RELATIONAL ENCODING EFFECTS ON EVENT COHERENCE IN LANGUAGE PRODUCTION

Sandra A. Zerkle

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the Department of Psychology and Neuroscience (Cognitive Program).

Chapel Hill
2018

Approved by:

Jennifer Arnold

Peter Gordon

Jessica Cohen

ABSTRACT

Sandra A. Zerkle: Relational encoding effects on event coherence in language production
(Under the direction of Jennifer Arnold)

When describing a series of events, speakers can explicitly signal the connection between events by using coherence markers (pronouns and connector words). Current models of reference production do not fully account for observed variation in reference form. This study attempts to solve this problem by testing if qualities of the discourse context and strength of event connectedness in memory can affect speakers' production choices. In two experiments, relational encoding of events was manipulated such that some participants had a higher degree of connection than others. Both experiments found that the encoding manipulation did not affect coherence marker use. In Experiment 1, it did not have the expected effect on memory nor language production. In Experiment 2, it did affect event relatedness representations (recall accuracy), but did not affect coherence marker use on accurate trials. Further research is needed to determine whether relational encoding can affect language production in other situations.

ACKNOWLEDGEMENTS

This work was partially supported by NSF grant #1348549 awarded to Jennifer Arnold. Thank you to my advisor, Jennifer Arnold, for her mentorship and support on this project. I also extend my thanks to my committee members for their insightful comments. Thank you to my cohort members, my fellow lab members, and the undergraduate research assistants who helped with collecting and coding the data. Finally, thank you to my parents for their unwavering support and encouragement on my academic journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A speaker makes many choices when planning and producing language. Planning each message involves decisions about what information to mention, and in what order. Speakers have a choice about whether they explicitly identify the relation between events in their discourse or not. Linguistic devices like reduced referential forms and temporal connector words like *and, then, so* can be used to signal coherence within the discourse (Grosz et al., 1995; Ariel, 1990; Givon, 1983; Gundel et al., 1993; Chafe 1976, 1994; Fraser, 1999). What determines whether speakers choose to signal the connection between events in these ways?

This study investigates whether qualities of the discourse context and strength of event coherence in memory can influence measurable changes in the structure of the discourse as it unfolds. I propose a model of reference form choice that is built on previous research, but additionally emphasizes the strength of event representations in memory as a constraint on reference. In the following sections, I first describe the current state of models of reference, I highlight the idea that these models have a problem that needs to be addressed, and I test my proposed model in two experiments by manipulating memory for events directly and measuring language production.

Background

Most scholars agree that one of the major goals of communication is to be informative: give as much information as is needed, and no more (Grice, 1975). If you have one pet animal in your home you might refer to it as “it”, but if there are several different pets, you will have to be more informative to signal which one you mean by saying “the black cat” or “Spooky.” Several models of language processing suggest that we maximize the likelihood of efficient and successful message communication in similar ways (Frank & Goodman, 2012; Frank & Jaeger, 2008; Kehler et al., 2008; Rohde & Kehler, 2014).

One major component of communication is the production of referential expressions, which function to link entities in different utterances together, which might result in memory representations where these entities are connected for the listener. Speakers must make a decision about how explicit each of their references will be: they must choose between pronouns and more explicit names or descriptions (*it* vs. *the cat*) based on certain linguistic constraints. Previous work on discourse coherence seems to assume that if there is a salient relation between references, the speaker will signal it: Speakers select shorter expressions when a referent is predictable in context (Levy & Jaeger, 2007; Mahowald et al., 2013), and the probability of using a reduced reference is represented by the salience of that referent within the context (Frank & Goodman, 2012). The more salient or in focus the referent is in the context of the discourse, the more likely it is to be referred to with less explicit referring expressions like pronouns and null anaphors (Almor, 2000; Ariel, 1990; Brennan et al., 1987; Chafe, 1976; Gundel et al., 1993; Givon, 1983; Prince 1981; Rohde & Kehler, 2014; Rosa & Arnold, 2017).

Existing models tend to agree that if a speaker mentions a character in salient, topical, and accessible ways, then the speaker would likely use a pronoun to refer to it in subsequent utterances, and they assume that the effect of the discourse context is constant (see Arnold, 2016 for a review). Consider the following story:

Anna went to the grocery store with Lily. She brought a list with her.

In the second sentence, repeating *Anna* in place of *she* would seem unnatural, because it is reasonable to assume that *she* refers to *Anna*. Anna is mentioned in a syntactically prominent position in the first sentence, therefore the representation of her in the speaker's mental model of the discourse is more accessible, and the speaker is more likely to use a reduced form in the second sentence to refer to her (Arnold, 2008, 2010; Ariel, 1990, 1996, 2001; Chafe, 1976, Gundel et al., 1993).

Co-reference is due in part to coherence; when coherence devices are used, there is a high degree of overlap in the information conveyed (Hobbs, 1979). The specific coherence devices that I refer to here tend to be that of Occasion/Result (initial state -> final state; *John gave a book to Zach. He put it on the shelf;* Kehler, 2002). Many events in a discourse tend to be structured temporally, and efficient communication establishes co-reference by using reduced forms and connector words like *and, then, so*. People tend to produce pronouns more often in sentences that contain an explicit connector, likely because the presence of connectors reflects a stronger conceptual connection between the events being described, and a greater preference for the speaker to treat the utterance as a part of the prior discourse context (Arnold & Griffin, 2007; Arnold & Nozari, 2017). Thus, one of the main goals of efficient communication is to be informative when representations are accessible by signaling connection in the discourse.

In order to produce an utterance, a speaker must have a representation of the event or message that they plan to say. How does a speaker get from a conceptual representation in their mental model to the physical production of each sequential word in a sentence? The time course of utterance planning has been well modeled in the field of psycholinguistics, but only one model specific to referent production has been proposed. Schmitt, Meyer, & Levelt (1999) built off of Levelt's (1989) model of utterance planning to create a model of lexical access in the production of pronouns. Traditional models of utterance planning generally agree that the generation of an utterance begins with the transformation of a communicative intention into a preverbal conceptual message, the speaker then encodes the grammatical formulation of this message (which includes lexical/lemma selection and functional assignment), phonological encoding and syllabification occurs, and then the phonemes are encoded and articulated (see Figure 1; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). Visual input can affect the early conceptual message level, such as in picture naming tasks (Roelofs, 1992; Indefrey & Levelt, 2004).

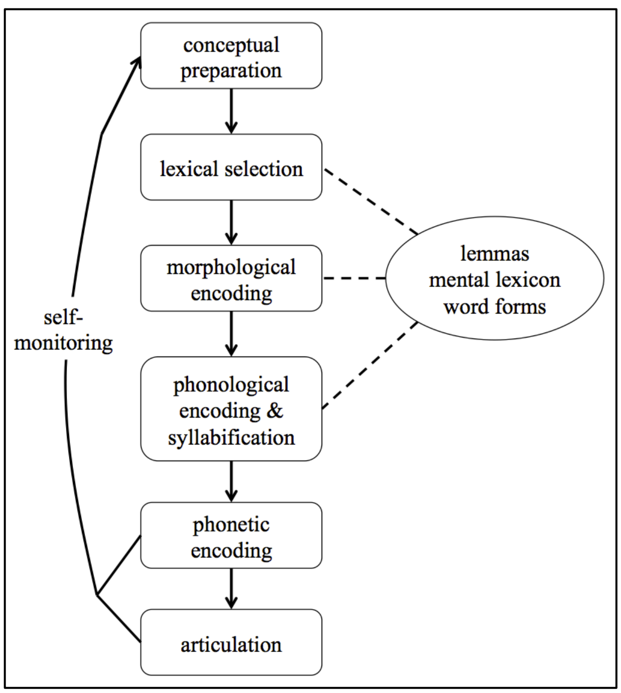

Figure 1. A model of utterance planning recreated from Levelt, Roelofs, & Meyer (1999).

Schmitt et al. (1999) extend this model to the production of pronouns, and they propose that speakers must fit each message fragment into the current discourse, which involves marking certain concepts as old information (*in focus*), and others as new (*not in focus*) (Chafe, 1976). These types of information can be referred to linguistically in different ways, from explicit noun phrases (*the red flower*) to reduced pronouns (*it*). Such marking helps speakers and listeners maintain a coherent discourse accessibility record (Levelt, 1989). In order to lexically access pronouns, lexical concepts of the entities are marked as being in focus or not in focus depending on the discourse context. A pronoun is selected if and only if the corresponding lexical concept is accessible and marked as in focus within the discourse; otherwise, a noun is selected. The authors propose that a referent's discourse accessibility status affects the message at the conceptual level; this accessibility node then affects the anaphoric link to lexical selection (whether a reduced form vs. an explicit form is chosen). See Figure 2 for a visualization of this model.

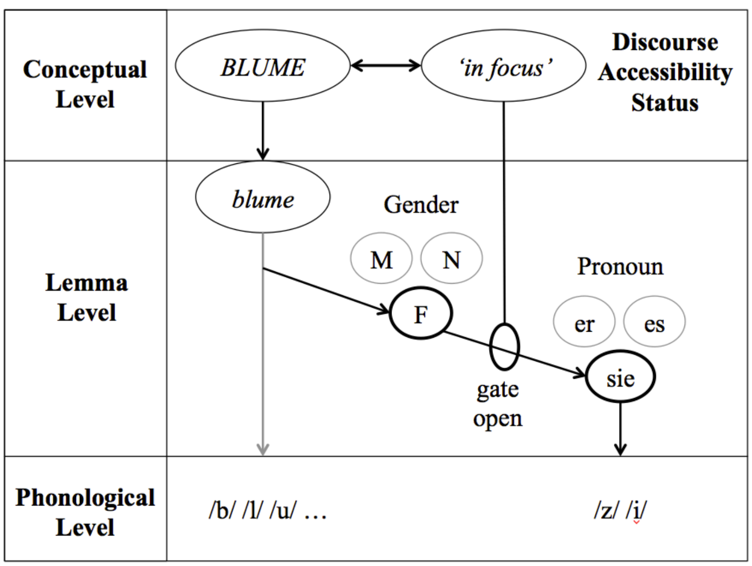

Figure 2. A model of reference form choice recreated from Schmitt, Meyer, & Levelt (1999).

