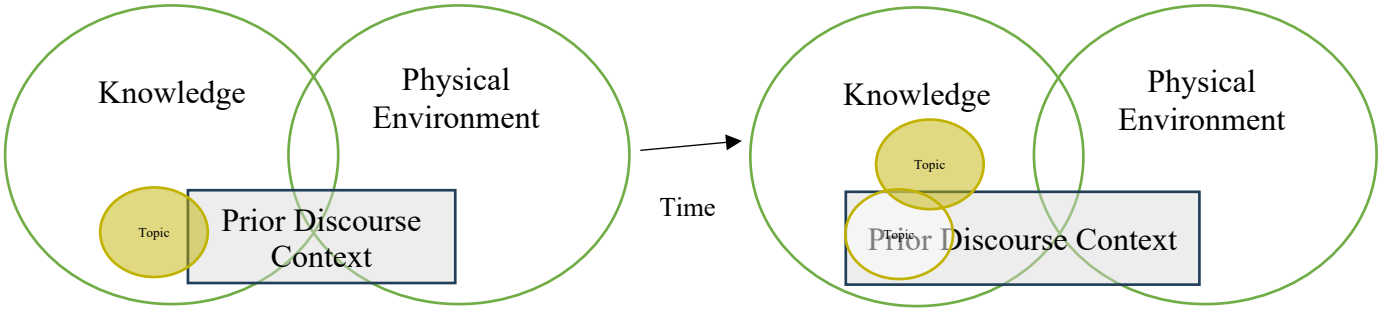


Figure 37. Graphical representation of top-down information structure available to influence language comprehension as the linguistic signal unfolds.

Future Directions

The results of the current study demonstrated the feasibility of using this task design to examine real-time tracking of topics and the interaction of topics with ongoing bottom-up language processing. Future studies could vary different parameters of the conversation input in order to examine discourse structure or real-time language processing. For example, varying the explicitness of the topic shifts could reveal whether more subtle topic shifts change the facilitation of topic-relevant referents or later reduction in consideration of topic-irrelevant referents. Future work may also examine whether the participant's role in the conversation changes the effect of topic structure. For example, when participants must both plan and produce utterances as well as listen, there may be differences in their use or ability to track topic transitions. Participants could engage in conversation about the same sketch-based task while their eye movements are tracked. Additionally, because this study has demonstrated sensitivity to topic shifts

within the study materials, the same conversations without the visual context could be used to examine cognitive effort at known topic transitions using pupillometry.

The representation of the status of all of the contextual information structures addressed in Figure 37 should correspond to the concept of common ground (Clark, 2021; Heller & Brown-Schmidt, 2023). Within the scope of this study, only the influences of these information sources for one individual within the conversation were discussed. However, an individual's concept of common ground also includes representations about what other communication partners think or know. Future work may examine how topic information is integrated into what speakers believe about each other's contextual interpretations. For example, do conversation partners share the same topic representations at the same time? Or can one's current topic representation differ from the representation about what the other partner thinks the topic is?

Additionally, a key objective of the current study was to establish a paradigm that could be used to examine possible differences in language comprehension related to discourse deficits in individuals with TBI. While the pace and visual complexity of the task could be overwhelming for individuals with more severe TBIs, participants in the chronic stages of recovery after moderate to severe TBI have been able to engage in unscripted language games with similarly complex visual displays which yield good quality eye tracking data (Lord et al., 2026). Examining the performance of different groups of individuals with TBI could answer important open questions about their possible discourse deficits. For example, if individuals with moderate to severe TBI do have deficits in tracking conversation topics, as suggested by smaller shift costs in Chapter 2, they would be expected to have a smaller difference in performance on both kinds of conversation context trials versus No Context trials.

While the results of Chapter 3 suggest that individuals with (primarily) mild TBI are able to engage in important component processes that should support tracking topic transitions, such as tracking prior discourse information and using cues to make predictions about upcoming given versus new information, the current task would allow for the examination of actually tracking topic transitions in a more complex

discourse context. If this group does not have difficulty tracking topics, using this task could reveal other differences in online language processing that may contribute to subjective complaints of discourse difficulty, such as the possibility of strong commitment to initial predictions discussed in Chapter 3. This could manifest as reduced consideration of contextually relevant possible referents on Within trials or failure to stop considering the initially considered topic-irrelevant object on Cross trials. In other words, even if individuals with mild TBI can identify topic transitions, they may not be able to effectively use topic information to facilitate or constrain online language processing.

Limitations

While this task was designed to strike a balance between unstructured conversation and decontextualized language, additional studies are needed to tease apart task-specific effects from effects related to topic comprehension. One factor is that the content of the conversations is somewhat unconventional. While comedy, fiction, and unexpected but real scenarios are things that people converse about, participants may have different expectations or modulate their expectations about possible referents differently than if they were discussing something more mundane. It could be the case that the unusual content would cause participants to take an approach that any items are possible referents, or that topics should not be constraining. The results demonstrated that participants did use topic information in ways consistent with theories of conversation topic, but there may still be differences between the use of topic information in this versus more naturalistic contexts. In addition to being unconventional, the content or question under discussion in each topic was similar across the experiment. While the sketches were describing different sets of events, the content was primarily about what event should happen for a particular sketch. In naturalistic conversation, a larger variety of topics would be addressed, including things like discussing past experiences, explaining tasks or processes, or discussing feelings or opinions. The type of information discussed within a topic may also influence how that topic interacts with the ongoing task and with incoming language input.

Similarly, the unconventional content and the use of descriptive adjectives required familiarizing participants with the names of the items before the experiment. In real-world contexts, listeners may have more uncertainty about how speakers would refer to any given item or idea and thus it may be more difficult to predict likely referents from incoming input.

An additional limitation is that participants in this study were not actively participating in the conversation. Other studies of discourse have demonstrated that referring behavior and interpretation can differ based on whether an individual is an active or passive conversation partner (Brown-Schmidt et al., 2015). Future studies may use a similar paradigm where participants converse with other participants or research assistants about the contents of the sketches to examine whether the influence of the topic changes when it is being actively constructed by the participant.

The current analysis also did not address certain factors that may impact when and how topic information is applied to influence online language comprehension. For example, some of the objects were better fits, or more likely to be thought of as relevant to the topic, than others. Thus, the facilitating effect of topic on the consideration of any particular object may be related to how relevant an item is given a particular topic. Additionally, the current analysis assumed that all time points within a topic are equally "on" the same topic. It is possible that topic representations emerge over time, such that the effect of topics may be weaker closer to topic transitions and stronger as confidence in the current topic representation has accumulated. Alternatively, it could be that topic representations are stronger at the point of transition and weaken as more, potentially unrelated, information is introduced as interlocutors run out of things to talk about. Using longer conversations, examining the within and cross effects as a function of time since topic onset could address this open question.

General Conclusions

This study demonstrated that healthy adult participants keep track of conversation topics in real-time and that topics interact with online language processing to facilitate consideration of topic-relevant referents. However, the current topic appears not to affect consideration of topic-irrelevant referents

relative to when there is no context. This study provides empirical evidence to support the key assumption in the discourse literature that interlocutors are representing something about topics moment to moment. These results demonstrate not only that discourse context influences online language processing, but also that listeners represent which subsets of the discourse context are relevant at any given point within an ongoing discourse. Importantly, these results reflect participants' own interpretations of naturalistically generated conversations, without the need for third-party raters to make assumptions about topic content or boundaries. Additionally, this study adds to the current literature on conversation by examining the effect of topics on language comprehension, while the field has been predominantly interested in how topics shape language production. Finally, this task design provides a novel method to supplement barrier tasks and reference games (Brown-Schmidt et al., 2015; Brown-Schmidt & Konopka, 2008; Brown-Schmidt & Tanenhaus, 2008) as ways to use non-invasive eye tracking methods to examine real-time language comprehension in more complex and more naturalistic contexts.

Chapter 5: General Discussion

The aims of this dissertation were (1) to test whether listeners keep track of changes in conversation topic in real-time, (2) to examine whether and how tracking these topic changes interacts with ongoing language comprehension and production, and (3) to delineate differences in these topic tracking processes in individuals with TBI. Together, these aims provide a framework for investigating whether individuals with TBI's subjective reports of difficulty following along in conversation are related to the temporal dynamics of language processing in complex communicative tasks. This key component of cognitive-linguistic competence has, so far, been difficult to examine because of the theoretical and empirical complexity of studying conversation and a focus on broader discourse organization deficits versus online language processing in TBI. The primary hypotheses were that listeners do track discourse topics in real-time, that individuals with TBI differ from healthy adults in their ability to track topic changes due to one or more components of the discourse structure building process (i.e., memory, discourse status prediction, or model updating), and that differences in topic tracking abilities affects ongoing production and comprehension during conversations. Chapter 2 examined sensitivity to topic transitions, interactions between topic tracking and ongoing language production, and differences in these abilities between healthy adults and adults with moderate to severe TBI by testing whether there was a cost to language production in unstructured discourse on conversational turns that followed a topic shift versus those that continued on the same topic of conversation. Chapter 3 compared the moment-to-moment language comprehension of individuals with mild to severe TBI and healthy adults using constrained, short discourse tasks that addressed component cognitive-linguistic abilities hypothesized to support topic tracking. Chapter 4 introduced a novel eye tracking paradigm addressing the interaction of topics and ongoing language comprehension during longer discourses with multiple topic transitions.

Investigating Topic Comprehension in Real-Time

Engaging in conversation requires individuals to be both speakers and listeners, co-constructing the structure of the discourse event as it unfolds over time. The high variability in content, goals, physical context, and partner relationships in spontaneous conversation make the form and time course of the information structure much less certain than in rehearsed discourse or in monologues (e.g., narratives, exposition) (Brown-Schmidt et al., 2015; Clark, 2014; Togher et al., 1997). This variability could make it inefficient for interlocutors to track topics moment-to-moment, suggesting it may be more advantageous to take an approach such as waiting for enough discourse content to emerge to be sure of the structure before updating a higher-level representation. There has thus far been little empirical evidence demonstrating whether interlocutors keep track of topic representations moment-to-moment such that topic structures could influence ongoing language production and comprehension.

In Chapter 2, evidence that encountering a topic shift initiated by a communication partner led to shorter and less syntactically complex utterances with more instances of utterance planning difficulty suggests that processing topic shifts uses some limited cognitive resource that disrupts word- and sentence-level processes. While these studies relied on raters to determine where topic shifts were likely to have occurred, in Chapter 4, effects of topic shifts occurred at locations where those shifts were known to be intended by the speakers/experimenters. These results also demonstrated effects of the current topic on listeners' real-time interpretations of ambiguous referential utterances, suggesting that listeners must have been keeping track of which topic was currently under discussion in order to use this information to guide parsing decisions. By addressing questions about the availability and effect of topic shifts in real-time from multiple angles—testing effects on both comprehension and production and testing effects in highly naturalistic and more controlled contexts—we can be more confident that the observed effects are due to topic-level representations rather than to experimental idiosyncrasies, rater perceptions, or contextual variability.

In addition to suggesting that listeners are indeed tracking topic structures moment-to-moment, the observed effects of shifting topics on ongoing language production and comprehension in Chapters 2 and 4 are consistent with theoretical proposals of how individuals can impose higher-level structures on continually unfolding language or event information. The costs to language production following topic shifts relative to turns staying on the same topic are consistent with a model in which establishing new components of the higher-level structure occupies some cognitive resources, and models that propose that certain resources are shared between discourse- and word- or sentence-level language processes (Gernsbacher, 1997; Peach & Hanna, 2021; Zacks et al., 2007). The results of Chapter 4 suggest that once a topic is established, it facilitates listeners' predictions about upcoming on-topic linguistic input, while topic either has no effect, or possibly constrains consideration of possible online interpretations to exclude off-topic information. Modulation of which subsets of information are relatively more active, but without pre-emptive inhibition of off-topic information, is partially consistent with both Structure Building Framework and event segmentation theories. Increased expectation of other on-topic information in Chapter 4 could be consistent with the spreading of activation to semantically similar information posited to occur by Structure Building Framework, as objects relevant to the same topic may share some semantic or world-knowledge features. However, the care taken to group items by topic-relevance with equivalent semantic similarity between and across topics suggests something closer to the active prediction of subset-relevant information in event segmentation theories is more likely. Additionally, the fact that topic information did not inhibit consideration of off-topic items is also inconsistent with the active inhibition posited by Structure Building Framework. Although conversation is a type of discourse, the greater similarity to event-related theories than discourse-related theories on points where the models differ may be attributable to the fact that narrative and expository structures are more circumscribed and there are fewer possible elements to the discourse macrostructure, while both events in the physical world and conversations can have more possible generators of change (e.g., multiparty conversation, multiparty events) and possibilities for when and how the higher-level structure will change.

While additional studies are needed to test specific predictions from each of the models, a benefit of comparing the observed topic-related effects with these models is that the respective literatures investigating discourse structure and event cognition may provide clues as to the possible cognitive and neural representations of topics (Baldassano et al., 2017). Understanding both the effects of topic changes and possible neural correlates could help better understand why and how neural damage from TBI might affect conversation processing. One route to better address the problem of heterogeneity in drawing conclusions in group studies of TBI could be to investigate neural damage profiles that would be likely to affect systems found to be related to topic representation to identify which subset of individuals with TBI may have topic-related issues and what might be the most useful treatment approaches.

Differences in Comprehending Topic Shifts in TBI

Evidence of the interaction between topic representations and ongoing language production and comprehension confirms that the temporal dynamics of topic processing are a reasonable area of study when thinking about cognitive communication disorders. This is especially the case in TBI, where contextual complexity and/or cognitive resource allocation in more complex communication environments have been posited as a primary cause of cognitive and communication impairment (Kekes-Szabo et al., 2025; Lord et al., 2026; Turkstra et al., 2025).

Experiments in Chapters 2 and 3 tested differences between healthy controls and individuals with TBI in comprehension of topic shifts. The study of differences in shift costs in Chapter 2 demonstrated no overall group differences between healthy controls and individuals with moderate-to-severe TBI in the extent of topic shift-related costs in subsequent language production complexity and planning difficulty. However, individuals with more severe TBIs showed smaller shift costs than those with more moderate TBIs, suggesting that those with more severe TBIs may not have been tracking when the topic changed as frequently or as well. Analyses in Chapter 3 demonstrated no significant differences in the pattern of results between participants with mild to severe (but primarily mild) TBI and healthy adults in two visual world eye tracking tasks that tested the ability to track prior discourse information and speaker cues in

order to make predictions about whether upcoming information would be given or new. Taken together, these results suggest that individuals with TBI, except perhaps those with very severe injuries, do not have difficulty in detecting topic shifts or differences in some of the component processes that support detecting when topics will shift. This result is unexpected given the centrality of higher-level discourse structure deficits in discussions of cognitive-communication deficits in TBI (Coelho et al., 2002; Hill et al., 2018; Lê & Coelho, 2023; Levin et al., 2014) and difficulty with similar memory, prediction, and subset shifting abilities in other domains (Bamdad et al., 2003; Barlow et al., 2018; Gilmore et al., 2018; Kavé et al., 2011a; Macdonald & Johnson, 2005; Ord et al., 2010; Perianez et al., 2007; Solbakk et al., 2002; Zakzanis et al., 2011). However, it remains to be explored whether heterogeneity in the patient population has masked the fact that some individuals with TBI do have difficulty with tracking topic changes. Further investigation is also needed to determine if differences in how topic structures interact with ongoing language processing, and not identification of the structures per se, differentiate individuals with TBI from healthy adults. Ultimately, these results narrow the scope of possible hypotheses about what factors contribute to subjective reports of conversation difficulty and objective differences in descriptive studies of discourse behavior (Bond & Godfrey, 1997; Coelho et al., 1993, 2002; Grayson et al., 2021; Hill et al., 2018; Mentis & Prutting, 1991; Snow et al., 1997; Togher et al., 2010).