For the current study, I have adapted this model of reference form choice and re-modeled it into the one depicted in Figure 3. It follows the same utterance planning sequence, such that a reference starts as a conceptual representation of the discourse context, then moves on to message encoding, and finally lexical selection and formulation. The main constraint on reference form choice is the referent's conceptual status, being either *accessible* or not *accessible;* this maps directly on to whether a reduced or explicit form is selected. Additionally, Schmitt et al.'s (1999) model was created for the purpose of thinking about gender assignment, and not for the purpose of accounting for discourse effects.

However, I must acknowledge some assumptions that I am making in extending this model. The notion of discourse accessibility is probably on a continuous gradient, rather than merely binary (accessible vs. not accessible). I also use the node of the referent's conceptual status to represent concepts like accessibility, salience, prominence, and other linguistic constraints on the referent's representation, which in turn affect the referent's "focus" status (Ariel, 1990; Chafe, 1976). Critically, these representations of conceptual mechanisms are not definitively placed in specific stages of production; therefore I have collapsed these underneath the nodes of both conceptual representation and message encoding. More work needs to be done to determine at what point in the sequence of utterance planning these constraints take hold. For the present purposes, this depiction is sufficient for the claims that I will make. Importantly, it assumes that the speaker's conceptual representation of the discourse is only constrained by the linguistic properties of the context, which as I describe in the following section, is not enough to predict variation that has been observed.

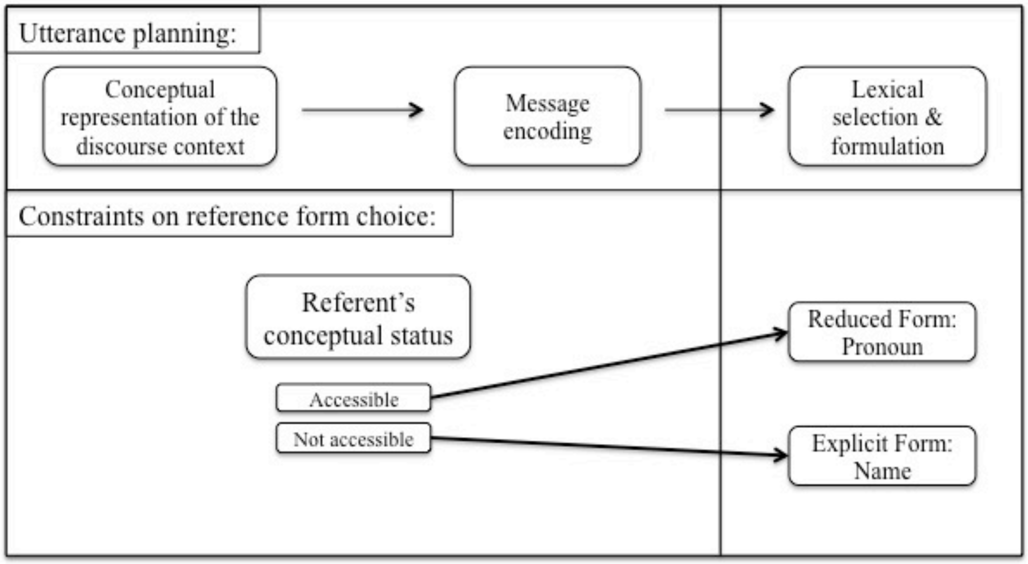

Figure 3. A model of reference form choice, adapted from Schmitt, Meyer, & Levelt (1999).

Problem: more variation in pronoun use than accounted for

The above model predicts fairly uniform effects of reference form choice. However, pronoun usage is not necessarily automatic, and just because the linguistic context is constrained to make a referent accessible, reduced forms are not always chosen. Alternative to the previously discussed models, I propose that speakers have to be interested in signaling the discourse connections in the first place in order for this linguistic device to be used. This is the main idea proposed and tested in the current paper. If the speaker is distracted, planning incrementally, or has low semantic link between entities, they may not use linguistic devices that mark coherence as much, such as pronouns and connector words like “and” and “then” (Zerkle & Arnold, 2016; Arnold & Nozari, 2017; Brown-Schmidt, Byron, & Tannenhaus, 2005). Reduced forms promote conceptual ties between events, and connector words have been shown to be highly correlated with pronoun use because they mark connectivity between events within the discourse and signal the contribution of these connections (Arnold & Griffin, 2007; Arnold & Nozari, 2017).

Further evidence for this point comes from Zerkle & Arnold (2016), where we found that some speakers used reduced forms and connector words (*and*, *then*) and pre-planned their speech while paying attention to the discourse context (called Context-Users), while other speakers used only noun phrase descriptions and virtually no connectors and initially directed their attention more to the event being described (called Context-Ignorers). We concluded that some speakers pay attention to the context while planning their utterances more than others, and that this led to variation in the use of linguistic coherence markers overall. We speculated that individuals responded differently to the competing goals of the production task: Some speakers might be more motivated than others to put their descriptions within a discourse story context than others, and consequently use more linguistic markers of event coherence like pronouns and connector words.

Similarly, Arnold & Nozari (2017) found variation in the use of these linguistic markers based on discourse connectedness. In an utterance production task, they found that speakers' use of pronouns and zeros increased on trials with longer latencies, suggesting that this enabled more pre-planning of the message and greater discourse connectedness. The strongest predictor of reduced reference form usage was the presence of a connector word like *and* or *then*, which also tended to occur on trials with longer latencies. However, temporal breaks in the visual input for this task created breaks in the discourse, and discouraged reduced forms. These findings suggest that discourse connectivity and reference form are tightly related, such that speakers can adopt a mode of production in which utterances are pre-planned, but this also depends on the availability of the conceptual information to be described. From these recent pieces of evidence, I can conclude that speakers will not always use reduced forms in ways that are assumed by most models of referential probability, and that speakers must be motivated by a broader communicative goal to convey event coherence in order for these markers to be used successfully.

Discourse connectivity has also been shown to be supported by activity in the prefrontal cortex (PFC). Arnold & Nozari (2017) also used transcranial direct current stimulation (tDCS) to stimulate the PFC, an area shown to be involved in executive functions, goal-directed behavior, and working memory (D'Esposito, 2007), to understand how it modulates utterance planning and reference. In stimulated participants, this increased their ability to produce utterances with greater discourse connectedness, even while planning incrementally. The latency effect found in sham participants described above was not present in participants who received stimulation. The authors conclude that the left PFC contributes to a speaker's ability to produce coherent, pragmatically appropriate discourses. The PFC has also been shown to be involved in relational memory functions (described below, Giovanello & Dew, 2015), which I speculate may also be important for discourse connectivity. The role of the PFC in this relationship is not tested directly in the current study, but it may have interesting implications for future research.

Proposed solution: a new model for variation in pronoun use

There are two related reasons why there might be variation in pronoun usage. One is that a speaker's communicative goals are influenced by how they represented the events originally; the second is that a speaker must have the cognitive abilities to recognize and communicate relations in order to successfully use discourse coherence devices. For an example of event representation, consider the two events depicted as illustrations in Figure 4:

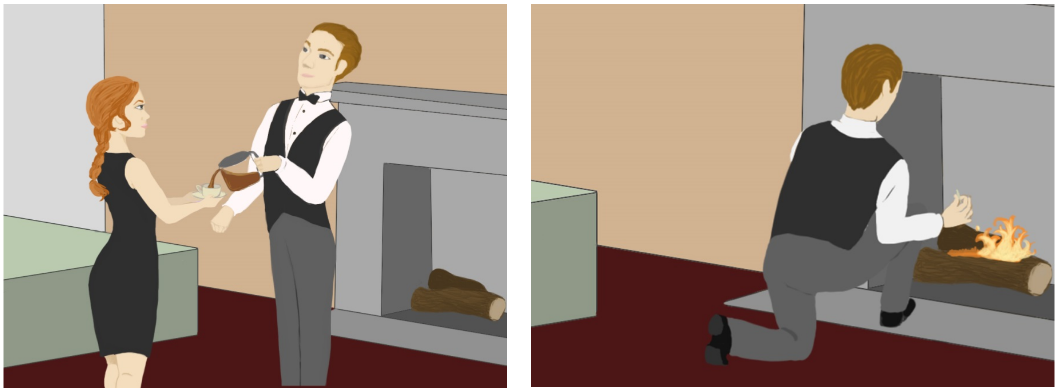

Figure 4. Events involving the maid and the butler.

These events are seemingly unrelated; the butler could either be lighting the fire or putting it out, which could either happen before or after he pours the coffee for the maid. Knowing the actual order of the events (that pouring coffee happens before starting the fire) creates a representation of these events as a connected pair in memory. When it comes time to tell the story of these characters, some speakers might pay more attention to the temporal relationship between the events, while others might focus less on the events as a connected set and more on the details of each individual event. If a speaker created a strong representation of these events in memory, they might be motivated to communicate this relationship and be more likely to describe it as: "The butler brought some coffee to the maid, and then he lit the fire," using linguistic markers of discourse coherence like pronouns and connectors to indicate the change from initial state to final state. However, another speaker might not have created a strong representation of these events as an ordered pair, which could decrease the likelihood that this speaker is motivated to communicate this connection. They might describe the events as completely separate, using explicit forms of reference and no connector words: "The butler brought some coffee to the maid. The butler lit the fire." In sum, this is a possible reason why a speaker's original strength of the representation of events affects their subsequent motivation to use of coherence markers in their utterances.

However, some events have inherent relatedness based on the semantics. There are well-known effects of discourse coherence that are related to the linguistic structure; this includes both grammatical role and thematic role effects. For example:

The Duke received a painting from the Duchess.

{and then/He/The Duke} threw it in the closet.

In the second sentence, many forms of reference are acceptable to use (the noun phrase *The Duke*, the pronoun *He*, or the verb phrase coordination "zero" preceded by a connector *and then*). The first sentence contains some structures that have been shown to affect the choice of reference form in the second sentence. The grammatical role of the target character affects this choice: if the target is mentioned in the subject role of the context, the reference is more likely to be reduced (pronoun or zero) (Arnold 1998; Brennan, 1995). The context also utilizes a transfer verb, such that an object (the painting) is transferred from the source (the Duchess) to the goal (the Duke). Several studies have found that goals tend to be more predictable in these contexts, and also tend to elicit more reduced references than if the source were the target character (Arnold, 2001; Rosa & Arnold, 2017; Zerkle, Rosa, & Arnold, 2015; Zerkle & Arnold, under revision). Models of coherence relations with respect to thematic roles tend to assume that speakers are representing some relation in memory in order to make these connections (Kehler, 2002). Therefore, I hypothesize that goal continuation events are inherently more related than source continuation events, and that having highly related representations for sets of transfer events might make goal continuations even more connected, driving coherence marker use even higher.

I propose to add this set of constraints to the aforementioned reference form choice model, such that the events' inherent relatedness (reflecting the difference between goal and source continuations) affects the strength of the relational encoding in memory, which in turn affects the likelihood of using a reduced form. See Figure 5 for this model, with the proposed constraints in the purple box. Based on findings from Zerkle & Arnold (2016) and Arnold & Nozari (2017), I hypothesize that speakers must have strong relational encoding and be motivated to convey event coherence in order for discourse markers to be used. The purpose of the current studies is to test whether reference form choice is influenced by the strength of the connection between events made in memory. In the model figure, I simplify this to the strength of relational encoding of events being 'strong' vs. 'weak'. In the current paper, I directly test this with manipulations meant to induce differences in event representation, and measure the variability in language production. I also hypothesize that the accessibility constraints described above still play a role, but the experiments in this paper test whether that variation in the strength event representations has an additional effect.

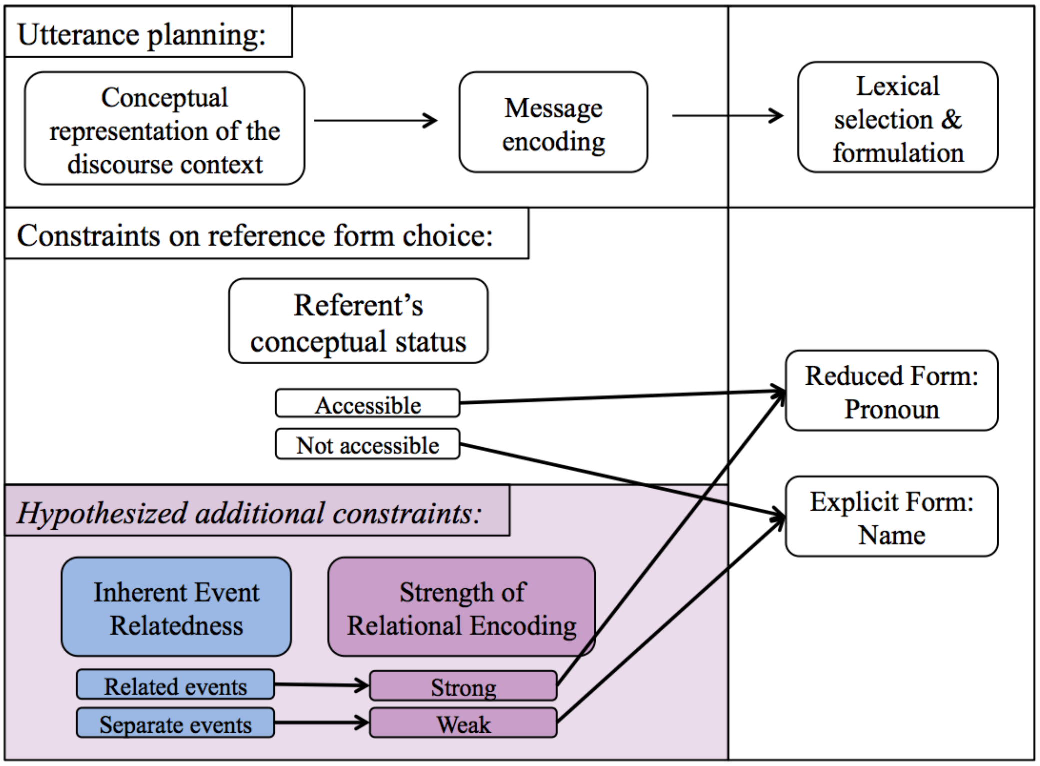

Figure 5. Hypothesized constraints to the model of reference form choice.

One of the benefits of this expanded model is that it may offer a straightforward mechanism to explain how reference form is influenced by executive dysfunction. The second reason for variation in pronoun usage is that a speaker must have the cognitive abilities to recognize and communicate relations in order to successfully use discourse coherence devices. Arnold & Nozari (2017) found that the PFC is important for maintaining discourse coherence, which may suggest that someone who has deficits in this area might be less likely to use coherence markers in their speech, in addition to having other executive dysfunctions. People with disorders like autism have been shown to have a deficit in the appropriate use of language in context (Tager-Flusberg, 1999; Kjelgaard & Tager-Flusberg, 2001). This is not the focus of the current project, but could be the basis for further investigation of this question. Thus, I conclude that both strong representations of connected events in memory and the cognitive abilities to do so may be necessary for the successful use of discourse coherence markers to communicate connected events. Importantly, I am analyzing the behaviors of healthy young adults, and I make the assumption that they all possess the cognitive abilities necessary to recognize and communicate these relations. Thus, variation in pronoun use would stem from linguistic constraints, inherent event relatedness constraints on encoding strength, or some combination of the two.

Representations of event relatedness

The current set of studies assesses the effect of event representations in memory on the production of connected discourse. I test the hypothesis that variation in discourse connectivity is in part due to the speaker's relational representations for the underlying events being conveyed (see Figure 4). To do this, I have manipulated the speaker's goal while encoding a set of events before describing them, such that some speakers were encouraged to pay more attention to the temporal relationship between the events, while others were encouraged to focus less on the events as a connected set and more on the details of each individual event. Most studies in the area of relational memory focus on memory for objects or words (Wang & Giovanello, 2016; Addis et al., 2014), but here I am extending this to memory of events. Memory for past events typically involves remembering the various details of the event (item memory), as well as the associations formed between the details (relational memory) (Giovanello & Dew, 2015). In order to retrieve memories of our experiences, we must recall not only the individual components but also the way in which all of the components were bound together. This partially supports the hypotheses of the current study, however there are some additional factors to consider. One event can have several details that need to be bound together to create a representation of that single event (intra-event associations). In addition, one can bind two events together to create a representation of these events as related and connected in some way (inter-event associations). Thus far, inducing the encoding of two pictured events as connected has not been tested directly.

Relational memory depends on several brain regions (primarily PFC and medial temporal lobes (MTL)) to interact in order to successfully encode each relationship. Relational memory is also more sensitive to healthy aging than item memory, because it is a more involved, difficult cognitive process. Since the PFC has been shown to influence the use of discourse coherence devices (Arnold & Nozari, 2017), it seems straightforward to assume that relational memory and discourse coherence are related to some extent and/or in certain situations. This motivates the current study, which tests whether it is possible to induce variation in representations of event relatedness, and whether this affects the use of discourse coherence devices.

The current studies

In this set of experiments, I tested whether I could experimentally manipulate the connection between events in memory, and whether this affects choices of linguistic devices used in language production. Experiment 1 tests the question of whether relational representations of events support the use of pronouns and connector words. I manipulated the perceived connection between events that speakers make when first encoding a set of events: In one between-subjects condition participants were encouraged to think of each picture as a single event (**Sequential condition**), and in another condition participants were encouraged to think about how the pictured events related to each other (**Side-by-Side condition**). They were told that they will take a subsequent memory test on these pictures in order to encourage effortful encoding. This manipulation mimics the item/relational processing manipulations used in other relational memory studies (Wang & Giovanello, 2016; Addis et al., 2014). Then, the speakers performed a language production task where they described the same set of events with an interlocutor. I measured the referential form choice of each utterance, as well as the use of connectors as indicators of connection with the previous discourse context. I predicted that if the encoding manipulation affected event representations, then participants in the Side-by-Side condition would use more coherence markers than participants in the Sequential condition, and that they would have better memory for the events overall. Experiment 2 also pursued the idea that a memory-encoding manipulation could modulate a speaker's production, but used slightly different methods and instructions.

This set of studies employs a well-understood production paradigm that has been shown to elicit varied pronoun use. One reason why it was used here is that the items are already designed such that the linguistic context affects pronoun use in controlled ways. By including a manipulation of thematic roles in the stimuli, I assessed a secondary question of whether the effect of thematic role on reference form choice also depends on the level of event relatedness in memory (blue box in Figure 4). Previous models of coherence relations and thematic roles tend to assume that speakers are representing some relation in memory in order to make these connections (Kehler, 2002). However, most of these studies have used narrative continuation paradigms, which may not be representative of natural language production mechanisms. The current paradigm is a natural language production task, which has rich properties of discourse that are more similar to everyday conversations. In Zerkle & Arnold (2016), we used this same paradigm and found that there are individual differences in how speakers attend to the context, which has effects on pronoun use. This difference is above and beyond the difference found for the goal bias on reduced forms. Any test of event relations must take into account the linguistic context, especially when using a paradigm that already manipulates both grammatical role and thematic role.

EXPERIMENT 1

The current set of studies draws on an existing paradigm that was designed to investigate thematic role effects on pronoun use (Rosa, 2015; Rosa & Arnold, 2017). One advantage of this is that I can test relational encoding effects on reference form against the backdrop of a well-characterized discourse context task. This paradigm controls for grammatical role and thematic role of the linguistic context, as well as gender matching between characters. Participants first view 106 pictures, which together create a story about a murder mystery (see Figure 6).

Figure 6. Illustration pictures (context on left, target on right) for one trial.