One additional takeaway from the experiments in Chapter 3 is that factors that influence cognition and language processing, such as age, post-traumatic stress symptoms, and neurobehavioral symptoms, can be accounted for in paradigms that assess online language processing. The results of Chapter 3 demonstrate that each of these factors affects the use of prior discourse information and disfluency cues to guide online comprehension, and that the effects are not necessarily the same as the effects of TBI. Online language comprehension measures like eye tracking may therefore be useful tool to differentiate the cognitive-linguistic effects of TBI per se from comorbid factors, particularly when other behavioral and self-report measures cannot (Lange et al., 2020, 2021; Lippa et al., 2021b). This is especially important in the military population because PTSD and TBI are highly comorbid but have different treatment

pathways. For the civilian population, being able to better differentiate effects of TBI from other individual factors also improves the ability to identify factors that may contribute to heterogenous outcomes in chronic TBI (Haarbauer-Krupa et al., 2021).

While hypothesized differences in tracking topic changes and the component processes that support topic tracking were not observed, individuals with TBI did show some differences in online language processing. In Chapter 3, when the given or new discourse status of an upcoming referent was predictable based on prior information in the discourse, participants with TBI initially looked less to unexpected referents than did healthy adults. Participants with TBI did not show lags in predictive looks to possible referents in other conditions, suggesting that this effect is not likely to be attributable to general processing speed deficits or generally less looking around the visual display. Rather, these results suggest that individuals with TBI may make stronger commitments to initial predictions about upcoming referents based on information they are able to track from the prior discourse and linguistic context. Given that the Chapter 3 experiments were not designed to test this particular effect, future studies are needed to directly examine initial commitments to interpretations during online language processing. However, these results highlight the importance of examining language processing over time, as the observed effects occurred early and TBI and healthy control groups became more similar over time. This short-lived effect could not be observed with methods typically used for assessing discourse abilities, such as discourse analysis or offline comprehension questions in standardized language assessments.

More generally, the absence of TBI versus healthy control group effects in regards to discourse-level topic structure tracking, alongside the presence of word- and sentence- level differences in production and online language processing differences in comprehension across Chapters 2 and 3, are consistent with emerging perspectives on how researchers and clinicians should approach language abilities in individuals with TBI. Lord et al. (2026) argue that a focus on de-contextualized language processes such as confrontation naming or single sentence interpretation, which is frequently the format of standardized language testing, has led to misunderstandings about the nature of language abilities in

TBI. They suggest that this focus has led to the prevailing view that TBI does not typically lead to language deficits and that communication changes after TBI are a result of cognitive difficulties unrelated to core language systems. Instead, they suggest that almost all real-world language use is contextual, meaning it takes place during tasks in the physical world, with other people, within discourse contexts, or a combination of these factors. Lord et al. (2026) demonstrate differences between participants with moderate to severe TBI and healthy adults in sentence planning and comprehension when a) completing a simple task with another person in a physical environment and b) incremental language processing can be measured over time. While other groups have not used temporally sensitive measures, several discourse analysis studies also demonstrate sentence-level differences between individuals with TBI and healthy adults when language use is examined within discourse tasks (Ghayoumi et al., 2015; Norman et al., 2022; Peach, 2013; Peach & Coelho, 2016; Peach & Hanna, 2021; Stubbs et al., 2018). While these results certainly do not suggest that non-linguistic cognitive processes are irrelevant in cognitive-communication disorders following TBI, they do suggest that clinicians and researchers should consider linguistic deficits below the level of discourse—particularly the use of word- and sentence-level language abilities in context—as possible contributors to the clinical presentation of individuals with TBI.

Looking forward, the results of the current studies suggest that visual world eye tracking and pupillometry can be useful methods for studying contextualized language processing in individuals with TBI. Chapter 3 demonstrated that both of these methods are feasible and able to detect relevant language processing signatures in individuals in the chronic stages of recovery from mild to severe TBIs. Feasibility with patients who are earlier in the recovery process, who may have more behavioral symptoms, light or sound sensitivity, or fatigue, is still to be determined. Although the current study used a research-grade, stationary eye tracker, the increasing availability of less costly and more portable commercial devices with eye tracking capabilities (e.g., virtual reality displays, phones or tablets) suggests that eye tracking or pupillometry measures may soon be feasible in clinical settings (Phillips et al., 2023). Chapter 4 demonstrated that visual world eye tracking can provide useful information about

online language processing even in visually and linguistically complex environments. Examining online language processing during the kinds of real-world environments individuals with TBI report difficulty in, such as in background noise or within multiparty conversation, therefore seems to be a promising future direction for understanding cognitive-linguistic difficulties.

Conclusion

The theoretical complexity and empirical difficulty of examining online language processing during conversations have made it difficult for individuals with TBI, clinicians, and clinical researchers to conceptualize and treat this aspect of cognitive-communication disorders. By using a variety of techniques, the current studies have demonstrated the feasibility of studying the effects of shifting topics on ongoing language production and comprehension. The current work has demonstrated that healthy adults and individuals with TBI are sensitive to rapidly shifting topic information in real-time. While individuals with TBI appear to perform similarly to healthy adults when using information from the previous discourse and from interlocutors to determine when discourse information is likely to shift, there may be differences in how individuals with TBI use this discourse-level information in conjunction with bottom-up language input. These results point to the importance of examining not only the effect of TBI on discourse structure, but also on interactions between discourse-level and word- or sentence-level processes.

Appendices

Appendix 1

Primary Analysis, Chapter 2

Supplementary Table 1. Model outputs for the effects of group and subtopic status on the words/utterance measure. Words = number of words, Group = factor representing TBI or NBI (reference level – NBI), New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), ID = participant, Total_Utts = number of utterances.

Model: glmer(Words ~ Group*New + Age + Education + offset(log(Total_Utts)) + (1+New|Group/ID), data=data, family="poisson")				
Term	**Parameter Estimate**	**Standard Error**	**Z Statistic**	**P Value**
Intercept	1.94	0.01	194.49	<0.001
GroupTBI	-0.11	0.01	-7.84	<0.001
New1	-0.10	0.01	-10.91	<0.001
Age	0.01	0.00	9.31	<0.001
Education	0.02	0.00	6.37	<0.001
Group*New	-0.02	0.01	-1.74	0.08

Supplementary Table 2. Model outputs for the effects of group and subtopic status on the syntactic complexity measure. Verbs = number of verbs, Group = factor representing TBI or NBI (reference level – NBI), New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), ID = participant, Total_Utts = number of utterances.

Model: glmer(Verbs ~ Group*New + Age + Education + offset(log(Total_Utts)) + (1+New|Group/ID), data=data, family="poisson")				
Term	**Parameter Estimate**	**Standard Error**	**Z Statistic**	**P Value**
Intercept	0.16	0.02	7.12	<0.001
GroupTBI	-0.09	0.03	-2.97	<0.01
New1	-0.13	0.02	-6.22	<0.001
Age	0.01	0.00	4.33	<0.001
Education	0.02	0.01	2.56	0.01
Group*New	-0.04	0.03	-1.26	0.21

Supplementary Table 3. Model outputs for the effects of group and subtopic status on the semantic complexity measure. Density = idea density measure, Group = factor representing TBI or NBI (reference level – NBI), New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), ID = participant.

Model: lmer(arcsine(sqrt(Density) ~ Group*New + Age + Education + (1|Group/ID), data=data)					
Term	**Parameter Estimate**	**Standard Error**	**df**	**t Statistic**	**P Value**
Intercept	0.74	0.01	115.20	56.29	<0.001
GroupTBI	-0.01	0.02	115.00	-0.51	0.61
New1	-0.01	0.01	99.00	-0.86	0.39
Age	0.00	0.00	97.00	0.83	0.41
Education	0.00	0.00	97.00	1.55	0.12
Group*New	-0.00	0.01	99.00	-0.40	0.69

Supplementary Table 4. Model outputs for the effects of group and subtopic status on the revision measure. RepRetrace = number of repetitions + number of retracings, Group = factor representing TBI or

NBI (reference level – NBI), New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), ID = participant, Total_Syll = number of syllables.

Model: glmer(RepRetrace ~ Group*New + Age + Education + offset(log(Total_Syll)) + (1|Group/ID), data=data, family="poisson")

Term	Parameter Estimate	Standard Error	Z Statistic	P Value
Intercept	-3.92	0.07	-56.66	<0.001
GroupTBI	0.08	0.10	0.78	0.43
New1	-0.11	0.05	-2.43	0.02
Age	0.00	0.00	0.43	0.66
Education	-0.01	0.02	-0.48	0.63
Group*New	0.09	0.06	1.46	0.14

Supplementary Table 5. Model outputs for the effects of group and subtopic status on the planning measure. FilledPauses = number of filled pauses, Group = factor representing TBI or NBI (reference level – NBI), New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), Total_Syll = number of syllables, ID = participant.

Model: glmer(FilledPauses ~ Group*New + Age + Education + offset(log(Total_Syll) + (1|Group/ID)), data=data, family="poisson")

Term	Parameter Estimate	Standard Error	Z Statistic	P Value
Intercept	-4.15	0.08	-51.30	<0.001
GroupTBI	-0.09	0.11	-0.76	0.45
New1	0.35	0.04	8.30	<0.001
Age	0.00	0.00	0.51	0.61
Education	0.04	0.02	1.71	0.09
Group*New	-0.08	0.06	-1.23	0.22

Effect of Severity in the TBI Group, Chapter 2

Supplementary Table 6. Model outputs for the effects of severity and subtopic status on the words/utterance measure in the TBI group. Words = number of words, LOC_days = duration of length of coma in days (centered), representing injury severity, New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), ID = participant, Total_Utts = number of utterances.

Model: glmer(Words ~ LOC_days*New + Age + Education + offset(log(Total_Utts)) + (1+New|ID), data=data_tbi, family="poisson")

Term	Parameter Estimate	Standard Error	Z Statistic	P Value
Intercept	1.81	0.01	191.60	<0.001
LOC_days	-0.00	0.00	-7.53	<0.001
New1	-0.12	0.01	-12.87	<0.001
Age	0.00	0.00	2.66	<0.01
Education	0.03	0.00	5.80	<0.001
LOC_Days*New	0.00	0.00	4.70	<0.001

Supplementary Table 7. Model outputs for the effects of severity and subtopic status on the syntactic complexity measure in the TBI group. Verbs = number of verbs, LOC_days = duration of length of coma in days (centered), representing injury severity, New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), ID = participant, Total_Utts = number of utterances.

Model: glmer(Verbs ~ LOC_days*New + Age + Education + offset(log(Total_Utts)) + (1+New|ID), data=data_tbi, family="poisson")

Term	Parameter Estimate	Standard Error	Z Statistic	P Value
Intercept	0.07	0.02	3.21	<0.01

LOC_days	-0.00	0.00	-3.59	<0.001
New1	-0.18	0.02	-8.20	<0.001
Age	0.00	0.00	1.57	0.12
Education	0.02	0.01	1.55	0.12
LOC_Days*New	0.00	0.00	2.16	0.03

Supplementary Table 8. Model outputs for the effects of severity and subtopic status on the revision measure in the TBI group. RepRetrace = number of repetitions + number of retracings, LOC_days = duration of length of coma in days (centered), representing injury severity, New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), ID = participant, Total_Syll = number of syllables.

Model: glmer(RepRetrace ~ LOC_days*New + Age + Education + offset(log(Total_Syll)) + (1+New|ID), data=data_tbi, family="poisson")

Term	Parameter Estimate	Standard Error	Z Statistic	P Value
Intercept	-3.88	0.07	-57.89	<0.001
LOC_days	0.00	0.00	1.01	0.31
New1	0.02	0.05	0.44	0.66
Age	-0.00	0.01	-0.28	0.78
Education	0.03	0.03	1.04	0.30
LOC_Days*New	0.00	0.00	0.43	0.67

Supplementary Table 9. Model outputs for the effects of severity and subtopic status planning measure in the TBI group. FilledPauses = number of filled pauses, LOC_days = duration of length of coma in days (centered), representing injury severity, New = factor representing subtopic status (1=shift, 0=no shift, reference level = no shift), Age = age in years (centered), Education = years of education (centered), Total_Syll = number of syllables, ID = participant.

Model: glmer(FilledPauses ~ LOC_days*New + Age + Education + offset(log(Total_Syll)) + (1+New|ID), data=data_tbi, family="poisson")				
Term	**Parameter Estimate**	**Standard Error**	**Z Statistic**	**P Value**
Intercept	-4.29	0.08	-54.10	<0.001
LOC_days	-0.01	0.00	-2.60	<0.01
New1	0.31	0.06	5.49	<0.001
Age	0.00	0.01	0.50	0.62
Education	0.04	0.03	1.03	0.30
LOC_Days*New	-0.01	0.00	-3.01	<0.01

Shift Costs for Investigators, Chapter 2

In order to assess whether shift costs were consistent across conversational roles, we completed a complementary exploratory analysis on utterances produced by the researcher conversation partner (investigator) in response to ongoing subtopics vs. to subtopic shifts initiated by the TBI or NBI group participants. We used the same modeling procedures as described for the analysis of NBI and TBI group participants. However, we could not include covariates of age and education because these measures are not included in the corpus dataset for the investigator. We also could not include a random slope by subtopic status because the investigator did not always have both IPNEW and IPSAME codes for each participant.

If subtopic shift effects are a reflection of increased demands on cognitive processes, all conversation partners should show similar patterns of costs when responding to a shift initiated by the other partner. However, if subtopic shift effects are instead capturing something else about conversation dynamics (e.g., one partner needing to come up with more contributions to keep the conversation going, responding to social cues, differences in topic shifting strategies or cues, expectancy about what the conversation partner can understand), we expect differences in shift costs between partners with different roles in the conversation within this corpus (i.e., participants and research investigators).

We observed differences in the research investigator's responses to participant-generated subtopic shifts versus responses continuing on the same subtopic varied according to the participant's group (Table 10). When responding to a TBI group participant, the pattern of shift costs was similar to that observed for the participants. Namely, when the subtopic changed, the investigator's responses had fewer repetitions and retracings and more filled pauses per syllable than when remaining on the same subtopic (Table 10). When the investigator was responding to the NBI group, there was also a pattern of more filled pauses per syllable (though to a lesser extent than when talking to the TBI group), but more repetitions and retracings when the subtopic shifted (Table 10). The model for idea density did not converge.

These results demonstrate some consistency with the shift costs identified across NBI and TBI groups in the main analysis. In particular, it appears that in turns following a partner-initiated subtopic shift, utterance planning, as indexed by filled pause frequency, is more difficult than in turns that remain on the same subtopic. While not apparent in the main analyses, this result is consistent with a cost to semantic complexity at the word- or sentence level when cognitive resources must be devoted to shifting to a new subtopic. This pattern of increased planning difficulty is consistent with the conclusions of the main analysis, demonstrating effects of increased demand on discourse structure processes on lower-level language production.

The differences in the pattern of shift effects in investigators' responses when talking to the two groups occurred only in the frequency of repetitions and retracings. One possibility is that the investigators had a different approach to the two groups, as they were not blinded to TBI status. Decreased repetitions and retracings after subtopic shifts when speaking with the TBI group may reflect more careful planning in an effort to improve communication, while increased repetitions and retracings when speaking with the NBI group may suggest planning difficulty but not masked by attempts to improve clarity. Another possibility is that repetitions and retracings are not particularly sensitive to subtopic shifting and differences reflect random variation. In the main analysis, repetitions and retracings were the only metric that did not show differences in shift cost size by TBI severity.

The lack of effects in total output and syntactic complexity may be a floor effect resulting from the generally low conversational input from the investigator (Main Text, Table 4).

Overall, the results of this exploratory analysis provide some additional evidence of subtopic shift effects in the investigators, suggesting that subtopic shift costs are not solely accounted for by non-topic conversational factors that might have been imposed by the investigators on the participants in the main analysis. However, these results should be interpreted with caution, as the structure of the current corpus sample is not conducive to the analysis of the investigators. This corpus focuses on participant characteristics in order to study effects of TBI on language production, and is not optimized for studying

dyad interactions. A small set of two investigators were used here to create a consistent conversational environment for the participants, while the variability in conversational partner was much higher when considering participants' effect on investigators. The small number of investigators means that individual idiosyncrasies in language production may play a larger role in the outcome measures of interest, while this variability could be accounted for by the large number of participants in the sample. Also, information such as age and education is provided about participants, but not investigators. These are key covariates to consider for speech/language output. Future work should examine more balanced dyad interactions, so that analysis conditions can be equivalent between investigators (or other types of conversation partners) and participants. Future work may also more robustly investigate the contribution of conversation partner status, knowledge, strategies, and other factors on subtopic shift costs by systematically varying these factors, rather than relying on exploratory analysis.

Supplementary Table 10. Model outputs for the effects of participant group and subtopic status on each dependent measure for the investigator's responses. Statistical models were of the form: glmer(Dependent Measure ~ Group*New + offset(log(Total_Utts)) + (1|ID/Group), family="poisson")where *Group* = participant's group (reference level = NBI), *New* = factor representing subtopic status (reference level = no shift), *ID* = participant, and *Total_Utts* = number of utterances.

Dependent Measure	Term	Parameter Estimate	Standard Error	Z/t Statistic	P Value
A) Words per Utterance					
	Intercept	1.67	0.20	8.13	<0.001
	Group	-0.18	0.29	-0.64	0.52
	New	-0.06	0.07	-0.79	0.43
	Group*New	0.10	0.10	0.96	0.34
B) Verbs per Utterance					
	Intercept	-0.08	0.02	-3.50	<0.001

	Group	0.04	0.03	1.51	0.13
	New	-0.10	0.06	-1.74	0.08
	Group*New	0.01	0.08	0.11	0.91
C) Repetitions + Retracings per Syllable					
	Intercept	-4.13	0.19	-21.41	<0.001
	Group	-0.01	0.27	-0.02	0.98
	New	0.18	0.24	0.73	0.47
	Group*New	-1.07	0.42	-2.51	0.01
D) Filled Pauses per Syllable					
	Intercept	-4.33	0.09	-46.40	<0.001
	Group	-0.60	0.13	-4.38	<0.001
	New	-0.04	0.16	-0.24	0.81
	Group*New	0.49	0.24	2.07	0.04

Appendix 2

Models used to analyze data from the Simple Prediction Task, Chapter 3, Study 1

Model examining the effect of verb type and object type on the proportion of looks in the Healthy Control group.

```
m_int_diss_hc.ar1<-bam(elog_PropLooks~

#parametric effects
      IsUnconstrained +
      IsOther +
      IsUnconstrainedIsOther

#smooth effects
      + s(Time, bs="tp", k=20) #reference smooth
      + s(Time, by=IsUnconstrained, bs="tp", k=20) #ordered smooth
      + s(Time, by=IsOther, bs="tp", k=20) #ordered smooth
      + s(Time, by=IsUnconstrainedIsOther, bs="tp", k=20) #ordered smooth
```

```
#random effects
    + s(Time, Session_Name_, bs = "fs", k = 20, m = 1) # Factor smooth for Participant.
    + s(Time, Session_Name_, by=IsUnconstrained, bs = "fs", k = 20, m = 1) # Factor smooth
for slope of constrained by Participant.
    + s(Time, Session_Name_, by=IsOther, bs = "fs", k = 20, m = 1) # Factor smooth for slope
of object type by Participant.
    + s(Time, Session_Name_, by=IsUnconstrainedIsOther, bs = "fs", k = 20, m = 1) # Factor
smooth for slope of interaction by Participant.

    ,data=bysubj_t50_gammswindow_hc, method="fREML",
    , AR.start = start.event
    , rho = -0.35)
```

Model examining the effect of verb type and group on the proportion of looks to the target.

```
m_group_constrained.ar1<-bam(elog_PropLooks~

#parametric effects
    IsUnconstrained +
    IsTBI +
    IsUnconstrainedIsTBI

#smooth effects
    + s(Time, bs="tp", k=20) #reference smooth
    + s(Time, by=IsUnconstrained, bs="tp", k=20) #ordered smooth
    + s(Time, by=IsTBI, bs="tp", k=20) #ordered smooth
    + s(Time, by=IsUnconstrainedIsTBI, bs="tp", k=20) #ordered smooth

#random effects
    + s(Time, Session_Name_, bs = "fs", k = 20, m = 1) # Factor smooth for Participant.
    + s(Time, Session_Name_, by=IsUnconstrained, bs = "fs", k = 20, m = 1) # Factor smooth
for slope of constrained by Participant.

    ,data=bysubj_t50_gammswindow_compare_target, method="fREML"
    , AR.start = start.event
    , rho = -0.02)
```

Models examining the interaction of Age, PTSS, and NSI and information status on looks to the target on constrained versus unconstrained verb trials, TBI group only

```
m_tbi_unconstrained_age.ar1<-bam(elog_PropLooks~

#parametric effects
     IsUnconstrained +

#smooth effects
    + te(Time, Age, bs="tp", k=10) #reference smooth, covariate
    + te(Time, Age,  by=IsUnconstrained, bs="tp", k=10) #ordered smooth
```

```
#random effects
        + s(Age, Session_Name_, bs = "fs", k = 10, m = 1) # Factor smooth for covariates
        + s(Time, Session_Name_, bs = "fs", k = 10, m = 1) # Factor smooth for Participant.
        + s(Time, Session_Name_, by=IsUnconstrained, bs = "fs", k = 10, m = 1) # Factor smooth
for slope of constrained by Participant.

        , AR.start = start.event
        , rho = -0.03
        ,data=bysubj_t50_gammswindow_tbi_target_cov, method="fREML")



m_tbi_unconstrained_ptsd.ar1<-bam(elog_PropLooks~

#parametric effects
       IsUnconstrained +

#smooth effects
        + te(Time, PCL5_Total, bs="tp", k=10) #reference smooth, covariate
        + te(Time, PCL5_Total,  by=IsUnconstrained, bs="tp", k=10) #ordered smooth

#random effects
        + s(PCL5_Total, Session_Name_, bs = "fs", k = 10, m = 1) # Factor smooth for covariates
        + s(Time, Session_Name_, bs = "fs", k = 10, m = 1) # Factor smooth for Participant.
        + s(Time, Session_Name_, by=IsUnconstrained, bs = "fs", k = 10, m = 1) # Factor smooth
for slope of constrained by Participant.
        , AR.start = start.event
        , rho = -0.03
        ,data=bysubj_t50_gammswindow_tbi_target_cov, method="fREML")


m_tbi_unconstrained_nsi.ar1<-bam(elog_PropLooks~

#parametric effects
       IsUnconstrained +

#smooth effects
        + te(Time, NSI_Total, bs="tp", k=10) #reference smooth, covariate
        + te(Time, NSI_Total,  by=IsUnconstrained, bs="tp", k=10) #ordered smooth

#random effects
        + s(NSI_Total, Session_Name_, bs = "fs", k = 10, m = 1) # Factor smooth for covariates
        + s(Time, Session_Name_, bs = "fs", k = 10, m = 1) # Factor smooth for Participant.
        + s(Time, Session_Name_, by=IsUnconstrained, bs = "fs", k = 10, m = 1) # Factor smooth
for slope of constrained by Participant.

        , AR.start = start.event
        , rho = -0.03
        ,data=bysubj_t50_gammswindow_tbi_target_cov, method="fREML")
```

Models used to analyze data from the Discourse Task, Chapter 3, Study 2

Model examining the effect of fluency and information status on looks to the competitor in the Healthy Control Group.

```
m_hc.ar1<-bam(elog_MeanPropCompetitor~

#parametric effects
      IsDisfluent +
      IsNew +
      IsDisfluentIsNew

#smooths
      + s(Time, bs="tp", k=10) #reference smooth
      + s(Time, by=IsDisfluent, bs="tp", k=10) #ordered smooth
      + s(Time, by=IsNew, bs="tp", k=10) #ordered smooth
      + s(Time, by=IsDisfluentIsNew, bs="tp", k=10) #ordered smooth

#random effects
      + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
      + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) #
Factor smooth for slope of disfluency by Participant.
      + s(Time, RECORDING_SESSION_LABEL, by=IsNew, bs = "fs", k = 10, m = 1) # Factor
smooth for slope of info status by Participant.
      + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluentIsNew, bs = "fs", k = 10, m =
1) # Factor smooth for slope of interaction by Participant.

      ,data=bysubj_t50_targetcomp_gammswindow_hc, method="fREML"
      , AR.start = start.event
      , rho = -0.08)
```

Model examining the effect of information status and group on looks to the competitor for fluent trials.

```
m_group_new.ar1<-bam(elog_MeanPropCompetitor~

#parametric effects
      IsTBI +
      IsNew +
      IsNewIsTBI

#smooths
      + s(Time, bs="tp", k=10) #reference smooth
      + s(Time, by=IsTBI, bs="tp", k=10) #ordered smooth
      + s(Time, by=IsNew, bs="tp", k=10) #ordered smooth
      + s(Time, by=IsNewIsTBI, bs="tp", k=10) #ordered smooth
```

```
#random effects
      + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
      + s(Time, RECORDING_SESSION_LABEL, by=IsNew, bs = "fs", k = 10, m = 1) # Factor
smooth for slope of info status by Participant.

      , AR.start = start.event
      , rho = -0.09
      ,data=data_comparegamms_fluent, method="fREML")
```

Models examining the interaction of Age, PTSS, and NSI and information status on looks to the competitor on fluent trials, TBI group only.

```
m_tbi_new_age.ar1<-bam(elog_MeanPropCompetitor~

#parametric effects
      IsNew+

#smooth terms
      + te(Time, Age, bs="tp", k=10) #reference smooth, age covariate
      + te(Time, Age,  by=IsNew, bs="tp", k=10) #ordered smooth

#random effects
       + s(Age, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
covariates
       + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
      + s(Time, RECORDING_SESSION_LABEL, by=IsNew, bs = "fs", k = 10, m = 1) # Factor
smooth for slope of new by Participant.

       , AR.start = start.event
       , rho = -0.15
       ,data=data_comparegamms_cov_tbi_fluent, method="fREML")
m_tbi_new_ptsd.ar1<-bam(elog_MeanPropCompetitor~

#parametric effects
      IsNew+

#smooth effects
      + te(Time, PCL5_Total, bs="tp", k=10) #reference smooth, PTSS covariate
      + te(Time, PCL5_Total,  by=IsNew, bs="tp", k=10) #ordered smooth

#random effects
       + s(PCL5_Total, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor
smooth for covariates
       + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
```

```
        + s(Time, RECORDING_SESSION_LABEL, by=IsNew, bs = "fs", k = 10, m = 1) # Factor
smooth for slope of new by Participant.

        , AR.start = start.event
        , rho = -0.15
        ,data=data_comparegamms_cov_tbi_fluent, method="fREML")


m_tbi_new_nsi.ar1<-bam(elog_MeanPropCompetitor~

#parametric effects
      IsNew+

#smooth effects
      + te(Time, NSI_Total, bs="tp", k=10) #reference smooth, NSI covariate
      + te(Time, NSI_Total,  by=IsNew, bs="tp", k=10) #ordered smooth

#random effects
       + s(NSI_Total, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor
smooth for covariates
       + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
       + s(Time, RECORDING_SESSION_LABEL, by=IsNew, bs = "fs", k = 10, m = 1) # Factor
smooth for slope of new by Participant.

       , AR.start = start.event
        , rho = -0.15
        ,data=data_comparegamms_cov_tbi_fluent, method="fREML")
```

Models examining the interaction of fluency and group on looks to the competitor on new trials (original and releveled models)

```
m_group_fluency.ar1<-bam(elog_MeanPropCompetitor~

#parametric effects
      IsTBI +
      IsDisfluent +
      IsDisfluentIsTBI

#smooth effects
      + s(Time, bs="tp", k=10) #reference smooth
      + s(Time, by=IsTBI, bs="tp", k=10) #ordered smooth
      + s(Time, by=IsDisfluent, bs="tp", k=10) #ordered smooth
      + s(Time, by=IsDisfluentIsTBI, bs="tp", k=10) #ordered smooth

#random effects
```

```
        + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
    Participant.
        + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) #
    Factor smooth for slope of fluency by Participant.

        , AR.start = start.event
        , rho = -0.10
        ,data=data_comparegamms_new, method="fREML")


    m_group_fluency.ar1_relevel<-bam(elog_MeanPropCompetitor~

    #parametric effects
        IsHealthy +
        IsDisfluent +
        IsDisfluentIsHealthy

    #smooth effects
        + s(Time, bs="tp", k=10) #reference smooth
        + s(Time, by=IsHealthy, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsDisfluent, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsDisfluentIsHealthy, bs="tp", k=10) #ordered smooth

    #random effects
        + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
    Participant.
        + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) #
    Factor smooth for slope of fluency by Participant.
        , AR.start = start.event
        , rho = -0.10
        ,data=data_comparegamms_new, method="fREML")
```

Model examining the interaction of fluency and group on looks to the competitor on given trials.

```
m_group_fluencyg<-bam(elog_MeanPropCompetitor~

#parametric effects
    IsTBI +
    IsDisfluent +
    IsDisfluentIsTBI

#smooth effects
    + s(Time, bs="tp", k=10) #reference smooth
    + s(Time, by=IsTBI, bs="tp", k=10) #ordered smooth
    + s(Time, by=IsDisfluent, bs="tp", k=10) #ordered smooth
    + s(Time, by=IsDisfluentIsTBI, bs="tp", k=10) #ordered smooth

 #random effects
```

```
    + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
    + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) # Factor
smooth for slope of fluency by Participant.

   , AR.start = start.event
    , rho = -0.14
    ,data=data_comparegamms_given, method="fREML")
```

Models examining the effect of information status and group on looks to the competitor for disfluent trials.

```
    m_group_newd.ar1<-bam(elog_MeanPropCompetitor~

    #parametric effects
        IsTBI +
        IsNew +
        IsNewIsTBI

    #smooth effects
        + s(Time, bs="tp", k=10) #reference smooth
        + s(Time, by=IsTBI, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsNew, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsNewIsTBI, bs="tp", k=10) #ordered smooth

    #random effects
        + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
    Participant.
        + s(Time, RECORDING_SESSION_LABEL, by=IsNew, bs = "fs", k = 10, m = 1) # Factor
    smooth for slope of info status by Participant.
        , AR.start = start.event
        , rho = -0.18
        ,data=data_comparegamms_disfluent, method="fREML")


    m_group_newd.ar1_relevel<-bam(elog_MeanPropCompetitor~

    #parametric effects
        IsHealthy +
        IsNew +
        IsNewIsHealthy

    #smooth effects
        + s(Time, bs="tp", k=10) #reference smooth
        + s(Time, by=IsHealthy, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsNew, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsNewIsHealthy, bs="tp", k=10) #ordered smooth
```

```
#random effects
    + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
    + s(Time, RECORDING_SESSION_LABEL, by=IsNew, bs = "fs", k = 10, m = 1) # Factor
smooth for slope of info status by Participant.
    , AR.start = start.event
    , rho = -0.18
    ,data=data_comparegamms_disfluent, method="fREML")
```

Model examining the interaction of fluency and group on looks to the other (non-target, non-competitor) object on new trials.

```
m_group_fluencyOtherB.ar1<-bam(elog_MeanPropOtherB~

#parametric effects
    IsTBI +
    IsDisfluent +
    IsDisfluentIsTBI

#smooth efffects
    + s(Time, bs="tp", k=10) #reference smooth
    + s(Time, by=IsTBI, bs="tp", k=10) #ordered smooth
    + s(Time, by=IsDisfluent, bs="tp", k=10) #ordered smooth
    + s(Time, by=IsDisfluentIsTBI, bs="tp", k=10) #ordered smooth

#random effects
    + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
    + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) #
Factor smooth for slope of fluency by Participant.
    , AR.start = start.event
    , rho = -0.20
    ,data=data_comparegamms_new, method="fREML")
```