Then, participants are told that they will meet with a detective (a confederate experimenter), and they will help describe these pictures with the detective in order to solve the crime. The detective and the participant go through the pictures in 53 pairs; the detective describes the event in the first picture, and the participant's task is to describe the event in the second picture. The linguistic structure of the detective's sentence is manipulated such that the target character (the character who continues to perform an action in the second picture) is either the subject or non-subject (object of the preposition), and is either the goal or the source of the transfer event. Additionally, the gender matching is controlled for such that on half of the trials the gender of both characters is the same, and the on the other half they are different. Participants are instructed to describe the second picture after the detective describes the first, and utterances are recorded, transcribed, and coded for reference form choice and connector use. These measures are then analyzed to test whether there are systematic differences between the relational and thematic conditions of these items. Previous studies using this paradigm have replicated the subject bias to use reduced forms, as well as found a goal bias to use reduced forms (Rosa & Arnold, 2017; in the Context-Users in Zerkle & Arnold, 2016; Zerkle & Arnold, under revision).

Experiment 1 tests the hypothesis that related representations of events supports the use of pronouns and connector words. The between-subjects manipulation for the current set of studies is designed to test the role of relational encoding of the events during the first viewing. I have added instructions before this viewing, such that half of the participants were shown the pictures in 53 item pairs and were told to learn the picture story with the goal of remembering the order of the events (**Side-by-Side condition**); and half of the participants were shown the 106 pictures one at a time and were told to learn the picture story with the goal of remembering the details of each picture (**Sequential condition**). Then participants did the production task, viewing each pair of events, listening to the detective's context sentence, and continuing with their own description of the second picture. For the current set of studies, I only used the detective context sentences that have the target character in the subject position, because we have found that these trials promote the most pronoun variation overall in previous versions of this task (Zerkle & Arnold, 2016; under revision). I predict that if the memory manipulation leads to differences in event representations, then it will also affect pronoun and connector production. If so, participants in the Side-by-Side condition would use more pronouns overall than participants in the Sequential condition, and perhaps especially more for goal continuations than source continuations. At encoding, these participants were encouraged to represent the events of the story in a coherent way, and I believe that this memory representation will have effects on later production, particularly reference form choice and connector use. Participants in the Sequential condition might show patterns of pronoun use similar to that of the Context-Ignorers in Zerkle & Arnold (2016) – since they were encouraged to encode these events separately, they may not have strong event representations in memory, and thus do not have the communicative goal to describe them as connected. After the production task, all participants were given two memory tests. The Relational test showed the filler trial pictures in pairs, either in the correct ordering or the reversed ordering, and participants were asked to select between the two options. The Item test showed one panel of each filler trial pair (half first panel, half second panel, randomized), either in the correct orientation or the mirror image orientation, and participants were asked to select between the two options. These tests were designed to act as a manipulation check: accuracy was measured for each participant on these questions in order to assess whether participants in each condition successfully encoded the pictures in the way they were instructed to.

Therefore, my main hypothesis is that if the encoding manipulation affects event representations, then it would have a direct effect on reference form and connector production. Since this paradigm is also designed to test effects of thematic role on reference form, this hypothesis also predicts interactions between relational encoding and thematic role on reference form choice.

Methods

Participants

Data from 43 native English participants were collected: 22 in the Sequential condition and 21 in the Side-by-Side condition. Participants were UNC undergraduates in the Summer (N=12) and Fall (N=11) of 2017 who received SONA subject pool credit, or residents of Chapel Hill in the Summer of 2017 who received $10 compensation (N=20).

Materials and Measures

Stimuli come from http://jaapstimuli.web.unc.edu/transfer-verb-stimuli-and-paradigm/ (published in Rosa & Arnold, 2017). Participants first watched an introduction slideshow of illustrated pictures with narration that describes the murder mystery story and the participant's role as a tabloid photographer, introduces the 6 characters and the location (manor house), and explains that the participant will meet with a detective to help solve the crime (in Microsoft PowerPoint). Participants in the Side-by-Side condition were then told that they will first view all of the pictures in order, and were instructed to "remember the order of the events, because you will be tested on them later." Then all 106 pictures were presented in their 53 pair sets (see Figure 5), such that each pair was presented on the screen for 5 seconds and automatically advances. Participants in the Sequential condition were told that they will first view all of the pictures in order, and were instructed to "remember the details of each picture, because you will be tested on them later." Then all 106 pictures were presented one at a time, such that each picture was presented on the screen for 2.5 seconds and automatically advanced.

After all pictures were presented, participants in both conditions were shown an example trial of the production task, with instructions from the same narrator (in Microsoft PowerPoint).

This shows one pair of pictures, gives a description of the first picture, and then gives an example of what the participant might say to describe the second picture, using both connectors (*and then*) and a pronoun to refer to the character. In Zerkle & Arnold (2016), we found that this type of example, along with guided instructions from the experimenter to "make the descriptions sound like a story," leads to an increase in coherent discourse overall. Therefore, I have used similar instructions here in both conditions.

The participants were then instructed that the main production task will begin (in MATLAB). Regardless of their relational encoding condition, this task was the same for all participants. They were presented with a pair of pictures on the screen in the correct order of events. A voice played over headphones that is the detective describing the first picture. These sentences were controlled for thematic role, such that half of the critical items (24) have the goal character in the subject position, and half have the source character in the subject position. The character in the subject position of this sentence is always the target character in the second picture, such that half of the items are goal continuations and half are source continuations. See examples below:

Goal continuation: *The maid took a basket of laundry from the Duchess.*

{She/The maid} poured bleach in the basket.

Source continuation: *The butler handed an apron to the maid.*

{He/The butler} helped tie it on.

Participants then described the second picture, in any way they want. They were audio recorded from the beginning of the trial until after their response, and their responses were transcribed by research assistants and then coded for pronoun/zero/description choice and connector use. Reference form type and connector use are binary variables (either 0 or 1). Trials were only included if the participant referred to the target character in the subject position of their utterance; all other trials were excluded from all analyses. After all 53 items were described, the participants were told that the production task is over.

After the production task, the participants took a memory test on Qualtrics. Participants in both conditions were first instructed: “Now you will be asked what you remember about these pictures. Please do your best to answer the following questions. Each page will stay up until you move ahead; you can take as long as you need to answer each one. Please ask the experimenter now if you have any questions.” First, all participants took the Relational test which showed the filler items (29) in pairs of pictures, in the order in which they were presented in the first viewing of the story. On half of the questions, the order of the pair was in reverse such that the second event is shown on the left, and the first event is shown on the right. On the other half of the questions, the correct order of events was shown. Each pair was presented along with the question, “Are these pictures in the correct order?” and the multiple choice options underneath were “Correct Order” and “Incorrect Order.” Participants must have selected one of these two options in order to move on to the next question, and they were not presented with feedback about their accuracy. Then all participants took the Item test, which showed the filler items (29) in the order in which they were presented in the first viewing, but only one of the pictures will be shown for each question. Half of the questions showed the first event, half showed the second event. Each picture was presented along with the question, “Is this picture identical to what you saw before, or the mirror image?” and the multiple choice options underneath were “Identical” and “Mirror Image.” Participants must have selected one of these two options in order to move on to the next question, and they were not presented with feedback about their accuracy (see Figure 7 for example questions for both tests). They then took a post-experiment questionnaire that asked about their knowledge of the story and their strategies in the production task and the memory test.

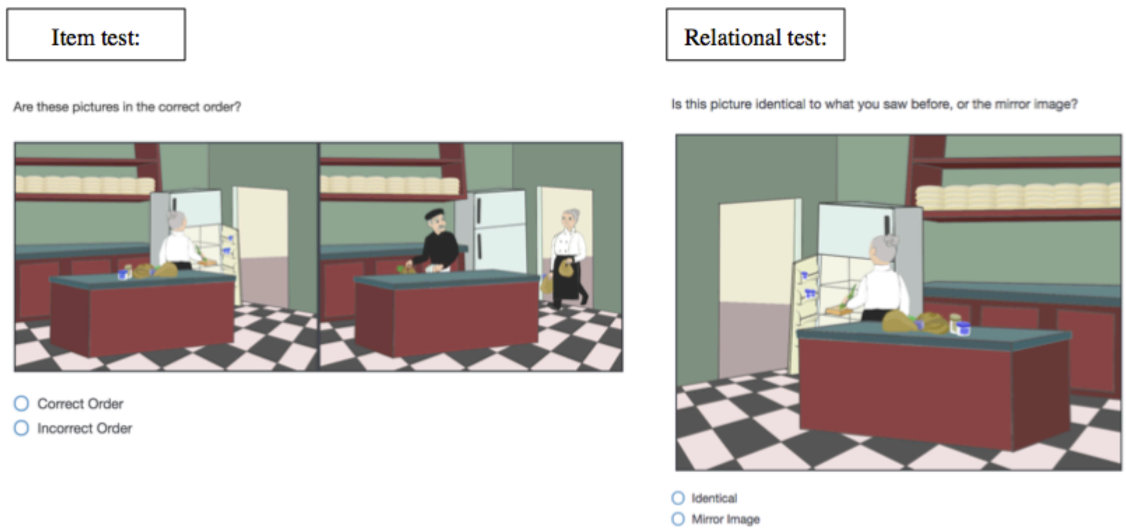

Figure 7. Item test and Relational test example questions.

Procedure

Participants were briefly introduced to the task by an experimenter, and they filled out a brief demographics survey. They first watched the narrated introduction slideshow on a computer screen, and then viewed all of the pictures in order silently. The experimenter played an example trial, and then started the main production experiment. Participants put on a pair of headset headphones with a microphone attachment, and this played the audio context sentences as well as recorded the participants' responses. After all 53 items, the experimenter then administered the memory tests and the post-questionnaire, and then the participant left the lab.

Analytic approach

Generalized linear mixed effects models were used to analyze the data, in order to account for the dependencies in this repeated measures design. SAS Proc Glimmix was used to analyze dichotomous outcomes with a logit link (reference form, connectors). Models of each dependent variable were constructed with random intercepts of participant and item to account for these differences. Random slopes of primary predictors by participant and by item were included when appropriate (if they improved model fit), and excluded if estimated to be zero (Searle et al., 1992). The two memory tests accuracy measures only used the filler items, so these variables must be each participant's average score, as well as the overall difficulty scores for each participant. These were analyzed by linear regression using SAS Proc Mixed.

Primary predictors of interest were encoding condition (Side-by-Side /Sequential), thematic role continuation (goal/source), reference form type (reduced/explicit), and connector use (used/not). Interactions between these predictors were increasingly added to each model, and only retained in each final model if nearly significant at $|t| > 1.5$. Effects coding was used for all binary predictors. All predictors were grand-mean centered.