Models (original and re-leveled) examining looks to competitor on trials when the competitor was expected versus unexpected by group.

```
m_group_expected<-bam(elog_MeanPropCompetitor~

#parametric effects
    IsTBI +
    IsExpected +
    IsExpectedIsTBI

#smooth effects
    + s(Time, bs="tp", k=10) #reference smooth
    + s(Time, by=IsTBI, bs="tp", k=10) #ordered smooth
```

```
        + s(Time, by=IsExpected, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsExpectedIsTBI, bs="tp", k=10) #ordered smooth

    #random effects
        + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
    Participant.
        + s(Time, RECORDING_SESSION_LABEL, by=IsExpected, bs = "fs", k = 10, m = 1) #
    Factor smooth for slopes by Participant.
        , AR.start = start.event
        , rho = -0.09
        ,data=data_comparegamms, method="fREML")


    m_group_unexpected<-bam(elog_MeanPropCompetitor~

    #parametric effects
        IsTBI +
        IsUnexpected +
        IsUnexpectedIsTBI

    #smooth effects
        + s(Time, bs="tp", k=10) #reference smooth
        + s(Time, by=IsTBI, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsUnexpected, bs="tp", k=10) #ordered smooth
        + s(Time, by=IsUnexpectedIsTBI, bs="tp", k=10) #ordered smooth

    #random effects
        + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
    Participant.
        + s(Time, RECORDING_SESSION_LABEL, by=IsUnexpected, bs = "fs", k = 10, m = 1) #
    Factor smooth for slopes by Participant.
        , AR.start = start.event
        , rho = -0.10
        ,data=data_comparegamms, method="fREML")
```

Model for the effect of whether the initial prediction for the competitor was confirmed (expected) or unconfirmed (unexpected) and group on pupil size.

```
    m_pupil.ar1k <- bam(mean_pupil ~

    #parametric effects
            IsExpected +
            IsTBI +
            IsExpectedIsTBI

    # Fixed Smooths
        + s(as.numeric(mean_gaze_x), as.numeric(mean_gaze_y), k = 20)  # Smooth for where
    gaze is
        + s(as.numeric(MeanBaselinePupil), k = 20)  # Smooth for baseline
        + s(TimeRelCriticalOnset, bs="tp" , k = 20) # reference smooth (unexpected, healthy)
```

```
        + s(TimeRelCriticalOnset, by = IsExpected, bs="tp", k = 20) # E-U for healthy
        + s(TimeRelCriticalOnset, by = IsTBI, bs="tp", k = 20) # T-H for unexpected
        + s(TimeRelCriticalOnset, by = IsExpectedIsTBI, bs="tp", k = 20) # interaction

    # Random Smooths
        + s(TimeRelCriticalOnset, SubjectNumber, bs = "fs",m = 1, k = 20)
        + s(TimeRelCriticalOnset, item, bs = "fs",m = 1, k = 20)
        + s(TimeRelCriticalOnset, SubjectNumber, by = IsExpected, bs = "fs",m = 1, k = 20)
        + s(TimeRelCriticalOnset, item, by = IsExpected, bs = "fs",m = 1, k = 20)
        ,
        data = sampledata_downsampled_compare_criticaltimes_inst2_noex_analysiswindow ,
        #family = "scat",
        method = "fREML",
        #discrete = TRUE,
        , AR.start = start.event
        , rho = 0.60,
        nthreads = 4
        )
```

Models examining the interaction of Age, PTSS, and NSI and information status on looks to the competitor on fluent trials, TBI group only.

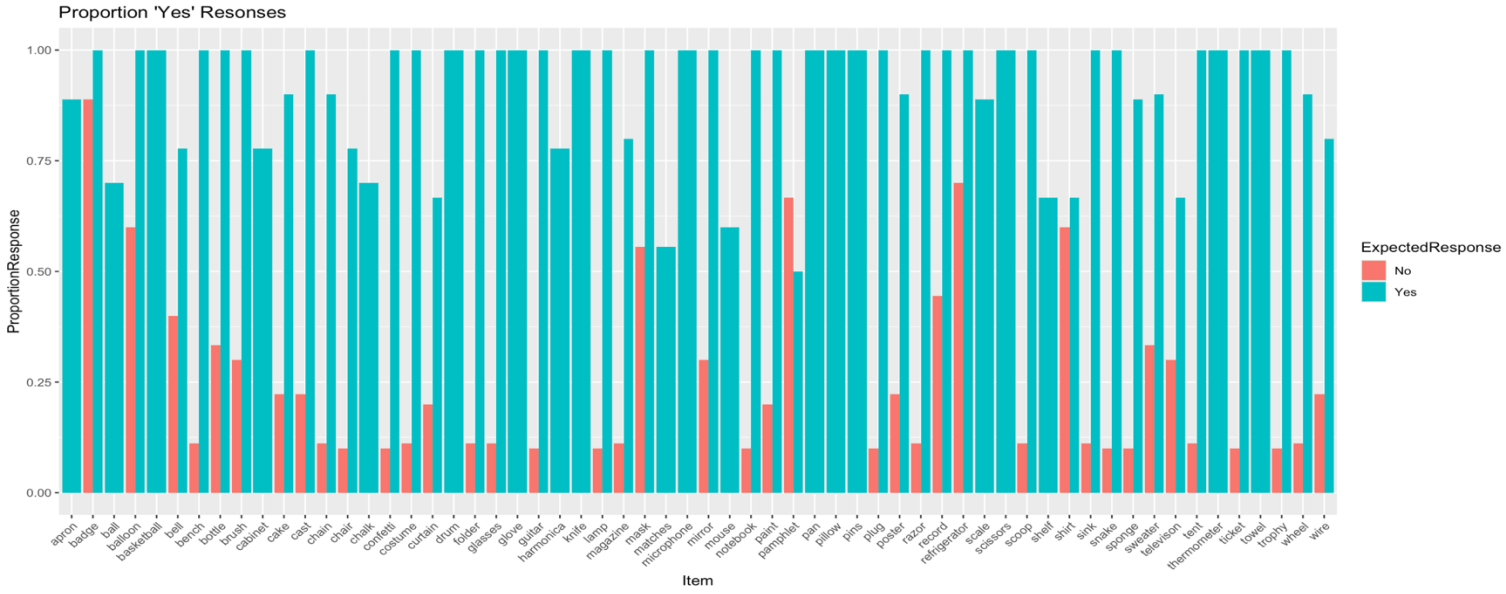
```
    m_tbi_disfluent_age.ar1<-bam(elog_MeanPropCompetitor~

    #parametric effect
        IsDisfluent+

    #smooth effects
        + te(Time, Age, bs="tp", k=10) #reference smooth, age covariate
        + te(Time, Age,  by=IsDisfluent, bs="tp", k=10) #ordered smooth

    #random effects
         + s(Age, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
covariates
        + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
        + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) #
Factor smooth for slope of disfluent by Participant.
        , AR.start = start.event
        , rho = -0.10
        ,data=data_comparegamms_cov_tbi_new, method="fREML")


    m_tbi_disfluent_ptsd.ar1<-bam(elog_MeanPropCompetitor~

    #parametric effect
        IsDisfluent+

    #smooth effects
```

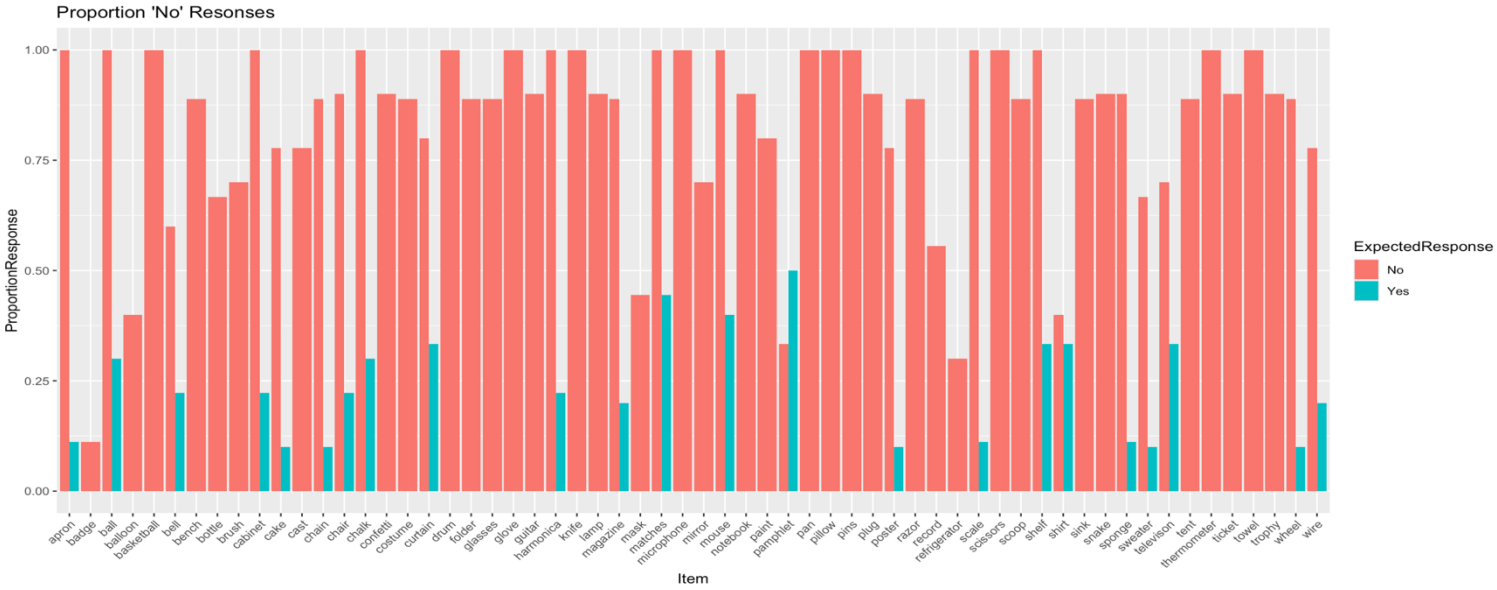
```
        + te(Time, PCL5_Total, bs="tp", k=10) #reference smooth, age covariate
        + te(Time, PCL5_Total,  by=IsDisfluent, bs="tp", k=10) #ordered smooth

#random effects
         + s(PCL5_Total, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor
smooth for covariates
         + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
         + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) #
Factor smooth for slope of disfluent by Participant.
         , AR.start = start.event
         , rho = -0.10
         ,data=data_comparegamms_cov_tbi_new, method="fREML")


m_tbi_disfluent_nsi.ar1<-bam(elog_MeanPropCompetitor~

#parametric effect
        IsDisfluent+

#smooth effects
        + te(Time, NSI_Total, bs="tp", k=10) #reference smooth, age covariate
        + te(Time, NSI_Total,  by=IsDisfluent, bs="tp", k=10) #ordered smooth

#random effects
         + s(NSI_Total, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor
smooth for covariates
         + s(Time, RECORDING_SESSION_LABEL, bs = "fs", k = 10, m = 1) # Factor smooth for
Participant.
         + s(Time, RECORDING_SESSION_LABEL, by=IsDisfluent, bs = "fs", k = 10, m = 1) #
Factor smooth for slope of disfluent by Participant.
         , AR.start = start.event
         , rho = -0.10
         ,data=data_comparegamms_cov_tbi_new, method="fREML")
```

Appendix 3

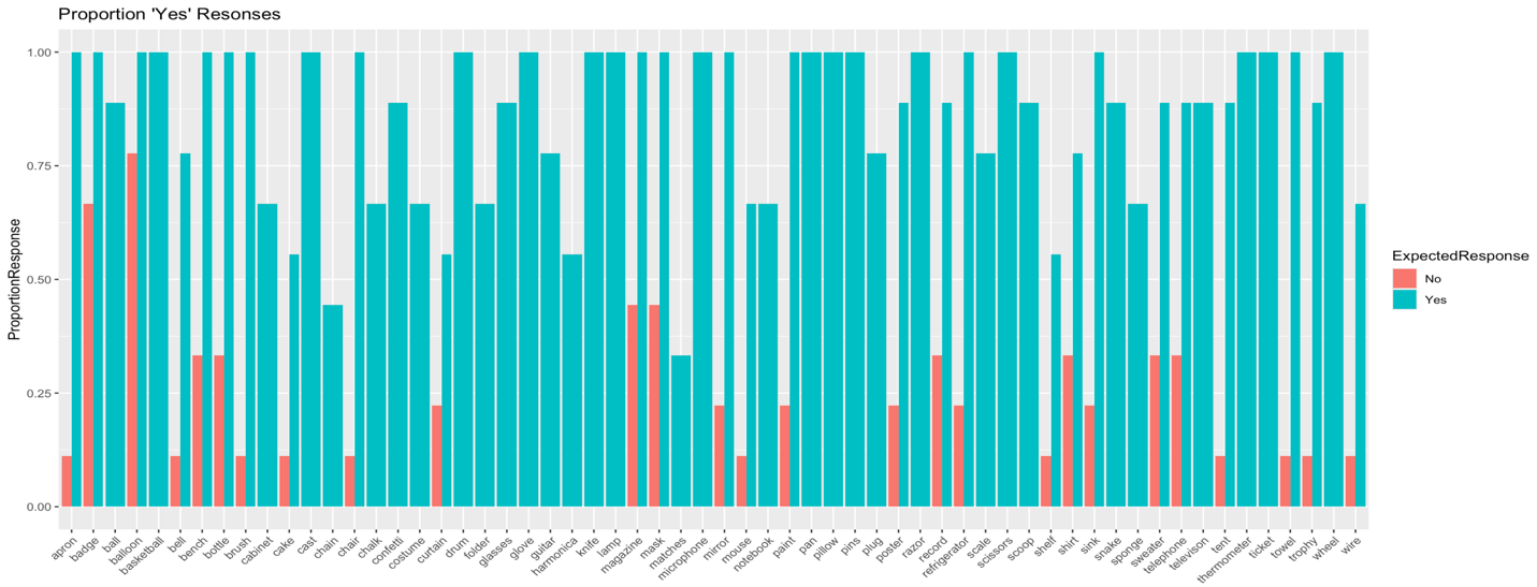
Norming data used to determine topic relevance, Chapter 4

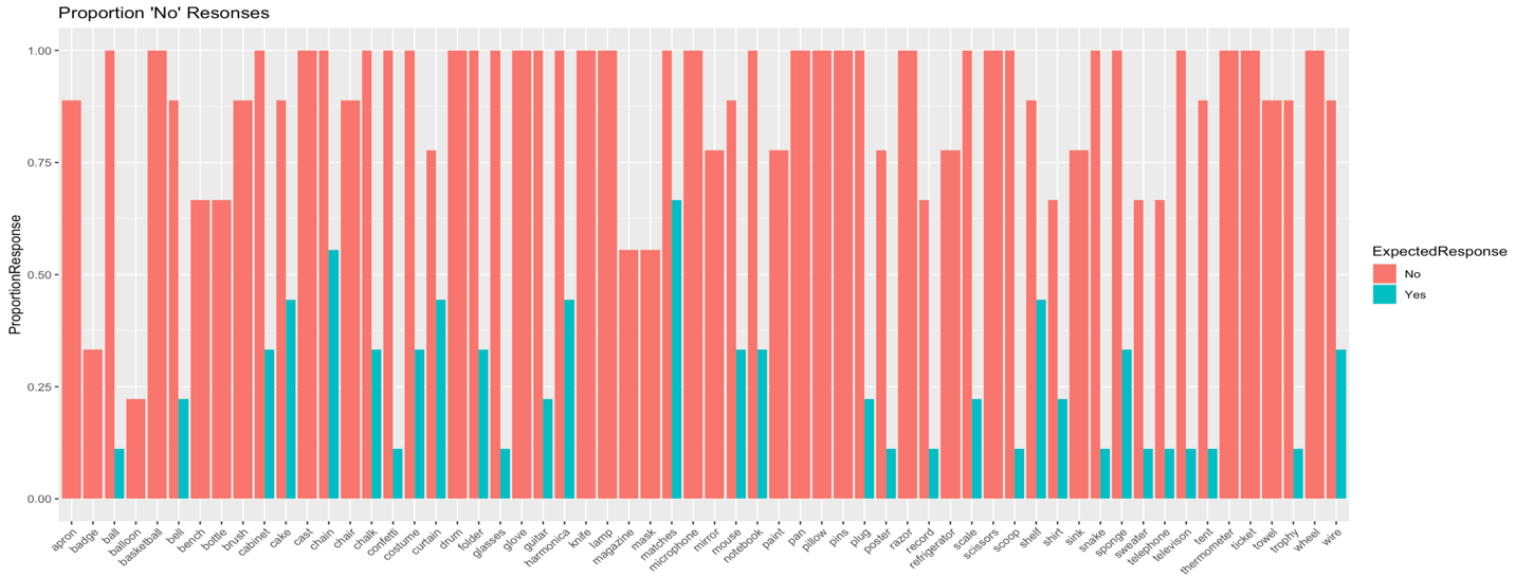
Round 1, N=19





Round 2, N=18


Proportion 'Yes' Resonses
ProportionResponse
1.00
0.75
0.50
0.25
0.00
apron badge ball balloon basketball bell bench bottle brush cabinet cake cast chain chair chalk confetti costume curtain drum folder glasses glove guitar harmonica knife lamp magazine mask matches microphone mirror mouse notebook paint pan pillow pins plug poster razor record refrigerator scale scissors scoop shelf shirt sink snake sponge sweater telephone television tent thermometer ticket towel trophy wheel wire
ExpectedResponse
No
Yes


Proportion 'No' Resonses
ProportionResponse
1.00
0.75
0.50
0.25
0.00
apron badge ball balloon basketball bell bench bottle brush cabinet cake cast chain chair chalk confetti costume curtain drum folder glasses glove guitar harmonica knife lamp magazine mask matches microphone mirror mouse notebook paint pan pillow pins plug poster razor record refrigerator scale scissors scoop shelf shirt sink snake sponge sweater telephone television tent thermometer ticket towel trophy wheel wire
ExpectedResponse
No
Yes

Example Context Script from Chapter 4

Critical trials are shown in bold. Fillers are shown in italics.

Topic	Speaker	Utterance
Radio	1	Okay, so for the radio show sketch, I'm thinking maybe the premise could be like a bird flies into the studio and messes up the radio show
Radio	2	So first we'll need a couple basic things to set up
Radio	2	**We can get the black microphone**
Radio	1	So maybe um the bird comes in and chews something up
Radio	1	**We should use the gold wire**
Radio	1	So the bird chews up the wire and then the microphone doesn't work and nobody listening to the radio can hear their voices
Radio	2	mhmm yeah
Radio	2	That's going wrong and the hosts are panicking
Radio	2	And I think like maybe the bird flies through and knocks over stuff in the studio or something like that
Radio	2	**We can use the blue mouse**
Radio	2	**We can use the striped curtain**
Kitchen	1	mhmm
Kitchen	1	And then that kind of reminds me of something we should do for the high-end kitchen set
Kitchen	1	Maybe like we could also have something come into the kitchen and disrupt something
Kitchen	1	Maybe like a rodent comes in and goes through like the stuff on the counters.
Kitchen	2	mhmm
Kitchen	2	Kind of like Ratatouille
Kitchen	1	mhmm
Kitchen	1	**We can get the yellow gloves**
Kitchen	1	**We should use the red towel**
Kitchen	2	*We could get the pink gloves too*
Kitchen	2	Yeah
Kitchen	2	And like maybe maybe I don't know they find it like if someone opens a door and it pops out or something like that
Kitchen	1	mhmm
Radio	1	And then going back to the radio show maybe the bird can come in and like rip some of the host's things up
Radio	2	mhmm
Radio	2	**We should get the green poster**
Radio	2	And then maybe it gets stuck in like it flies through one of the person's like hair the host's hair and like maybe stays there for a bit
Radio	1	And before it flies out it could hit some of the instruments they have in the studio

Radio	1	*We should use the wooden guitar*
Radio	2	Or um maybe like once it flies away the bird hits a button that plays song through like one of the speakers and it's broadcasting like normal again and the hosts sit down
Radio	1	**We can use the purple record**
Radio	2	mhmm
Radio	1	So then the hosts like sigh and finally relax
Radio	2	*We could get the leather chair*
Kitchen	1	And then back to the kitchen sketch, maybe the rat could somehow get into the stove, light something on fire and maybe that somehow like makes things fall
Kitchen	1	**We should use the orange knife**
Kitchen	1	And then the chef almost cuts his hand
Kitchen	2	mhmm yes
Kitchen	2	Yeah, and then…
Kitchen	2	And speaking of things falling, maybe it the chef knocks things over trying to capture the rat under something but it just makes everything worse
Kitchen	2	**We can use the white pan**
Kitchen	2	And they were trying to bake something, so there's maybe like flour
Kitchen	2	And like the the stuff gets knocked over and falls on the floor and spills everything in the kitchen
Kitchen	2	**We should get the silver scoop**
Kitchen	2	*We could also use the plastic scoop*
Kitchen	2	And, you know, the the flour that got spilled was like the last of their flour so they get it all up from the floor and back into the dish
Kitchen	1	mhmm
Kitchen	1	Okay yeah I think this will be good

Models used to analyze data from Chapter 4

Models examining the effect of condition and object type, releveled so that each object type and condition were used as the reference level

```
m1subj.ar1<-bam(elog_PropLooks~
      IsCompetitor +
      IsWithin +
      IsCross +
      IsWithinIsCompetitor +
      IsCrossIsCompetitor
      + s(TimeRelAdjective, bs="tp", k=25) #reference smooth
      + s(TimeRelAdjective, by=IsCompetitor, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsWithin, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsCross, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsWithinIsCompetitor, bs="tp", k=25) #ordered smooth
```

```
    + s(TimeRelAdjective, by=IsCrossIsCompetitor, bs="tp", k=25) #ordered smooth
   #random effects
    + s(TimeRelAdjective, Session_Name_, bs = "fs", k = 25, m = 1) # Factor smooth for
Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsCompetitor, bs = "fs", k = 25, m = 1) # Factor
smooth for slope of object type by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsWithin, bs = "fs", k = 25, m = 1) # Factor
smooth for slope of condition by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsCross, bs = "fs", k = 25, m = 1) # Factor
smooth for slope of condition by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsWithinIsCompetitor, bs = "fs", k = 25, m = 1)
# Factor smooth for slope of interaction by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsCrossIsCompetitor, bs = "fs", k = 25, m = 1) #
Factor smooth for slope of interaction by Participant.
    , AR.start = start.event
    , rho = -0.02 #increased from 0.01
    ,data=bysubj_t50_compccond_gammswindow_long, method="fREML")


m1subj.ar1_relevel1<-bam(elog_PropLooks~
    IsOther +
    IsWithin +
    IsCross +
    IsWithinIsOther +
    IsCrossIsOther
    + s(TimeRelAdjective, bs="tp", k=25) #reference smooth
    + s(TimeRelAdjective, by=IsOther, bs="tp", k=25) #ordered smooth
    + s(TimeRelAdjective, by=IsWithin, bs="tp", k=25) #ordered smooth
    + s(TimeRelAdjective, by=IsCross, bs="tp", k=25) #ordered smooth
    + s(TimeRelAdjective, by=IsWithinIsOther, bs="tp", k=25) #ordered smooth
    + s(TimeRelAdjective, by=IsCrossIsOther, bs="tp", k=25) #ordered smooth
   #random effects
    + s(TimeRelAdjective, Session_Name_, bs = "fs", k = 25, m = 1) # Factor smooth for
Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsOther, bs = "fs", k = 25, m = 1) # Factor
smooth for slope of object type by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsWithin, bs = "fs", k = 25, m = 1) # Factor
smooth for slope of condition by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsCross, bs = "fs", k = 25, m = 1) # Factor
smooth for slope of condition by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsWithinIsOther, bs = "fs", k = 25, m = 1) #
Factor smooth for slope of interaction by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsCrossIsOther, bs = "fs", k = 25, m = 1) #
Factor smooth for slope of interaction by Participant.
    , AR.start = start.event
    , rho = -0.02
    ,data=bysubj_t50_compccond_gammswindow_long, method="fREML")


      m1subj.ar1_relevel2<-bam(elog_PropLooks~
```

```
      IsOther +
      IsWithin +
      IsNoContext +
      IsWithinIsOther +
      IsNoContextIsOther
      + s(TimeRelAdjective, bs="tp", k=25) #reference smooth
      + s(TimeRelAdjective, by=IsOther, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsWithin, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsNoContext, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsWithinIsOther, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsNoContextIsOther, bs="tp", k=25) #ordered smooth
    #random effects
      + s(TimeRelAdjective, Session_Name_, bs = "fs", k = 25, m = 1) # Factor smooth
for Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsOther, bs = "fs", k = 25, m = 1) #
Factor smooth for slope of object type by Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsWithin, bs = "fs", k = 25, m = 1) #
Factor smooth for slope of condition by Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsNoContext, bs = "fs", k = 25, m = 1)
# Factor smooth for slope of condition by Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsWithinIsOther, bs = "fs", k = 25, m =
1) # Factor smooth for slope of interaction by Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsNoContextIsOther, bs = "fs", k = 25,
m = 1) # Factor smooth for slope of interaction by Participant.
      , AR.start = start.event
      , rho = -0.02
      ,data=bysubj_t50_compccond_gammswindow_long, method="fREML")


m1subj.ar1_relevel3<-bam(elog_PropLooks~
      IsOther +
      IsCross +
      IsNoContext +
      IsCrossIsOther +
      IsNoContextIsOther
      + s(TimeRelAdjective, bs="tp", k=25) #reference smooth
      + s(TimeRelAdjective, by=IsOther, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsCross, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsNoContext, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsCrossIsOther, bs="tp", k=25) #ordered smooth
      + s(TimeRelAdjective, by=IsNoContextIsOther, bs="tp", k=25) #ordered smooth
    #random effects
      + s(TimeRelAdjective, Session_Name_, bs = "fs", k = 25, m = 1) # Factor smooth
for Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsOther, bs = "fs", k = 25, m = 1) #
Factor smooth for slope of object type by Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsCross, bs = "fs", k = 25, m = 1) #
Factor smooth for slope of condition by Participant.
      + s(TimeRelAdjective, Session_Name_, by=IsNoContext, bs = "fs", k = 25, m = 1)
# Factor smooth for slope of condition by Participant.
```

```
    + s(TimeRelAdjective, Session_Name_, by=IsCrossIsOther, bs = "fs", k = 25, m =
1) # Factor smooth for slope of interaction by Participant.
    + s(TimeRelAdjective, Session_Name_, by=IsNoContextIsOther, bs = "fs", k = 25,
m = 1) # Factor smooth for slope of interaction by Participant.
    , AR.start = start.event
    , rho = -0.02
    ,data=bysubj_t50_compccond_gammswindow_long, method="fREML")
```

Bibliography

Adornetti, I. (2014). A Neuro-Cognitive Perspective on the Production and Comprehension of Discourse Coherence. In *Ways to protolanguage 3* (pp. 9–24). Wydawnictwo Wyższej Szkoły Filologicznej : Polska Akademia Nauk. Oddział.

Agimi, Y., Regasa, L. E., & Stout, K. C. (2019). Incidence of Traumatic Brain Injury in the U.S. Military, 2010–2014. *Military Medicine*, *184*(5–6), e233–e241. https://doi.org/10.1093/milmed/usy313

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Anderson, A., Garrod, S. C., & Sanford, A. J. (1983). The Accessibility of Pronominal Antecedents as a Function of Episode Shifts in Narrative Text. *The Quarterly Journal of Experimental Psychology Section A*, *35*(3), 427–440. https://doi.org/10.1080/14640748308402480

Arnold, J. E. (2016). Explicit and Emergent Mechanisms of Information Status. *Topics in Cognitive Science*, *8*(4), 737–760. https://doi.org/10.1111/tops.12220

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–930. https://doi.org/10.1037/0278-7393.33.5.914

Arnold, J. E., Tanenhaus, M. K., Altmann, R. J., & Fagnano, M. (2004). The Old and Thee, uh, New: Disfluency and Reference Resolution. *Psychological Science*, *15*(9), 578–582. https://doi.org/10.1111/j.0956-7976.2004.00723.x

Asher, N., & Vieu, L. (2005). Subordinating and coordinating discourse relations. *Lingua*, *115*(4), 591–610. https://doi.org/10.1016/j.lingua.2003.09.017

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *The Journal of Neuroscience*, *14*(7), 4467–4480. https://doi.org/10.1523/JNEUROSCI.14-07-04467.1994

Audacity Team. (2021). *Audacity(R): Free Audio Editor and Recorder* (Version 3.1.3) [Computer software]. https://audacityteam.org/

Azios, J. H., Archer, B., Simmons-Mackie, N., Raymer, A., Carragher, M., Shashikanth, S., & Gulick, E. (2022). Conversation as an Outcome of Aphasia Treatment: A Systematic Scoping Review. *American Journal of Speech-Language Pathology*, *31*(6), 2920–2942. https://doi.org/10.1044/2022_AJSLP-22-00011

Azouvi, P., Arnould, A., Dromer, E., & Vallat-Azouvi, C. (2017). Neuropsychology of traumatic brain injury: An expert overview. *Revue Neurologique*, *173*(7–8), 461–472. https://doi.org/10.1016/j.neurol.2017.07.006

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Bamdad, M. J., Ryan, L. M., & Warden, D. L. (2003). Functional assessment of executive abilities following traumatic brain injury. *Brain Injury*, *17*(12), 1011–1020. https://doi.org/10.1080/0269905031000110553

Barlow, S. E., Medrano, P., Seichepine, D. R., & Ross, R. S. (2018). Investigation of the changes in oscillatory power during task switching after mild traumatic brain injury. *European Journal of Neuroscience*, *48*(12), 3498–3513. https://doi.org/10.1111/ejn.14231

Beattie, G. W., & Butterworth, B. L. (1979). Contextual Probability and Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech. *Language and Speech*, *22*(3), 201–211. https://doi.org/10.1177/002383097902200301

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for *DSM-5* (PCL-5): Development and Initial Psychometric Evaluation: Posttraumatic Stress Disorder Checklist for *DSM-5*. *Journal of Traumatic Stress*, *28*(6), 489–498. https://doi.org/10.1002/jts.22059

Bolker, B., mixedmodels-misc, (2024), GitHub repository,
https://github.com/bbolker/mixedmodels-misc

Bond, F., & Godfrey, H. P. D. (1997). Conversation with traumatically brain injured individuals: A controlled study of behavioural changes and their impact. *Brain Injury*, *11*(5), 319–330. https://doi.org/10.1080/026990597123476

Bosker, H. R., Van Os, M., Does, R., & Van Bergen, G. (2019). Counting 'uhm's: How tracking the distribution of native and non-native disfluencies influences online language comprehension. *Journal of Memory and Language*, *106*, 189–202. https://doi.org/10.1016/j.jml.2019.02.006

Boudewyn, M. A., & Carter, C. S. (2018). I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia*, *117*, 148–155. https://doi.org/10.1016/j.neuropsychologia.2018.05.024

Boudewyn, M. A., Carter, C. S., & Swaab, T. Y. (2012). Cognitive Control and Discourse Comprehension in Schizophrenia. *Schizophrenia Research and Treatment*, *2012*, 1–7. https://doi.org/10.1155/2012/484502

Boudewyn, M. A., Xu, Y., Rosenfeld, A. R., & Caines, N. P. (2025). Common sources of linguistic conflict engage domain-general conflict control mechanisms during language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/10.3758/s13415-025-01267-3

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177–220. https://doi.org/10.1016/0010-0285(79)90009-4

Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of Topic Models. *Foundations and Trends® in Information Retrieval*, *11*(2–3), 143–296. https://doi.org/10.1561/1500000030

Brennan, S. E., Kuhlen, A. K., & Charoy, J. (2018). Discourse and Dialogue. In J. T. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (1st ed., pp. 1–57). Wiley. https://doi.org/10.1002/9781119170174.epcn305

Brinton, B., & Fujiki, M. (1984). Development of Topic Manipulation Skills in Discourse. *Journal of Speech, Language, and Hearing Research*, *27*(3), 350–358. https://doi.org/10.1044/jshr.2703.350

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, *40*(2), 540–545. https://doi.org/10.3758/BRM.40.2.540

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language and Cognitive Processes*, *27*(1), 62–89. https://doi.org/10.1080/01690965.2010.543363

Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2*(1), 11–33. https://doi.org/10.5087/dad.2011.102

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-Time Investigation of Referential Domains in Unscripted Conversation: A Targeted Language Game Approach. *Cognitive Science*, *32*(4), 643–684. https://doi.org/10.1080/03640210802066816

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as Contexts in Conversation. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 62, pp. 59–99). Elsevier. https://doi.org/10.1016/bs.plm.2014.09.003

Brungart, D. S., Barrett, M. E., Schurman, J., Sheffield, B., Ramos, L., Martorana, R., & Galloza, H. (2019). Relationship Between Subjective Reports of Temporary Threshold Shift and the Prevalence of Hearing Problems in Military Personnel. *Trends in Hearing*, *23*, 2331216519872601. https://doi.org/10.1177/2331216519872601

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

Chang, Y., Chen, H., Barquero, C., Tsai, H. J., Liang, W., Hsu, C., Muggleton, N. G., & Wang, C. (2024). Linking tonic and phasic pupil responses to P300 amplitude in an emotional face-word Stroop task. *Psychophysiology*, *61*(4), e14479. https://doi.org/10.1111/psyp.14479

Cicerone, K. D., & Kalmar, K. (1995). Persistent postconcussion syndrome: The structure of subjective complaints after mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *10*(3), 1–17. https://doi.org/10.1097/00001199-199510030-00002

Cifu, D. X., Wares, J. R., Hoke, K. W., Wetzel, P. A., Gitchel, G., & Carne, W. (2015). Differential Eye Movements in Mild Traumatic Brain Injury Versus Normal Controls. *Journal of Head Trauma Rehabilitation*, *30*(1), 21–28. https://doi.org/10.1097/HTR.0000000000000036

Clark, H. H. (2014). Spontaneous Discourse. In M. A. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production*. Oxford University Press.