Results

All recordings were transcribed by research assistants, and coded for reference form and connector use. 954 trials were included in the following analyses. 78 critical trials were excluded: 74 began with reference to the incorrect character, and 4 had audio recording issues such that the responses could not be transcribed. See Table 1 for a summary of the model outputs, and see the Appendix for more details about each model.

Does relational encoding manipulation affect memory accuracy?

My first question was to see if the encoding manipulation was successful by analyzing the results of each memory test. Accuracy was at ceiling for participants in both conditions for both the Relational (99%) and Item (97%) tests, indicating that both tests were too easy. Statistical analyses showed no effect of condition, no effect of thematic role, and no interaction between the two for neither the Item test nor the Relational test (see Figure 7, Table 1).

One question in the post-experiment questionnaire was to rate the difficulty of the memory test on a 1-5 scale (1 being Very Easy and 5 being Very Difficult). Statistical analyses showed no effect of condition on this measure; the Sequential group had an average score of 2.17, and the Side-by-Side group had an average score of 2.16. This supports the conclusion that the memory tests were on average Somewhat Easy to participants in both groups.

Does relational encoding manipulation affect reference form production?

My next question was to see if the encoding manipulation affected reference form choice. There was a numerical difference between relational encoding conditions, however it was not in the predicted direction – participants in the Sequential condition used pronouns/zeros on 48% of the trials, while participants in the Side-by-Side condition used pronouns/zeros on 35% of the trials (see Figure 8). There was a significant thematic role effect in the predicted direction such that more pronouns/zeros were used for goal continuation trials (replicating previous findings: Arnold, 2001; Rosa & Arnold, 2017; Zerkle, Rosa, & Arnold, 2015; Zerkle & Arnold, under revision). There was also a significant effect of connector use such that more pronouns/zeros were used on trials where a connector was also used. No interactions between these predictors were significant when tested. See Table 1 for model output.

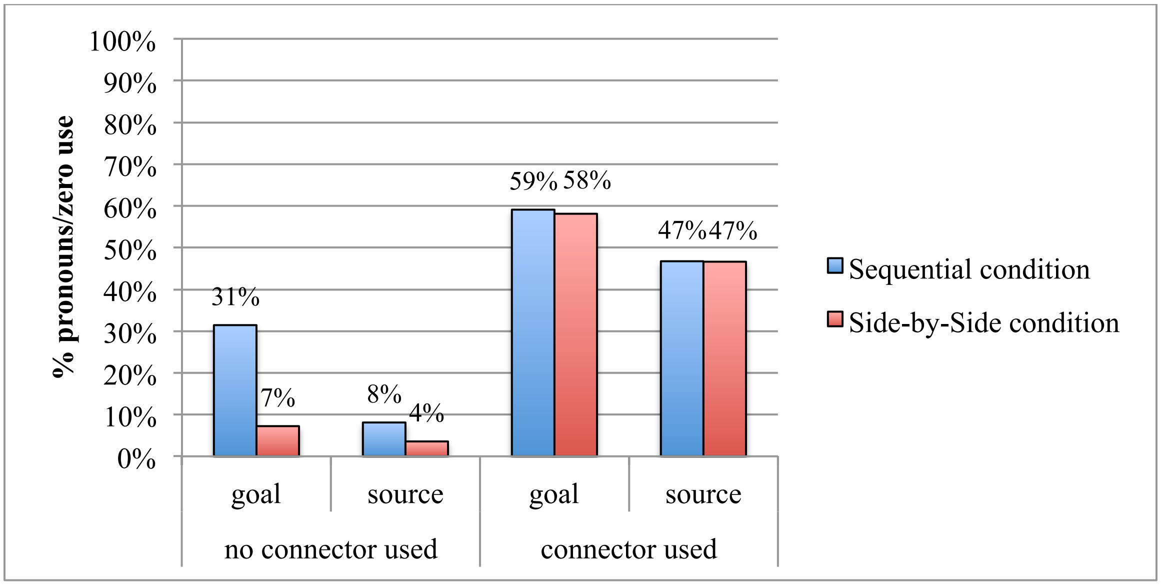

Figure 8. Experiment 1 pronouns/zeros use.

Does relational encoding manipulation affect connector production?

My final question was to see if the encoding manipulation affected connector use. There was a significant effect of relational encoding condition, however it was also not in the predicted direction – participants in the Sequential condition used more connectors (85%) than participants in the Side-by-Side condition (64%) (see Figure 9, Table 1).

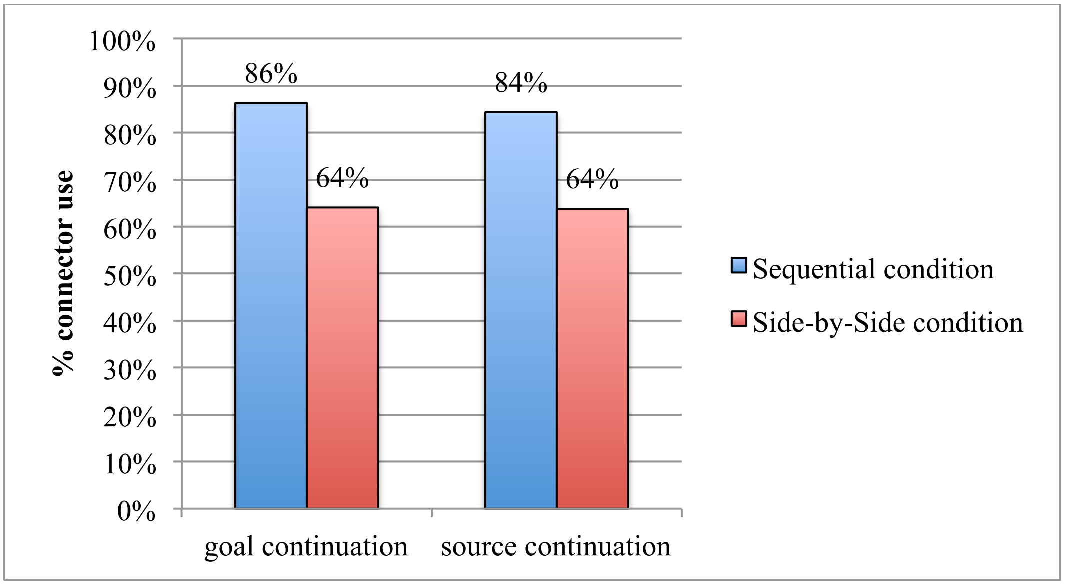

Figure 9. Experiment 1 connector use.

Table 1. Output from Experiment 1 statistical models.

Model	**Relational test**	**Item test**	**Test difficulty**	**Pronouns/Zeros**	**Connectors**
Encoding condition	-0.00034 (0.00138)	0.00309 (0.00244)	-0.00706 (0.07381)	-0.5698 (0.6131)	-1.5467 (0.3596)*
Thematic role	-0.00009 (0.00138)	-0.00043 (0.00244)	-	0.8862 (0.4341)*	0.1529 (0.3376)
Encoding condition* Thematic role	0.000067 (0.00277)	-0.00084 (0.00488)	-	-0.2566 (0.3686)	-001366 (0.4512)
Connector use	-	-	-	2.3236 (0.3236)***	-

Fixed effects reported: Coefficient (Standard Error). Reference condition for Encoding condition is Side-by-Side. Reference condition for Thematic role is goal. Reference value for connector use refers to connector was used for that trial. All predictors were grand-mean centered.

***p< .001. **p< 0.01. *p< 0.05. †p<0.10.

Experiment 1 Discussion

These results were unexpected: the Sequential group used numerically more reduced forms than the Side-by-Side group, and significantly more temporal connectors. This is the opposite of what I had predicted in terms of differences between the groups, and suggests that the Sequential group actually had more connected representations of the events in their memory than the Side-by-Side group. Additionally, all participants performed at ceiling on both memory tests. This suggests that the memory questions were too easy, and the difficulty measure in the post-experiment questionnaire supported this conclusion.

The main theoretical motivation for this set of experiments was to test the hypothesis that speakers use discourse coherence markers like reduced forms and connectors because the linguistic context makes certain representations more accessible, but also because the speaker has representations of these events in their memory that are more connected and temporally related. We've seen variation in the use of these coherence markers in previous experiments, so this study was designed to directly test the hypothesis that the degree of relational encoding affects the use of coherence markers in speech.

One possibility for these results is that the effects of pronouns and connectors were perhaps really due to memory differences, but that the encoding manipulation didn't have the expected effect. A post-hoc explanation for the inverse group effect found here is that the Sequential group actually created connected representations of the events during the initial picture viewing because of an "animation effect" occurring with the stimuli. The Sequential group viewed a PowerPoint presentation that was set up to display each picture at the center of the screen for 2.5 seconds, and then automatically advanced to the next picture. Upon viewing this presentation more closely, I noticed that the images replace each other in such a way that it looks like a 2-frame animation is playing for each pair of events. For example, the first stimulus item shows the Duchess handing a painting to the Duke in the first picture, and then the Duke throwing the painting into a closet in the second picture (Figure 5). The background room that they are in stays exactly the same, so when the two pictures are presented overlaying quickly, it looks like the Duke is performing one action of receiving the painting and then immediately throwing it away. Even though there is no discernable difference between each item pair, the animation-like action and the consistency between backgrounds makes the shift between items more noticeable.

This explanation is speculative. In Experiment 2, I empirically tested this idea that the animation effect increases relational encoding and use of coherence markers. In short, I altered the relational encoding manipulation such that the encoding task is even more distinct between the groups, and improved the memory test so that it encapsulates both coherence marker production and more difficult relational/item memory measures.