Clark, H. H. (2021). Anchoring Utterances. *Topics in Cognitive Science*, *13*(2), 329–350. https://doi.org/10.1111/tops.12496
Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3
Clark, H. H., & Murphy, G. L. (1982). Audience Design in Meaning and Reference. In J.-F. L. Ny & W. Kintsch (Eds.), *Advances in Psychology* (Vol. 9, pp. 287–299). Elsevier. https://doi.org/10.1016/S0166-4115(09)60059-5
Coelho, C. A. (2007). Management of Discourse Deficits following Traumatic Brain Injury: Progress, Caveats, and Needs. *Seminars in Speech and Language*, *28*(2), 122–135. https://doi.org/10.1055/s-2007-970570
Coelho, C. A., Lê, K., Mozeiko, J., Krueger, F., & Grafman, J. (2012). Discourse production following injury to the dorsolateral prefrontal cortex. *Neuropsychologia*, *50*(14), 3564–3572. https://doi.org/10.1016/j.neuropsychologia.2012.09.005
Coelho, C. A., Liles, B. Z., Duffy, R. J., & Clarkson, J. V. (1993). Conversational Patterns of Aphasic, Closed-head-injured, and Normal Speakers. *Clinical Aphasiology*, *21*, 183–192.
Coelho, C. A., Youse, K. M., & Le, K. N. (2002). Conversational discourse in closed-head-injured and non-brain-injured adults. *Aphasiology*, *16*(4–6), 659–672. https://doi.org/10.1080/02687030244000275
Cognitive Science Research Center. (2012). *ANAM4 military: Administration manual*. Cognitive Science Research Center, University of Oklahoma.
Contier, F., Wartenburger, I., Weymar, M., & Rabovsky, M. (2024). Are the P600 and P3 ERP components linked to the task-evoked pupillary response as a correlate of norepinephrine activity? *Psychophysiology*, *61*(7), e14565. https://doi.org/10.1111/psyp.14565
Corrigan, J. D., Yang, J., Singichetti, B., Manchester, K., & Bogner, J. (2018). Lifetime prevalence of traumatic brain injury with loss of consciousness. *Injury Prevention*, *24*(6), 396–404. https://doi.org/10.1136/injuryprev-2017-042371
Cosentino, E., Adornetti, I., & Ferretti, F. (2013). Processing Narrative Coherence: Towards a Top-Down Model of Discourse [Application/pdf]. *OASIcs, Volume 32, CMN 2013*, *32*, 61–75. https://doi.org/10.4230/OASICS.CMN.2013.61
Covington, N. V., & Duff, M. C. (2021). Heterogeneity Is a Hallmark of Traumatic Brain Injury, Not a Limitation: A New Perspective on Study Design in Rehabilitation Research. *American Journal of Speech-Language Pathology*, *30*(2S), 974–985. https://doi.org/10.1044/2020_AJSLP-20-00081
Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers* (Vol. 286). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.286
Dalgaard, P. (2008). *Introductory Statistics with R*. Springer New York. https://doi.org/10.1007/978-0-387-79054-1
Davis, E. E., & Campbell, K. L. (2023). Event boundaries structure the contents of long-term memory in younger and older adults. *Memory*, *31*(1), 47–60. https://doi.org/10.1080/09658211.2022.2122998
Defense Health Agency. (2025). *DOD Numbers for Traumatic Brain Injury Worldwide 2025 Q1-Q2*. Traumatic Brain Injury Center of Excellence. https://www.health.mil/Reference-Center/Reports/2026/01/05/2025-Q1Q2-DOD-Worldwide-Numbers-for-TBI
Diachek, E., Brown-Schmidt, S., & Duff, M. (2024). Attentional Orienting and Disfluency-Related Memory Boost Are Intact in Adults With Moderate-Severe Traumatic Brain

Injury. *Journal of Speech, Language, and Hearing Research: JSLHR*, *67*(6), 1803–1818. https://doi.org/10.1044/2024_JSLHR-23-00385

Diwakar, M., Harrington, D. L., Maruta, J., Ghajar, J., El-Gabalawy, F., Muzzatti, L., Corbetta, M., Huang, M.-X., & Lee, R. R. (2015). Filling in the gaps: Anticipatory control of eye movements in chronic mild traumatic brain injury. *NeuroImage: Clinical*, *8*, 210–223. https://doi.org/10.1016/j.nicl.2015.04.011

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature*, *8*(3), 231–246. https://doi.org/10.1007/BF02912493

Duncan, C. C., Summers, A. C., Perla, E. J., Coburn, K. L., & Mirsky, A. F. (2011). Evaluation of traumatic brain injury: Brain potentials in diagnosis, function, and prognosis. *International Journal of Psychophysiology*, *82*(1), 24–40. https://doi.org/10.1016/j.ijpsycho.2011.02.013

Dunning, D. L., Westgate, B., & Adlam, A.-L. R. (2016). A meta-analysis of working memory impairments in survivors of moderate-to-severe traumatic brain injury. *Neuropsychology*, *30*(7), 811–819. https://doi.org/10.1037/neu0000285

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*(6), 409–436. https://doi.org/10.1007/BF02143160

Elbourn, E., Kenny, B., Power, E., & Togher, L. (2019). Psychosocial Outcomes of Severe Traumatic Brain Injury in Relation to Discourse Recovery: A Longitudinal Study up to 1 Year Post-Injury. *American Journal of Speech-Language Pathology*, *28*(4), 1463–1478. https://doi.org/10.1044/2019_AJSLP-18-0204

Elman, J. A., Panizzon, M. S., Hagler, D. J., Eyler, L. T., Granholm, E. L., Fennema-Notestine, C., Lyons, M. J., McEvoy, L. K., Franz, C. E., Dale, A. M., & Kremen, W. S. (2017). Task-evoked pupil dilation and BOLD variance as indicators of locus coeruleus dysfunction. *Cortex*, *97*, 60–69. https://doi.org/10.1016/j.cortex.2017.09.025

Ettenhofer, M. L., Hershaw, J. N., Engle, J. R., & Hungerford, L. D. (2018). Saccadic impairment in chronic traumatic brain injury: Examining the influence of cognitive load and injury severity. *Brain Injury*, *32*(13–14), 1740–1748. https://doi.org/10.1080/02699052.2018.1511067

Evans, M. J., Clough, S., Duff, M. C., & Brown-Schmidt, S. (2024). Temporal organization of narrative recall is present but attenuated in adults with hippocampal amnesia. *Hippocampus*, *34*(8), 438–451. https://doi.org/10.1002/hipo.23620

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

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006

Fostick, L., Babkoff, H., & Zukerman, G. (2014). Effect of 24 Hours of Sleep Deprivation on Auditory and Linguistic Perception: A Comparison Among Young Controls, Sleep-Deprived Participants, Dyslexic Readers, and Aging Adults. *Journal of Speech, Language, and Hearing Research*, *57*(3), 1078–1088. https://doi.org/10.1044/1092-4388(2013/13-0031)

Fraundorf, S. H. (2022). *psycholing: R Functions for Common Psycholingusitic and Cognitive Designs* (Version 0.5.4) [Computer software]. https://github.com/sfraundorf/psycholing/tree/master

Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods*, *43*(4), 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

Gardner, R. (1987). The Identification and role of topic in spoken interaction. *Semiotica*, *65*(1–2). https://doi.org/10.1515/semi.1987.65.1-2.129

Gernsbacher, M. A. (1991). Cognitive Processes and Mechanisms in Language Comprehension: The Structure Building Framework. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 27, pp. 217–263). Elsevier. https://doi.org/10.1016/S0079-7421(08)60125-5

Gernsbacher, M. A. (1997). Two Decades of Structure Building. *Discourse Processes*, *23*(3), 265–304. https://doi.org/10.1080/01638539709544994

Gernsbacher, M. A., Tallent, K. A., & Bolliger, C. M. (1999). Disordered Discourse in Schizophrenia Described by the Structure Building Framework. *Discourse Studies*, *1*(3), 355–372. https://doi.org/10.1177/1461445699001003004

Ghayoumi, Z., Yadegari, F., Mahmoodi-Bakhtiari, B., Fakharian, E., Rahgozar, M., & Rasouli, M. (2015). Persuasive Discourse Impairments in Traumatic Brain Injury. *Archives of Trauma Research*, *4*(1), 1–7. https://doi.org/10.5812/atr.21473

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gilmore, C. S., Marquardt, C. A., Kang, S. S., & Sponheim, S. R. (2018). Reduced P3b brain response during sustained visual attention is associated with remote blast mTBI and current PTSD in U.S. military veterans. *Behavioural Brain Research*, *340*, 174–182. https://doi.org/10.1016/j.bbr.2016.12.002

Giora, R. (1996). Language comprehension as structure building. *Journal of Pragmatics*, *26*(3), 417–436. https://doi.org/10.1016/0378-2166(95)00071-2

Grayson, L., Brady, M. C., Togher, L., & Ali, M. (2021). The impact of cognitive-communication difficulties following traumatic brain injury on the family; a qualitative, focus group study. *Brain Injury*, *35*(1), 15–25. https://doi.org/10.1080/02699052.2020.1849800

Griffin, Z. M., & Bock, K. (1998). Constraint, Word Frequency, and the Relationship between Lexical Processing Levels in Spoken Word Production. *Journal of Memory and Language*, *38*(3), 313–338. https://doi.org/10.1006/jmla.1997.2547

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). *Centering: A framework for modeling the local coherence of discourse* (Vol. 21). MIT Press.

Haarbauer-Krupa, J., Pugh, M. J., Prager, E. M., Harmon, N., Wolfe, J., & Yaffe, K. (2021). Epidemiology of Chronic Effects of Traumatic Brain Injury. *Journal of Neurotrauma*, *38*(23), 3235–3247. https://doi.org/10.1089/neu.2021.0062

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic Unification. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (4th ed., pp. 819–835). The MIT Press. https://doi.org/10.7551/mitpress/8029.003.0072

Hakulinen, A. (2009). Conversation Types. In S. D'hondt, J.-O. Ostman, & J. Verschueren (Eds.), *The Pragmatics of Interaction* (pp. 55–66). John Benjamins Publishing Company.

Hald, L. A., Steenbeek-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, *1146*, 210–218. https://doi.org/10.1016/j.brainres.2007.02.054

Hard, B. M., Recchia, G., & Tversky, B. (2011a). The shape of action. *Journal of Experimental Psychology: General*, *140*(4), 586–604. https://doi.org/10.1037/a0024310

Hard, B. M., Recchia, G., & Tversky, B. (2011b). The shape of action. *Journal of Experimental Psychology: General*, *140*(4), 586–604. https://doi.org/10.1037/a0024310

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring New information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 512–521. https://doi.org/10.1016/S0022-5371(74)80003-4

Hawkins, P. R. (1971). The Syntactic Location of Hesitation Pauses. *Language and Speech*, *14*(3), 277–288. https://doi.org/10.1177/002383097101400308

Hayes, J. P., Bigler, E. D., & Verfaellie, M. (2016). Traumatic Brain Injury as a Disorder of Brain Connectivity. *Journal of the International Neuropsychological Society*, *22*(2), 120–137. https://doi.org/10.1017/S1355617715000740

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, *14*(10), e0223792. https://doi.org/10.1371/journal.pone.0223792

Heller, D., & Brown-Schmidt, S. (2023). The Multiple Perspectives Theory of Mental States in Communication. *Cognitive Science*, *47*(7), e13322. https://doi.org/10.1111/cogs.13322

Hengst, J. A., & Duff, M. C. (2007). Clinicians as Communication Partners: Developing a Mediated Discourse Elicitation Protocol. *Topics in Language Disorders*, *27*(1), 37–49. https://doi.org/10.1097/00011363-200701000-00005

Hershaw, J. N., & Ettenhofer, M. L. (2018). Insights into cognitive pupillometry: Evaluation of the utility of pupillary metrics for assessing cognitive load in normative and clinical samples. *International Journal of Psychophysiology*, *134*, 62–78. https://doi.org/10.1016/j.ijpsycho.2018.10.008

Hill, E., Claessen, M., Whitworth, A., Boyes, M., & Ward, R. (2018). Discourse and cognition in speakers with acquired brain injury (ABI): A systematic review: Discourse and cognition in speakers with ABI: a systematic review. *International Journal of Language & Communication Disorders*, *53*(4), 689–717. https://doi.org/10.1111/1460-6984.12394

Hobbs, J. R. (1985). *On the coherence and structure of discourse.*

Hobbs, J. R. (1990). Topic drift. *Conversational Organization and Its Development*, *38*, 3–22.

Hoddes, E., Dement, W. C., & Zarcone, V. (2015). *Stanford Sleepiness Scale* [Dataset]. https://doi.org/10.1037/t07116-000

Hoover, P. J., Nix, C. A., Llop, J. Z., Lu, L. H., Bowles, A. O., & Caban, J. J. (2023). Correlations Between the Neurobehavioral Symptom Inventory and Other Commonly Used Questionnaires for Traumatic Brain Injury. *Military Medicine*, *188*(7–8), e2150–e2157. https://doi.org/10.1093/milmed/usab559

Hox, J. J., & Roberts, J. K. (Eds.). (2011). *Handbook of advanced multilevel analysis*. Routledge,Taylor & Francis Group. https://doi.org/10.4324/9780203848852

Hsu, N. S., Jaeggi, S. M., & Novick, J. M. (2017). A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. *Brain and Language*, *166*, 63–77. https://doi.org/10.1016/j.bandl.2016.12.006

Hsu, N. S., Kuchinsky, S. E., & Novick, J. M. (2021). Direct impact of cognitive control on sentence processing and comprehension. *Language, Cognition and Neuroscience*, *36*(2), 211–239. https://doi.org/10.1080/23273798.2020.1836379

Hsu, N. S., & Novick, J. M. (2016). Dynamic Engagement of Cognitive Control Modulates Recovery From Misinterpretation During Real-Time Language Processing. *Psychological Science*, *27*(4), 572–582. https://doi.org/10.1177/0956797615625223

Jannace, K. C., Pompeii, L., Gimeno Ruiz De Porras, D., Perkison, W. B., Yamal, J.-M., Trone, D. W., & Rull, R. P. (2024). Lifetime Traumatic Brain Injury and Risk of Post-Concussive Symptoms in the Millennium Cohort Study. *Journal of Neurotrauma*, *41*(5–6), 613–622. https://doi.org/10.1089/neu.2022.0213

Ji, Y., & Papafragou, A. (2022). Boundedness in event cognition: Viewers spontaneously represent the temporal texture of events. *Journal of Memory and Language*, *127*, 104353. https://doi.org/10.1016/j.jml.2022.104353

Johansson, B., Berglund, P., & Rönnbäck, L. (2009). Mental fatigue and impaired information processing after mild and moderate traumatic brain injury. *Brain Injury*, *23*(13–14), 1027–1040. https://doi.org/10.3109/02699050903421099

Johns, M. A., Calloway, R. C., Karunathilake, I. M. D., Decruy, L. P., Anderson, S., Simon, J. Z., & Kuchinsky, S. E. (2024). Attention Mobilization as a Modulator of Listening Effort: Evidence From Pupillometry. *Trends in Hearing*, *28*, 23312165241245240. https://doi.org/10.1177/23312165241245240

Johnson, J. E., & Turkstra, L. S. (2012). Inference in conversation of adults with traumatic brain injury. *Brain Injury*, *26*(9), 1118–1126. https://doi.org/10.3109/02699052.2012.666370

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kavé, G., Heled, E., Vakil, E., & Agranov, E. (2011a). Which verbal fluency measure is most useful in demonstrating executive deficits after traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 358–365. https://doi.org/10.1080/13803395.2010.518703

Kavé, G., Heled, E., Vakil, E., & Agranov, E. (2011b). Which verbal fluency measure is most useful in demonstrating executive deficits after traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology*, *33*(3), 358–365. https://doi.org/10.1080/13803395.2010.518703

Kekes-Szabo, S., Clough, S., Brown-Schmidt, S., & Duff, M. C. (2025). Multiparty Communication: A New Direction in Characterizing the Impact of Traumatic Brain Injury on Social Communication. *American Journal of Speech-Language Pathology*, 1–14. https://doi.org/10.1044/2025_AJSLP-24-00151

Key-DeLyria, S. E. (2016). Sentence Processing in Traumatic Brain Injury: Evidence From the P600. *Journal of Speech, Language, and Hearing Research*, *59*(4), 759–771. https://doi.org/10.1044/2016_JSLHR-L-15-0104

Kim, A. E., Langlois, V. J., Ness, T., Wade, M., & Novick, J. M. (2025). Resolving conflicting interpretations: Theta band oscillations and the role of cognitive control. *Neuropsychologia*, *217*, 109214. https://doi.org/10.1016/j.neuropsychologia.2025.109214

King, J. H., Sweet, J. J., Sherer, M., Curtiss, G., & Vanderploeg, R. D. (2002). Validity Indicators Within the Wisconsin Card Sorting Test: Application of New and Previously

Researched Multivariate Procedures in Multiple Traumatic Brain Injury Samples. *The Clinical Neuropsychologist*, *16*(4), 506–523. https://doi.org/10.1076/clin.16.4.506.13912
Kintsch, W. (1974). *The representation of meaning in memory* (Vol. 15). Erlbaum.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163
Kowalski, R. G., Hammond, F. M., Weintraub, A. H., Nakase-Richardson, R., Zafonte, R. D., Whyte, J., & Giacino, J. T. (2021). Recovery of Consciousness and Functional Outcome in Moderate and Severe Traumatic Brain Injury. *JAMA Neurology*, *78*(5), 548–557. https://doi.org/10.1001/jamaneurol.2021.0084
Krautz, A. E., & Keuleers, E. (2022). LinguaPix database: A megastudy of picture-naming norms. *Behavior Research Methods*, *54*(2), 941–954. https://doi.org/10.3758/s13428-021-01651-0
Kroczek, L. O. H., & Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Scientific Reports*, *7*(1), 17581. https://doi.org/10.1038/s41598-017-17907-9
Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368. https://doi.org/10.1037/0033-295X.89.4.334
Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Brown University Press.
Kukona, A., Cho, P. W., Magnuson, J. S., & Tabor, W. (2014). Lexical interference effects in sentence processing: Evidence from the visual world paradigm and self-organizing models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 326–347. https://doi.org/10.1037/a0034903
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299
Lange, R. T., French, L. M., Lippa, S. M., Bailie, J. M., & Brickell, T. A. (2020). Posttraumatic Stress Disorder is a Stronger Predictor of Long-Term Neurobehavioral Outcomes Than Traumatic Brain Injury Severity. *Journal of Traumatic Stress*, *33*(3), 318–329. https://doi.org/10.1002/jts.22480
Lange, R. T., Lippa, S. M., Brickell, T. A., Yeh, P.-H., Ollinger, J., Wright, M., Driscoll, A., Sullivan, J., Braatz, S., Gartner, R., Barnhart, E., & French, L. M. (2021). Post-Traumatic Stress Disorder Is Associated with Neuropsychological Outcome but Not White Matter Integrity after Mild Traumatic Brain Injury. *Journal of Neurotrauma*, *38*(1), 63–73. https://doi.org/10.1089/neu.2019.6852
Langlois, V. J., Ness, T., Kim, A. E., & Novick, J. M. (2024). Does cognitive control modulate referential ambiguity resolution? A remote visual world study. *Language, Cognition and Neuroscience*, *39*(7), 924–938. https://doi.org/10.1080/23273798.2024.2369182
Lê, K., & Coelho, C. A. (2023). Discourse Characteristics in Traumatic Brain Injury. In A. P.-H. Kong (Ed.), *Spoken Discourse Impairments in the Neurogenic Populations* (pp. 65–80). Springer International Publishing. https://doi.org/10.1007/978-3-031-45190-4_5
Lê, K., Coelho, C. A., & Fiszdon, J. (2022). Systematic Review of Discourse and Social Communication Interventions in Traumatic Brain Injury. *American Journal of Speech-Language Pathology*, *31*(2), 991–1022. https://doi.org/10.1044/2021_AJSLP-21-00088

Leaman, M. C., & Edmonds, L. A. (2020). “By the Way”… How People With Aphasia and Their Communication Partners Initiate New Topics of Conversation. *American Journal of Speech-Language Pathology*, *29*(1S), 375–392. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0198

Leaman, M. C., & Edmonds, L. A. (2021). Measuring Global Coherence in People With Aphasia During Unstructured Conversation. *American Journal of Speech-Language Pathology*, *30*(1S), 359–375. https://doi.org/10.1044/2020_AJSLP-19-00104

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levin, H. S., Shum, D., & Chan, R. C. K. (Eds.). (2014). *Understanding traumatic brain injury: Current research and future directions*. Oxford University Press.

Levin, H. S., Williams, D., Crofford, M. J., High, W. M., Eisenberg, H. M., Amparo, E. G., Guinto, F. C., Kalisky, Z., Handel, S. F., & Goldman, A. M. (1988). Relationship of depth of brain lesions to consciousness and outcome after closed head injury. *Journal of Neurosurgery*, *69*(6), 861–866. https://doi.org/10.3171/jns.1988.69.6.0861

Lezak, M. D., & Lezak, M. D. (Eds.). (2004). *Neuropsychological assessment* (4th ed). Oxford University Press.

Li, H., Li, N., Xing, Y., Zhang, S., Liu, C., Cai, W., Hong, W., & Zhang, Q. (2021). P300 as a Potential Indicator in the Evaluation of Neurocognitive Disorders After Traumatic Brain Injury. *Frontiers in Neurology*, *12*, 690792. https://doi.org/10.3389/fneur.2021.690792

Lippa, Sara. M., French, L. M., Brickell, T. A., Driscoll, A. E., Glazer, M. E., Tippett, C. E., Sullivan, J. K., & Lange, R. T. (2021a). Post-Traumatic Stress Disorder Symptoms Are Related to Cognition after Complicated Mild and Moderate Traumatic Brain Injury but Not Severe and Penetrating Traumatic Brain Injury. *Journal of Neurotrauma*, *38*(22), 3137–3145. https://doi.org/10.1089/neu.2021.0120

Lippa, Sara. M., French, L. M., Brickell, T. A., Driscoll, A. E., Glazer, M. E., Tippett, C. E., Sullivan, J. K., & Lange, R. T. (2021b). Post-Traumatic Stress Disorder Symptoms Are Related to Cognition after Complicated Mild and Moderate Traumatic Brain Injury but Not Severe and Penetrating Traumatic Brain Injury. *Journal of Neurotrauma*, *38*(22), 3137–3145. https://doi.org/10.1089/neu.2021.0120

Lorch, R. F., Lorch, E. P., & Matthews, P. D. (1985). On-line processing of the topic structure of a text. *Journal of Memory and Language*, *24*(3), 350–362. https://doi.org/10.1016/0749-596X(85)90033-6

Lord, K. M., Duff, M. C., & Brown-Schmidt, S. (2024). Temporary ambiguity and memory for the context of spoken language in adults with moderate-severe traumatic brain injury. *Brain and Language*, *257*, 105471. https://doi.org/10.1016/j.bandl.2024.105471

Lord, K. M., Kekes-Szabo, S., Sherlock, P., Duff, M. C., & Brown-Schmidt, S. (2026). Moderate-severe traumatic brain injury disrupts core mechanisms of online language processing and use. *Cortex*, S0010945226000389. https://doi.org/10.1016/j.cortex.2026.01.010

Lucca, L. F., Lofaro, D., Pignolo, L., Leto, E., Ursino, M., Cortese, M. D., Conforti, D., Tonin, P., & Cerasa, A. (2019). Outcome prediction in disorders of consciousness: The role of coma recovery scale revised. *BMC Neurology*, *19*(1), 1–8. https://doi.org/10.1186/s12883-019-1293-7

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

MacDonald, S. (2017). Introducing the model of cognitive-communication competence: A model to guide evidence-based communication interventions after brain injury. *Brain Injury*, *31*(13–14), 1760–1780. https://doi.org/10.1080/02699052.2017.1379613

Macdonald, S., & Johnson, C. J. (2005). Assessment of subtle cognitive-communication deficits following acquired brain injury: A normative study of the Functional Assessment of Verbal Reasoning and Executive Strategies (FAVRES). *Brain Injury*, *19*(11), 895–902. https://doi.org/10.1080/02699050400004294

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates

Malone, C., Erler, K. S., Giacino, J. T., Hammond, F. M., Juengst, S. B., Locascio, J. J., Nakase-Richardson, R., Verduzco-Gutierrez, M., Whyte, J., Zasler, N., & Bodien, Y. G. (2019). Participation Following Inpatient Rehabilitation for Traumatic Disorders of Consciousness: A TBI Model Systems Study. *Frontiers in Neurology*, *10*, 1314. https://doi.org/10.3389/fneur.2019.01314

Mandler, J. M., & Goodman, M. S. (1982). On the psychological validity of story structure. *Journal of Verbal Learning and Verbal Behavior*, *21*(5), 507–523. https://doi.org/10.1016/S0022-5371(82)90746-0

Marino, A. N., Mui, M., Dobryakova, E., & Sandry, J. (2025). Classification Systems for Moderate-Severe Traumatic Brain Injury: Review and Recommendations. *Current Physical Medicine and Rehabilitation Reports*, *13*(1), 32. https://doi.org/10.1007/s40141-025-00504-7

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Martelli, M. F., Grayson, R. L., & Zasler, N. D. (1999). Posttraumatic Headache: Neuropsychological and Psychological Effects and Treatment Implications. *Journal of Head Trauma Rehabilitation*, *14*(1), 49–69.

Mason, R. A., & Just, M. A. (2006). Neuroimaging Contributions to the Understanding of Discourse Processes. In *Handbook of Psycholinguistics* (pp. 765–799). Elsevier. https://doi.org/10.1016/B978-012369374-7/50020-1

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

McMurray, B., Farris-Trimble, A., & Rigler, H. (2017). Waiting for lexical access: Cochlear implants or severely degraded input lead listeners to process speech less incrementally. *Cognition*, *169*, 147–164. https://doi.org/10.1016/j.cognition.2017.08.013

McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In R. P. G. Van Gompel (Ed.), *Sentence processing*. Psychology Press. https://doi.org/10.4324/9780203488454

Mentis, M., & Prutting, C. A. (1991). Analysis of Topic as Illustrated in a Head-Injured and a Normal Adult. *Journal of Speech, Language, and Hearing Research*, *34*(3), 583–595. https://doi.org/10.1044/jshr.3403.583

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567. https://doi.org/10.1016/j.jml.2012.01.001

Meyer, A. S. (2023). Timing in Conversation. *Journal of Cognition*, *6*(1), 1–17. https://doi.org/10.5334/joc.268

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed Representations of Words and Phrases and their Compositionality* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1310.4546

Millis, S. R., Rosenthal, M., Novack, T. A., Sherer, M., Nick, T. G., Kreutzer, J. S., High, W. M., & Ricker, J. H. (2001). Long-Term Neuropsychological Outcome After Traumatic Brain Injury: *Journal of Head Trauma Rehabilitation*, *16*(4), 343–355. https://doi.org/10.1097/00001199-200108000-00005

Morrow, E. L., Brown-Schmidt, S., & Duff, M. C. (2024). Memory for Conversation in Traumatic Brain Injury: A Feasibility Study and Preliminary Findings. *Journal of Speech, Language, and Hearing Research*, 1–10. https://doi.org/10.1044/2024_JSLHR-23-00420

Mozeiko, J., Le, K., Coelho, C. A., Krueger, F., & Grafman, J. (2011). The relationship of story grammar and executive function following TBI. *Aphasiology*, *25*(6–7), 826–835. https://doi.org/10.1080/02687038.2010.543983

Müller, A., Candrian, G., Dall'Acqua, P., Kompatsiari, K., Baschera, G.-M., Mica, L., Simmen, H.-P., Glaab, R., Fandino, J., Schwendinger, M., Meier, C., Ulbrich, E. J., & Johannes, S. (2015). Altered cognitive processes in the acute phase of mTBI: An analysis of independent components of event-related potentials. *NeuroReport*, *26*(16), 952–957. https://doi.org/10.1097/WNR.0000000000000447

Ness, T., Langlois, V. J., Novick, J. M., & Kim, A. E. (2024). Theta-band neural oscillations reflect cognitive control during language processing. *Journal of Experimental Psychology: General*, *153*(9), 2279–2298. https://doi.org/10.1037/xge0001621

Newtson, D., Engquist, G. A., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, *35*(12), 847–862. https://doi.org/10.1037/0022-3514.35.12.847

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When Peanuts Fall in Love: N400 Evidence for the Power of Discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098

Nolden, S., Turan, G., Güler, B., & Günseli, E. (2024). Prediction error and event segmentation in episodic memory. *Neuroscience & Biobehavioral Reviews*, *157*, 105533. https://doi.org/10.1016/j.neubiorev.2024.105533

Norman, R. S., Mueller, K. D., Huerta, P., Shah, M. N., Turkstra, L. S., & Power, E. (2022). Discourse Performance in Adults With Mild Traumatic Brain Injury, Orthopedic Injuries, and Moderate to Severe Traumatic Brain Injury, and Healthy Controls. *American Journal of Speech-Language Pathology*, *31*(1), 67–83. https://doi.org/10.1044/2021_AJSLP-20-00299

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, *26*(6), 527–567. https://doi.org/10.1080/02643290903519367

Nozari, N., Trueswell, J. C., & Thompson-Schill, S. L. (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychonomic Bulletin & Review*, *23*(6), 1942–1953. https://doi.org/10.3758/s13423-016-1068-8

Ord, J. S., Greve, K. W., Bianchini, K. J., & Aguerrevere, L. E. (2010). Executive dysfunction in traumatic brain injury: The effects of injury severity and effort on the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *32*(2), 132–140. https://doi.org/10.1080/13803390902858874

Ovans, Z. (2022). *Developmental Parsing and Cognitive Control* [Unpublished manuscript].