EXPERIMENT 2

The group manipulation was changed such that participants in the **Paired condition** viewed each picture sequentially for 2.5 seconds each (in correct story order), but with a crosshair slide between each item pair (between every 2 pictures). Participants in this group were also asked after every item pair if the same character performed both actions in the two pictures (yes/no button press). I predict that the combination of the animation effect for pairs of pictures and the questions that focus on character continuation will encourage this group to encode the events with higher degree of connection. Participants in the **Singular condition** viewed each picture sequentially for 2.5 seconds each (in correct story order), but there was a crosshair slide after every picture. Participants in this group were asked after every picture if there was a murder weapon in the picture (yes/no button press). I predict that the temporally and visually separated pictures will reduce the "animation effect", and that the questions focus on semantic detail awareness within each picture, rather than event connectedness between the pictures, and encourage this group to encode the events as individual items. This initial viewing task was coded in E-Prime, to better control for presentation timing and to collect button responses.

In order to test this manipulation, I designed two new memory tests:

(1) Continuation task: (for all items, in story order) Participants were shown the first picture of each pair, and asked, "what happened in the following picture?" Participants were asked to type in a sentence describing what they remember happened in the second picture of each item. One critical difference here is that

this is no longer a spoken language task, but language production can still be measured. This task served to test two hypotheses: 1) the Paired condition will have better memory accuracy, and 2) the Paired condition will use more reduced forms and connectors on accurately remembered trials.

(2) Change Detection task: (for all items, in story order) Participants were shown the second picture of each pair, and asked, "did anything change in this image? If so, what?" Some of the details were changed using Photoshop for half of the pictures in this task: an object changes color, an object is no longer present in the picture, or an object shifts location within the picture.

My main prediction is that if the encoding manipulation affects event relatedness representations, then participants in the Paired condition will perform better on both parts of this memory test. For the Continuation task, I hypothesize that this group will be more accurate at correctly describing the second picture because relational encoding is known to boost accuracy in associative recall tasks (Giovanello, Schnyer, & Verfaellie, 2004). Because they are typing in descriptions of the second event, I will also look at their use of pronouns and connectors, and predict to see a group difference here as well. For the Change Detection task, I predict that participants in the Paired condition will be more accurate at deciding whether the details in a picture have changed, because relational encoding has been shown to boost both relational and item memory test accuracy (Staresina, Gray, & Davachi, 2009).

Methods

Participants

Data from 54 native English participants were collected. Participants were UNC undergraduates in the Spring of 2018 and received SONA subject pool credit for participation. One criterion for a participant to be included in the analyses was at least 25% accuracy on the Continuation task, to ensure that all participants engaged in the task as expected and so that analyses of language production could be robust (Bell et al., 2010). 14 participants were excluded: 9 had <25% accuracy in the Continuation task, 2 were shown an old version of instructions before they were updated, 2 had a technology failure in the encoding task, and 1 had a technology failure in the Continuation task. This left 40 participants to be analyzed: 20 in the Singular condition and 20 in the Paired condition.

Materials and Procedure

Stimuli were the same as Experiment 1. Participants were briefly introduced to the task by an experimenter, and they filled out a brief demographics survey. Participants first watch an introduction slideshow of illustrated pictures with narration that describes the murder mystery story and the participant's role as a tabloid photographer, introduces the 6 characters and the location (manor house), and explains that the participant will meet with a detective to help solve the crime (in Microsoft PowerPoint).

<u>Encoding task</u>: All participants were then told that they will view the pictures individually and the screen will advance automatically. Participants in the Paired condition were told that after every two pictures, a question would appear asking if the same character performed both actions in the previous two pictures, and response is by Y/N button press. An

example item was shown with a correct response explained. Then each picture in an item appeared on the screen for 2.5 seconds each, and then the question screen appeared between item pairs (times out at 4000ms), and then a fixation screen appeared for 1000ms between pairs. Participants in the Singular condition were told that after every picture, a question would appear asking if there was a murder weapon shown in the previous picture, and response is by Y/N button press. An example item was shown with a correct response explained. Then each single picture appeared on the screen for 2.5 seconds, and then the question screen appeared after every picture (times out at 4000ms), and then a fixation screen appeared for 1000ms between pictures. See Figure 10 for an example of these procedures.

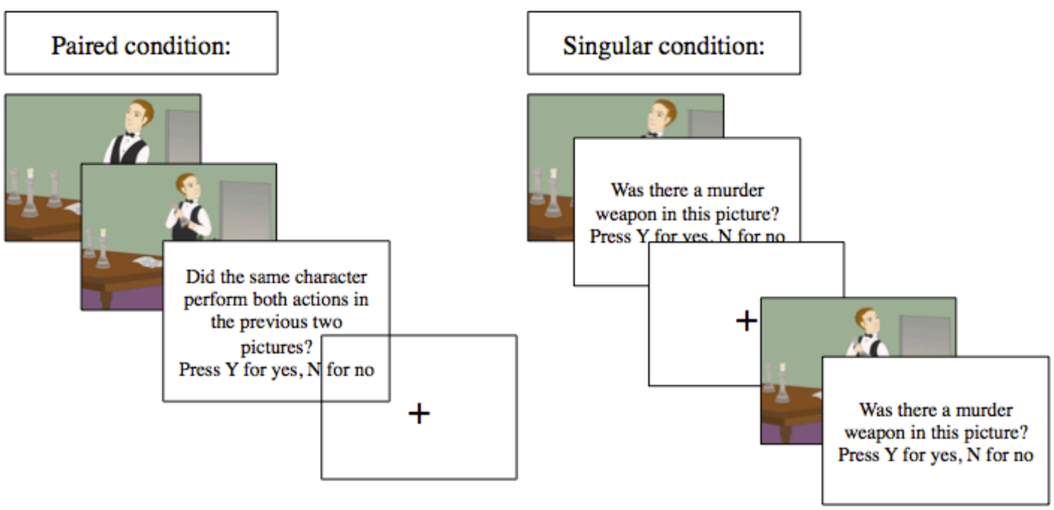

Figure 10: Encoding task screen procedures for the Paired condition and the Singular condition.

<u>Continuation task</u>: This task was the same for all participants. Through written instructions they were told: “you will see a selection of pictures one by one, presented in the same order as the initial viewing, but you’ll only see every other picture. On each page you will see a picture on the screen with a text box under it. You’ll listen to the detective describe the

picture. Then you will type in your description of what you remember the FOLLOWING picture was about. Do your best to remember what you saw in the previous task, and try to make your description sound like a story continuing from the detective's sentence. Imagine that you are going to say aloud your description, but type it instead. Press ENTER when you are finished typing to submit your response and continue to the next picture." They were shown an example trial, with an example description of the following picture provided. For each item, only the first event picture was shown, and the participants had to describe what they remember the second event picture was about. The detective sentences were played over headphones, and between each question there was a fixation screen for 1000ms.

<u>Change Detection task</u>: This task was also the same for all participants. Through written instructions they were told: "Now you will view a selection of pictures one by one, presented in the same order as the initial viewing, but you'll only see every other picture. Your task is to decide whether each image has any changes in it, or if it is the same as the image you saw in the first task. There are three possible types of change: Color – an item in the picture has changed color, Deletion – an item in the picture is now missing, or Shift – an item in the picture has moved positions. Half of the items will have a change, and half will not. If you think an item has a change, you will also be asked to describe the change after selecting one of the three types. After you're finished selecting and typing, press ENTER to move on." They are shown an example trial, with a Color change provided and a description provided on the following screen. Between each item was a fixation screen for 1000ms, then a picture was presented on the screen with 4 multiple choice options presented below it: Color, Deletion, Shift, None. Once one of these was selected, the screen advanced to show a text box under the picture with instructions to describe the change (or press ENTER if None), and then moved on to the next trial. The pictures

with changes were created in Adobe Photoshop by editing one of the layers of the original picture. Some changes were likely smaller/harder to detect than others, so this test was piloted by 4 naïve participants. Overall accuracy was 84%, with Deletion items at 50%, Color items at 83%, Shift items at 83%, and items with No Change at 95%. See Figure 11 for an example of this procedure, and Figure 12 for examples of the three change types. Then participants took a post-experiment questionnaire that asked about their knowledge of the story and their strategies in the each of the three tasks.

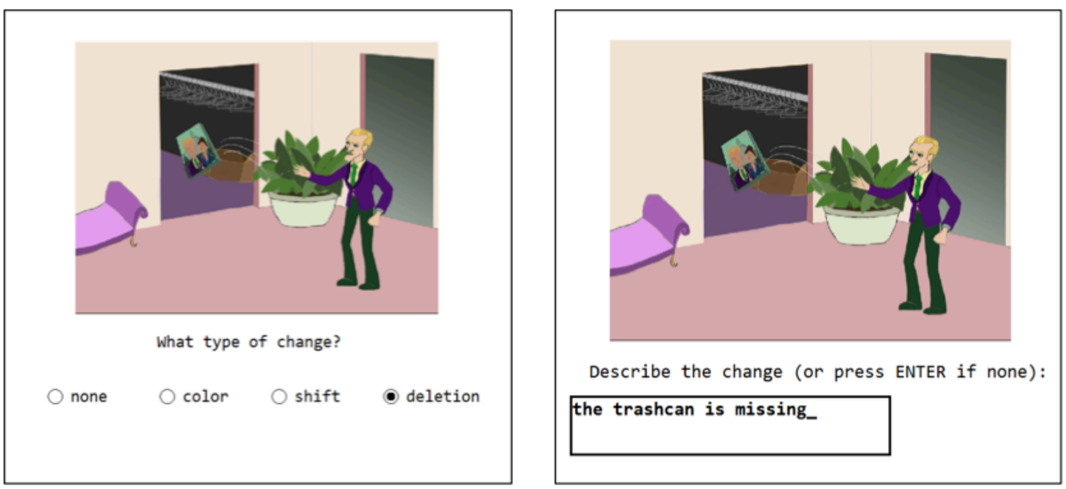

Figure 11. Change detection task screen procedure.

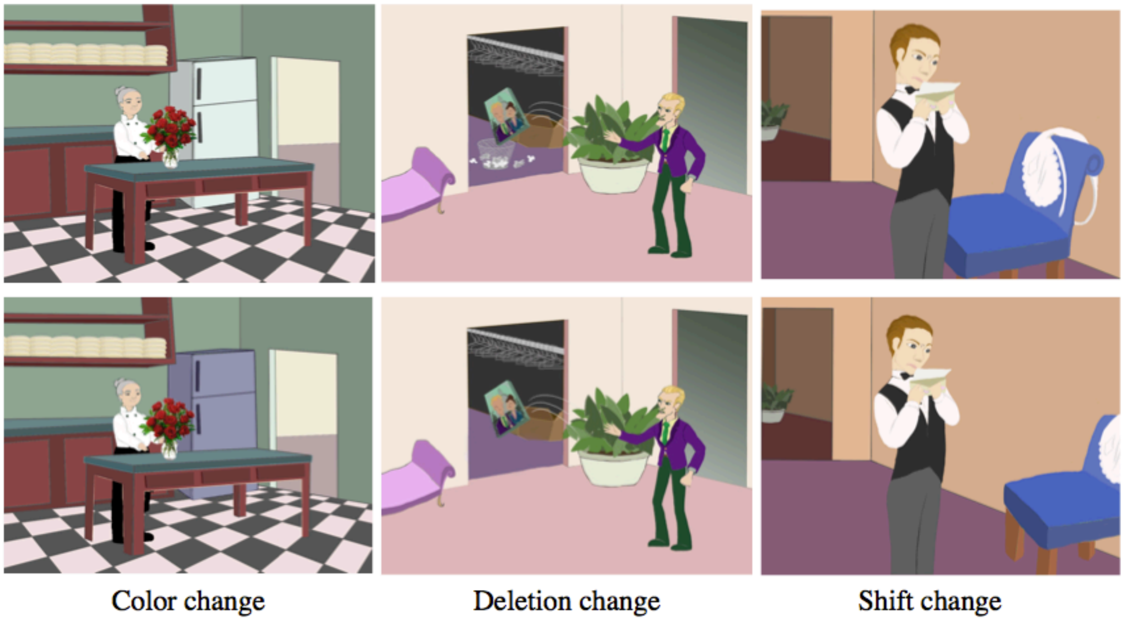

Figure 12. Examples of change types.

Measures and Analytic Approach

Accuracy was recorded for the Encoding task for both condition groups by the E-Prime tasks. For the Paired group, participants received either a 0 or 1 on the "same action" question. For the Singular group, participants received either a 0 or 1 on the "murder weapon" question for both pictures in an item, resulting in two accuracy scores for each item. The accuracy coding from each picture for the Paired group were averaged, and then combined with the average accuracy coding from each singular picture for the Singular group to create one by-item Encoding task accuracy variable, treated as continuous.

Accuracy was recorded for the Continuation task by a pair of trained research assistants who read through the typed responses and noted whether the description was accurate or not. If there were any questions about certain responses, these were flagged and I reviewed them. I then went through all of the responses and coded whether the response begins with reference to the correct target character or not, and whether the response describes the correct event or not. For a trial to have accurate memory recall in this task, the correct target must be referred to and the correct event must be described. Then I compared this combined accuracy score with the initial coded measures, and reviewed any discrepancies.

Research assistants also coded the responses for reference form choice, indicating if a pronoun, zero, or descriptive name was used; and for connector use, indicating whether one of the types of connectors was used.

Accuracy for the Change Detection task was recorded by the E-Prime program. If a participant responded with the correct change type, this measure was coded as 1. Then I went through the typed descriptions of the changes to check that the correct change was in fact described. If the response was different than the actual change, then accuracy was changed to 0.

In the post-experiment questionnaire, participants were asked to rate the difficulty of each of the three tasks on a 1-5 scale (1 being Very Easy, 5 being Very Difficult). These measures were treated as continuous in subsequent analyses.

Generalized linear mixed effects models were used to analyze the data, in order to account for the dependencies in this repeated measures design. SAS Proc Glimmix was used to analyze dichotomous outcomes with a logit link (reference form, connectors, accuracy on Continuation and Change Detection memory tests). Models of each dependent variable were constructed with random intercepts of participant and item to account for these differences. Effects coding was used for all binary predictors. Random slopes of primary predictors by participant and by item were included when appropriate (if they improved model fit), and excluded if estimated to be zero (Searle et al., 1992). The overall difficulty scores for each memory test must be each participant's singular score (1-5), so these were analyzed by linear regression using SAS Proc Mixed (no random effects). SAS Proc Mixed was also used for the Encoding task accuracy variable, since this was treated as continuous.

Primary predictors of interest were encoding condition (Paired/Sequential), thematic role continuation (goal/source), reference form type (reduced/explicit), and connector use (used/not). Interactions between these predictors were increasingly added to each model, and only retained in each final model if nearly significant at $|t| > 1.5$. All predictors were grand-mean centered.

Results

897 items were included in the following analyses. 87 critical items were excluded: 14 had no response typed in, 6 had no grammar in their response (shorthand), 6 had ambiguous pronouns whose referent could not be determined, 7 referred to an item instead of a character, 11 began with a preposition, 26 began with reference to both characters by "they", 1 began with reference to both character by "the two of them", 6 referred to both characters by mentioning both by name "X and Y". 1 began with "that", 1 began with "where", 1 began with "who", and 7 began with "which". See Tables 2 and 4 for summaries of the model outputs, and see the Appendix for more details about each model.

Does relational encoding manipulation affect memory accuracy?

My first question was to see if the encoding manipulation was successful by analyzing the accuracy measure for each task.

For the Encoding task, there was a significant difference between the two groups such that the Singular group was more accurate than the Paired group. There was no effect of thematic role, and no interaction. See Table 2 for model output. Importantly, this just shows that the

Singular encoding task was easier than the Paired encoding task, not that there was a difference in encoding: Participants who were asked if a murder weapon appeared in the picture were more accurate (92%) than participants who were asked if the same character performed both actions (54%). There was no significant difference between the groups for the by-participant difficulty score for this task: average response was 3.23 for the Paired condition, and 3.21 for the Singular condition.

For the Continuation task, there was a significant effect of encoding condition such that the Paired condition had higher accuracy (48%) in their recalled responses than the Singular condition (40%). There was also a significant main effect of thematic role such that accuracy was higher for goal continuation items, suggesting that the inherent connectivity of goal continuation events is reflected in memory representations. There was a marginal interaction between these predictors ($p=0.089$) (see Figure 13 and Table 2). There was also a significant difference between the groups for the by-participant difficulty score for this task such that participants in the Singular condition rated it as more difficult than participants in the Paired condition (4.00 and 3.49 respectively). This patterns with the actual accuracy measure: the Singular group rated this task as significantly harder, and they also had significantly lower accuracy.

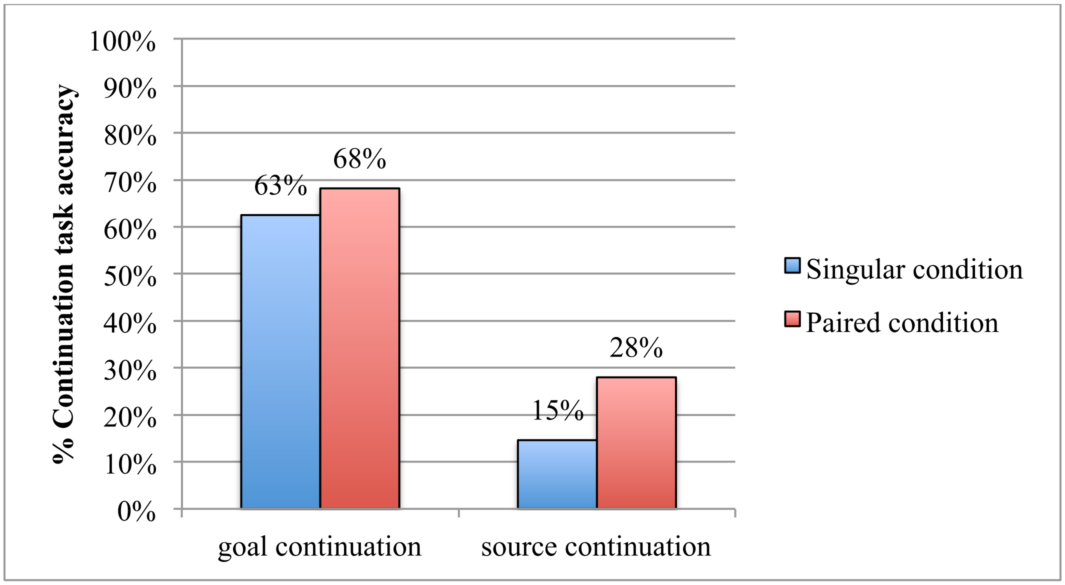

Figure 13. Experiment 2 Continuation task accuracy.

For the Change Detection task, there was no significant difference between the encoding conditions on this measure of accuracy (55% for Paired, 54% for Singular). There was also no effect of thematic role, nor an interaction between the two. There was a main effect of change type such that the "No Change" type had the highest accuracy; the "Color" change type was not significantly different from "No Change", but both "Deletion" and "Shift" were significantly different than "No Change". See Figure 14 and Appendix. There was a significant difference between the encoding conditions on the by-participant difficulty ratings for this task, such that participants in the Paired condition rated it as more difficult than participants in the Singular condition did (4.31 and 4.10 respectively).

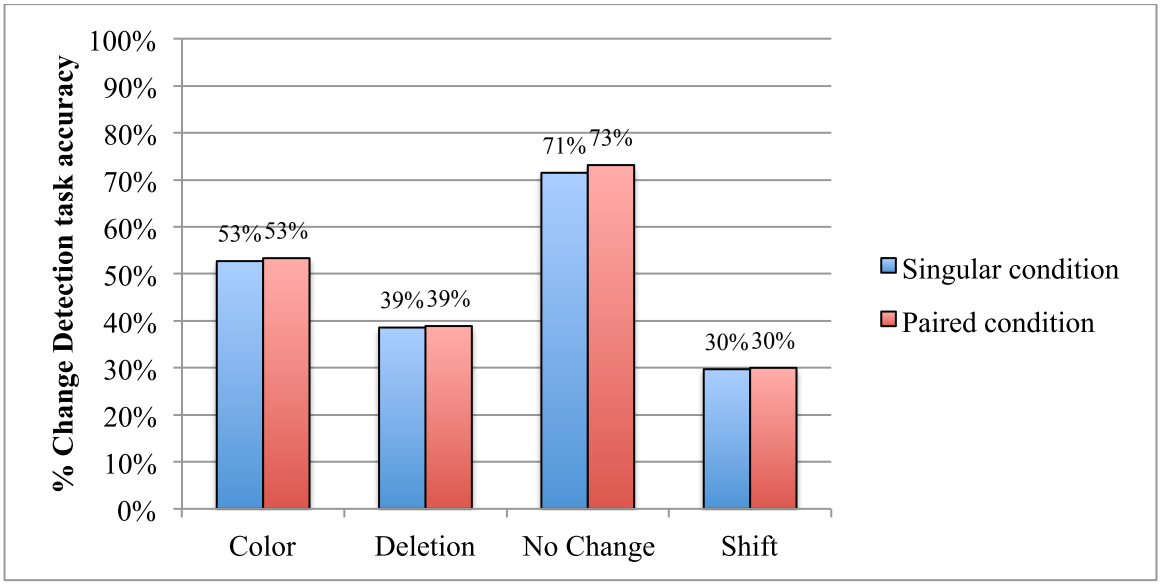

Figure 14. Experiment 2 Change Detection task accuracy.