Ovans, Z., Hsu, N. S., Bell-Souder, D., Gilley, P., Novick, J. M., & Kim, A. E. (2022). Cognitive control states influence real-time sentence processing as reflected in the P600 ERP. *Language, Cognition and Neuroscience*, 1–9. https://doi.org/10.1080/23273798.2022.2026422

Peach, R. K. (2013). The Cognitive Basis for Sentence Planning Difficulties in Discourse After Traumatic Brain Injury. *American Journal of Speech-Language Pathology*, *22*(2), S285–S297. https://doi.org/10.1044/1058-0360(2013/12-0081)

Peach, R. K., & Coelho, C. A. (2016). Linking inter- and intra-sentential processes for narrative production following traumatic brain injury: Implications for a model of discourse processing. *Neuropsychologia*, *80*, 157–164. https://doi.org/10.1016/j.neuropsychologia.2015.11.015

Peach, R. K., & Hanna, L. E. (2021). Sentence-level processing predicts narrative coherence following traumatic brain injury: Evidence in support of a resource model of discourse processing. *Language, Cognition and Neuroscience*, *36*(6), 694–710. https://doi.org/10.1080/23273798.2021.1894346

Perianez, J., Rioslago, M., Rodriguezsanchez, J., Adroverroig, D., Sanchezcubillo, I., Crespofacorro, B., Quemada, J., & Barcelo, F. (2007). Trail Making Test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology*, *22*(4), 433–447. https://doi.org/10.1016/j.acn.2007.01.022

Phillips, I., Ellis, G. M., McNamara, B., Lefler, J., DeRoy Milvae, K., Gordon-Salant, S., Kuchinsky, S. E., & Brungart, D. S. (2023). Examining the feasibility of integrating pupillometry measures of listening effort into clinical audiology assessments. *The Journal of the Acoustical Society of America*, *153*(3_supplement), A79–A79. https://doi.org/10.1121/10.0018230

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear & Hearing*, *37*(1), 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pilcher, J. J., McClelland, L. E., Moore, D. D., Haarmann, H., Baron, J., Wallsten, T. S., & McCubbin, J. A. (2007). Language performance under sustained work and sleep deprivation conditions. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl), B25-38.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Ponsford, J., Velikonja, D., Janzen, S., Harnett, A., McIntyre, A., Wiseman-Hakes, C., Togher, L., Teasell, R., Kua, A., Patsakos, E., Welch-West, P., & Bayley, M. T. (2023). INCOG 2.0 Guidelines for Cognitive Rehabilitation Following Traumatic Brain Injury, Part II:

Attention and Information Processing Speed. *Journal of Head Trauma Rehabilitation*, *38*(1), 38–51. https://doi.org/10.1097/HTR.0000000000000839
Porretta, V., Kyröläinen, A.-J., Van Rij, J., & Järvikivi, J. (2018). Visual World Paradigm Data: From Preprocessing to Nonlinear Time-Course Analysis. In I. Czarnowski, R. J. Howlett, & L. C. Jain (Eds.), *Intelligent Decision Technologies 2017* (Vol. 73, pp. 268–277). Springer International Publishing. https://doi.org/10.1007/978-3-319-59424-8_25
Postma, A., Kolk, H., & Povel, D.-J. (1990). On The Relation among Speech Errors, Disfluencies, and Self-Repairs. *Language and Speech*, *33*(1), 19–29. https://doi.org/10.1177/002383099003300102
R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Riou, M. (2017). Transitioning to a new topic in American English conversation: A multi-level and mixed-methods account. *Journal of Pragmatics*, *117*, 88–105. https://doi.org/10.1016/j.pragma.2017.06.015
Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 1–69. https://doi.org/10.3765/sp.5.6
Rossion, B., & Pourtois, G. (2001). Revisiting snodgrass and Vanderwart's object database: Color and texture improve object recognition. *Journal of Vision*, *1*(3), 413–413. https://doi.org/10.1167/1.3.413
Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition*, *181*, 141–150. https://doi.org/10.1016/j.cognition.2018.08.018
Salas, C. E., Casassus, M., Rowlands, L., Pimm, S., & Flanagan, D. A. J. (2018). "Relating through sameness": A qualitative study of friendship and social isolation in chronic traumatic brain injury. *Neuropsychological Rehabilitation*, *28*(7), 1161–1178. https://doi.org/10.1080/09602011.2016.1247730
Sankaran, N., Carlson, T. A., & Thompson, W. F. (2020). The Rapid Emergence of Musical Pitch Structure in Human Cortex. *The Journal of Neuroscience*, *40*(10), 2108–2118. https://doi.org/10.1523/JNEUROSCI.1399-19.2020
Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods*, *51*(3), 1258–1270. https://doi.org/10.3758/s13428-018-1099-3
Sedivy, J. C., K. Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6
Selassie, A. W., Zaloshnja, E., Langlois, J. A., Miller, T., Jones, P., & Steiner, C. (2008). Incidence of Long-term Disability Following Traumatic Brain Injury Hospitalization, United States, 2003. *Journal of Head Trauma Rehabilitation*, *23*(2), 123–131. https://doi.org/10.1097/01.HTR.0000314531.30401.39
Shahid, A., Shen, J., & Shapiro, C. M. (2010). Measurements of sleepiness and fatigue. *Journal of Psychosomatic Research*, *69*(1), 81–89. https://doi.org/10.1016/j.jpsychores.2010.04.001
Sharer, K., & Thothathiri, M. (2020). Neural Mechanisms Underlying the Dynamic Updating of Native Language. *Neurobiology of Language*, *1*(4), 492–522. https://doi.org/10.1162/nol_a_00023

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1503–1514. https://doi.org/10.1037/a0024350

Slevc, L. R., & Martin, R. C. (2016). Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology*, *28*(7), 773–790. https://doi.org/10.1080/20445911.2016.1202252

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. https://doi.org/10.1016/j.cogpsych.2004.03.001

Snow, P., Douglas, J., & Ponsford, J. (1997). Conversational assessment following traumatic brain injury: A comparison across two control groups. *Brain Injury*, *11*(6), 409–429. https://doi.org/10.1080/026990597123403

Snowdon, D. A. (1996). Linguistic Ability in Early Life and Cognitive Function and Alzheimer's Disease in Late Life: Findings From the Nun Study. *JAMA*, *275*(7), 528–532. https://doi.org/10.1001/jama.1996.03530310034029

Sokal, R. R., & Rohlf, F. J. (2012). *Biometry: The principles and practice of statistics in biological research* (4th ed). W. H. Freeman.

Solbakk, A.-K., Reinvang, I., & Andersson, S. (2002). Assessment of P3a and P3b after Moderate to Severe Brain Injury. *Clinical Electroencephalography*, *33*(3), 102–110. https://doi.org/10.1177/155005940203300306

Sorg, S. F., Werhane, M. L., Merritt, V. C., Clark, A. L., Holiday, K. A., Hanson, K. L., Jak, A. J., Schiehser, D. M., & Delano-Wood, L. (2021). Research Letter: PTSD Symptom Severity and Multiple Traumatic Brain Injuries Are Associated With Elevated Memory Complaints in Veterans With Histories of Mild TBI. *Journal of Head Trauma Rehabilitation*, *36*(6), 418–423. https://doi.org/10.1097/HTR.0000000000000659

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227–267. https://doi.org/10.1016/0010-0277(94)00647-4

SR Research Ltd. (2024). *Data Viewer* (Version 4.4.1) [Computer software]. https://www.sr-research.com/data-viewer/

Stubbs, E., Togher, L., Kenny, B., Fromm, D., Forbes, M., MacWhinney, B., McDonald, S., Tate, R., Turkstra, L., & Power, E. (2018). Procedural discourse performance in adults with severe traumatic brain injury at 3 and 6 months post injury. *Brain Injury*, *32*(2), 167–181. https://doi.org/10.1080/02699052.2017.1291989

Stuss, D. T. (2011). Traumatic brain injury: Relation to executive dysfunction and the frontal lobes. *Current Opinion in Neurology*, *24*(6), 584–589. https://doi.org/10.1097/WCO.0b013e32834c7eb9

Su, X., & Swallow, K. M. (2024). People can reliably detect action changes and goal changes during naturalistic perception. *Memory & Cognition*, *52*(5), 1093–1111. https://doi.org/10.3758/s13421-024-01525-8

Sullivan, K. W., Law, W. A., Loyola, L., Knoll, M. A., Shub, D. E., & French, L. M. (2020). A Novel Intervention Platform for Service Members With Subjective Cognitive Complaints: Implementation, Patient Participation, and Satisfaction. *Military Medicine*, *185*(Supplement_1), 326–333. https://doi.org/10.1093/milmed/usz218

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, *30*(4), 485–496. https://doi.org/10.1016/S0378-2166(98)00014-9

Taboada, M., & Zabala, L. H. (2008). Deciding on units of analysis within Centering Theory. *Corpus Linguistics and Linguistic Theory*, *4*(1). https://doi.org/10.1515/CLLT.2008.003

Tagliamonte, S. A. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Wiley-Blackwell.

Taing, A. S., Mundy, M. E., Ponsford, J. L., & Spitz, G. (2021). Aberrant modulation of brain activity underlies impaired working memory following traumatic brain injury. *NeuroImage: Clinical*, *31*, 102777. https://doi.org/10.1016/j.nicl.2021.102777

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tanev, K. S., Pentel, K. Z., Kredlow, M. A., & Charney, M. E. (2014). PTSD and TBI co-morbidity: Scope, clinical presentation and treatment options. *Brain Injury*, *28*(3), 261–270. https://doi.org/10.3109/02699052.2013.873821

Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods*, *52*(6), 2372–2382. https://doi.org/10.3758/s13428-020-01389-1

Thothathiri, M. (2018). Statistical experience and individual cognitive differences modulate neural activity during sentence production. *Brain and Language*, *183*, 47–53. https://doi.org/10.1016/j.bandl.2018.06.005

Thothathiri, M., Asaro, C. T., Hsu, N. S., & Novick, J. M. (2018). Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. *Cognition*, *178*, 162–177. https://doi.org/10.1016/j.cognition.2018.05.014

Thothathiri, M., & Braiuca, M. C. (2021). Distributional learning in English: The effect of verb-specific biases and verb-general semantic mappings on sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 113–128. https://doi.org/10.1037/xlm0000814

Togher, L., Elbourn, E., Kenny, B., Honan, C., Power, E., Tate, R., McDonald, S., & MacWhinney, B. (2023). Communication and Psychosocial Outcomes 2-Years After Severe Traumatic Brain Injury: Development of a Prognostic Model. *Archives of Physical Medicine and Rehabilitation*, *104*(11), 1840–1849. https://doi.org/10.1016/j.apmr.2023.04.010

Togher, L., Hand, L., & Code, C. (1997). Analysing discourse in the traumatic brain injury population: Telephone interactions with different communication partners. *Brain Injury*, *11*(3), 169–190. https://doi.org/10.1080/026990597123629

Togher, L., Power, E., Tate, R., McDonald, S., & Rietdijk, R. (2010). Measuring the social interactions of people with traumatic brain injury and their communication partners: The adapted Kagan scales. *Aphasiology*, *24*(6–8), 914–927. https://doi.org/10.1080/02687030903422478

Traumatic Brain Injury: Updated Definition and Reporting. (2015, April 6). Department of Defense. https://www.health.mil/Reference-Center/Policies/2015/04/06/Traumatic-Brain-Injury-Updated-Definition-and-Reporting

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Psychology Press.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic Influences On Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language*, *33*(3), 285–318. https://doi.org/10.1006/jmla.1994.1014

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 528–553. https://doi.org/10.1037/0278-7393.19.3.528

Truong, J. Q., & Ciuffreda, K. J. (2016). Comparison of pupillary dynamics to light in the mild traumatic brain injury (mTBI) and normal populations. *Brain Injury*, *30*(11), 1378–1389. https://doi.org/10.1080/02699052.2016.1195922

Tulsky, D. S., Kisala, P. A., Victorson, D., Carlozzi, N., Bushnik, T., Sherer, M., Choi, S. W., Heinemann, A. W., Chiaravalloti, N., Sander, A. M., Englander, J., Hanks, R., Kolakowsky-Hayner, S., Roth, E., Gershon, R., Rosenthal, M., & Cella, D. (2016). TBI-QOL: Development and Calibration of Item Banks to Measure Patient Reported Outcomes Following Traumatic Brain Injury. *Journal of Head Trauma Rehabilitation*, *31*(1), 40–51. https://doi.org/10.1097/HTR.0000000000000131

Turkstra, L. S., Brehm, S. E., & Montgomery, E. B. (2006). Analysing Conversational Discourse After Traumatic Brain Injury: Isn't It About Time? *Brain Impairment*, *7*(3), 234–245. https://doi.org/10.1375/brim.7.3.234

Turkstra, L. S., Ray, M. R., LeBlanc, M. M., Lu, L. H., Curtiss, G., Bowles, A. O., Eapen, B. C., & Cooper, D. B. (2025). Development and Pilot Implementation of a Theory-Based Cognitive Rehabilitation Protocol for Adults With Chronic Cognitive Complaints After Mild Traumatic Brain Injury. *American Journal of Speech-Language Pathology*, 1–18. https://doi.org/10.1044/2024_AJSLP-24-00306

Tyler, J. (2014). Prosody and the Interpretation of Hierarchically Ambiguous Discourse. *Discourse Processes*, *51*(8), 656–687. https://doi.org/10.1080/0163853X.2013.875866

Vakil, E., Greenstein, Y., Weiss, I., & Shtein, S. (2019). The Effects of Moderate-to-Severe Traumatic Brain Injury on Episodic Memory: A Meta-Analysis. *Neuropsychology Review*, *29*(3), 270–287. https://doi.org/10.1007/s11065-019-09413-8

van Dijk, T. A. (1979). Relevance assignment in discourse comprehension. *Discourse Processes*, *2*(2), 113–126. https://doi.org/10.1080/01638537909544458

van Dijk, T. A. (1998). Cognitive Context Models and Discourse. In *Language Structure, Discourse, and the Access to Consciousness* (pp. 11–52). Amsterdam: Benjamins.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Van Kuppevelt, J., (1995). Discourse Structure, Topicality and Questioning. *Journal of Linguistics*, *31*(1), 109–147. JSTOR.

Vanderploeg, R. D., Silva, M. A., Soble, J. R., Curtiss, G., Belanger, H. G., Donnell, A. J., & Scott, S. G. (2015). The Structure of Postconcussion Symptoms on the Neurobehavioral Symptom Inventory: A Comparison of Alternative Models. *Journal of Head Trauma Rehabilitation*, *30*(1), 1–11. https://doi.org/10.1097/HTR.0000000000000009

vanRij, J., Wieling, M., & van Rijn, H. (2022). *Itsadug: Interpreting time series and autocorrelated data using GAMMs* [Computer software]. https://cran.r-project.org/package= itsadug

Walker, M. A. (1998). Centering, anaphora resolution, and discourse structure. In *Centering theory in discourse* (pp. 401–435).

Wilde, E. A., Li, X., Hunter, J. V., Narayana, P. A., Hasan, K., Biekman, B., Swank, P., Robertson, C., Miller, E., McCauley, S. R., Chu, Z. D., Faber, J., McCarthy, J., & Levin, H. S. (2016). Loss of Consciousness Is Related to White Matter Injury in Mild Traumatic Brain Injury. *Journal of Neurotrauma*, *33*(22), 2000–2010. https://doi.org/10.1089/neu.2015.4212

Winn, M. B., & Teece, K. H. (2021). Listening Effort Is Not the Same as Speech Intelligibility Score. *Trends in Hearing*, *25*, 1–26. https://doi.org/10.1177/23312165211027688

Winn, M. B., & Teece, K. H. (2022). Effortful Listening Despite Correct Responses: The Cost of Mental Repair in Sentence Recognition by Listeners With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *65*(10), 3966–3980. https://doi.org/10.1044/2022_JSLHR-21-00631

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*, 233121651880086. https://doi.org/10.1177/2331216518800869

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. https://doi.org/10.1037/a0029206

Wood, S. N. (2003). Thin Plate Regression Splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood, S. N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Wüst, L. N., Capdevila, N. C., Lane, L. T., Reichert, C. F., & Lasauskaite, R. (2024). Impact of one night of sleep restriction on sleepiness and cognitive function: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *76*, 101940. https://doi.org/10.1016/j.smrv.2024.101940

Yeh, J.-F., Tan, Y.-S., & Lee, C.-H. (2016). Topic detection and tracking for conversational content by using conceptual dynamic latent Dirichlet allocation. *Neurocomputing*, *216*, 310–318. https://doi.org/10.1016/j.neucom.2016.08.017

Yu, Z., Gu, Z., Shen, Y., & Lu, J. (2025). The relationship between language features and PTSD symptoms: A systematic review and meta-analysis. *Frontiers in Psychiatry*, *16*, 1476978. https://doi.org/10.3389/fpsyt.2025.1476978

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zakzanis, K. K., McDonald, K., & Troyer, A. K. (2011). Component analysis of verbal fluency in patients with mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *33*(7), 785–792. https://doi.org/10.1080/13803395.2011.558496