Table 2. Output from Experiment 2 memory accuracy and difficulty statistical models.

Model	**Encoding task accuracy**	**Encoding task difficulty**	**Cont. task accuracy**	**Cont. task difficulty**	**Change Detection task accuracy**	**Change Detection task difficulty**
Encoding condition	-0.3794 (0.03215)***	0.0147 (0.06944)	0.5832 (0.1968)**	-.05047 (0.0554)***	0.1102 (0.275)	0.2145 (0.0466)***
Thematic role	-0.01636 (0.05048)	-	2.3679 (0.4794)***	-	0.989 (0.6855)	-
Encoding condition * Thematic role	0.05905 (0.1077)	-	-0.6174 (0.3548) †	-	0.08089 (0.4631)	-

Fixed effects reported: Coefficient (Standard Error). Reference condition for Encoding condition is Paired. Reference condition for Thematic role is goal. All predictors were grand-mean centered. ***p< .001. **p< 0.01. *p< 0.05. †p<0.10.

Does relational encoding manipulation affect reference form choice?

Now that I know that the relational encoding manipulation was successful, the next question is to see if the encoding manipulation affects pronoun/zero use in the typed responses of the Continuation task. For this analysis, I excluded all trials where participants' responses were incorrect on the Continuation task accuracy measure, because I only want to examine participants' referential form use when their recall of the event is correct (correct target is referred to and correct event is described). Thus, for the following analyses only 393 trials were included (see Table 3).

Table 3. Number of trials by encoding condition and Continuation task accuracy.

	Continuation task: correct	Continuation task: incorrect	Total
Singular Condition	182	277	459
Paired Condition	211	227	438
Total	393	504	897

There was a marginal effect of encoding group on pronoun/zero use, however it was in the opposite direction than I predicted: the Singular group used more pronouns/zeros (49%) than the Paired group (36%). There was no effect of thematic role on pronoun/zero use, nor an interaction between these predictors. There was a significant effect of connector use on pronoun/zero use such that more pronouns/zeros were used on trials where a connector was also used (see Figure 15 and Table 4).

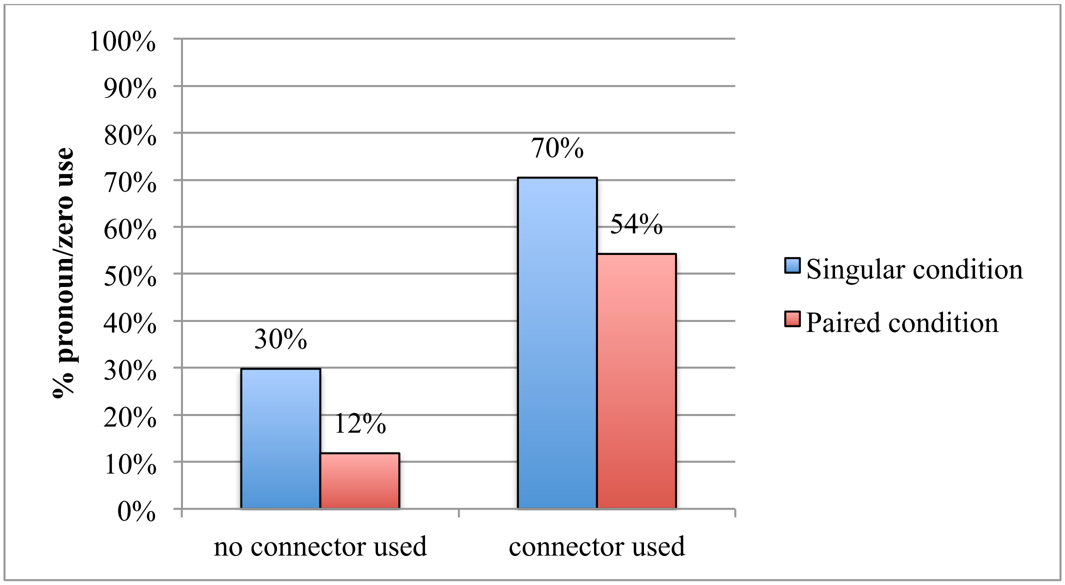

Figure 15. Experiment 2 pronoun/zero use for only accurately recalled trials.

Does relational encoding manipulation affect connector use?

My next question was to see if the encoding manipulation affected connector use in the typed responses of the Continuation task, again only including accurately recalled trials. There was no main effect of encoding group on connector use (56% for Paired condition, 48% for Singular condition), there was no main effect of thematic role, and there was no significant interaction between these predictors (see Figure 16, Table 4).

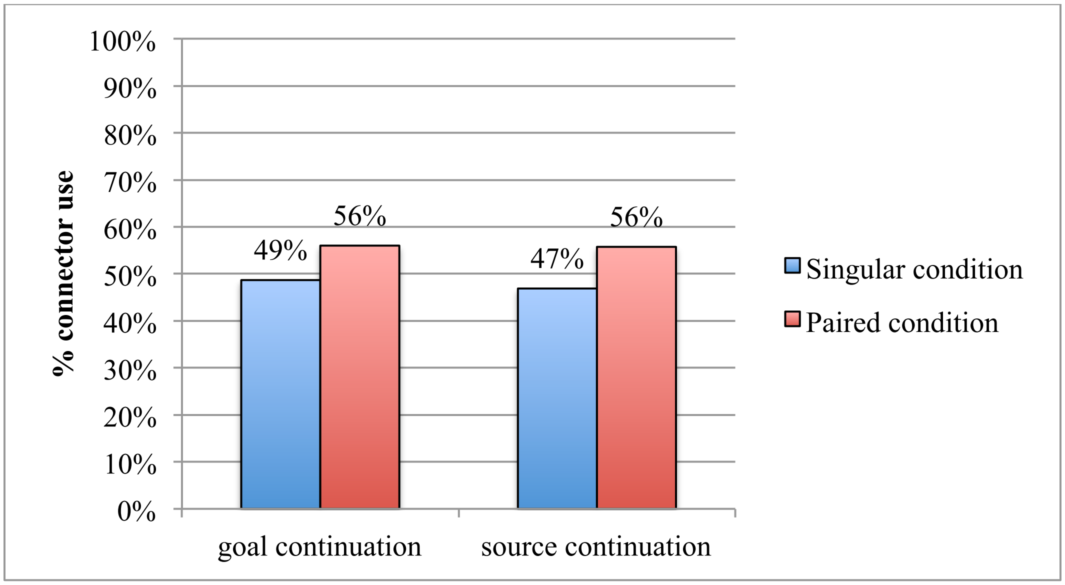

Figure 16. Experiment 2 connector use for only accurately recalled trials.

Table 4. Output from Experiment 2 reference form choice and connector statistical models.

Model	**Pronoun/Zero**	**Connector**
Encoding condition	-1.5652 (0.7843)†	0.13 (1.765)
Thematic role	0.5343 (0.5476)	0.3432 (0.6027)
Encoding condition*Thematic role	0.6981 (0.8325)	-0.02474 (1.0466)
Connector use	2.0243 (0.5587)***	-

Fixed effects reported: Coefficient (Standard Error). Reference condition for Encoding condition is Paired. Reference condition for Thematic role is goal. All predictors were grand-mean centered. ***p< .001. **p< 0.01. *p< 0.05. †p<0.10.

Experiment 2 Discussion

In sum, Experiment 2 found that the encoding manipulation does affect event relatedness representations, but that this manipulation does not have a direct effect on linguistic coherence marker use. Participants in the Paired condition who initially answered questions about characters in pairs of events had better accuracy when given the first event and asked to recall the second event. This shows support for a difference in event relatedness representations that the two groups made in the Encoding task. However, this experiment failed to show support for the proposal that the encoding manipulation would affect language production: there was no significant effect of encoding group on either pronoun/zero use or connector use. In fact, there was a marginal effect on pronoun/zero use but in the opposite direction than predicted. There was also no group difference in the Change Detection task, indicating that both groups noticed changes in single event details equally well. Speculative post-hoc explanations of these findings are discussed in the General Discussion.

GENERAL DISCUSSION

Both Experiments 1 and 2 found that the relational encoding manipulation did not affect discourse coherence marker use in the predicted ways. In Experiment 1, the encoding manipulation did not have the expected effect on memory or language production. In Experiment 2, the encoding manipulation did affect event relatedness representations (as measured by event recall accuracy), but this did not have an effect on reduced referential form or connector use on accurately recalled trials.

There are two possible implications from these results: either representations of event relatedness don not affect coherence marker use, or they do but this experiment failed to show support. Since the model proposed in the Introduction was supported by existing empirical evidence (Zerkle & Arnold, 2016; Arnold & Nozari, 2017), this particular study failed to find support for the model, but the model may still hold true. Below, I describe possible speculative reasons why this particular set of experiments did not entirely show support for the proposed model of reference production.

One potential problem with the Continuation task is that on some trials, participants typed in descriptions that did not match the current trial, but instead described another trial; in other words, false alarms that were ultimately coded as incorrect. This was not an expected finding nor a planned analysis, but it is interesting that some participants had “accurate” memory for certain events, but did not connect them with their respective pair correctly. This could be due in part to the nature of the paradigm: there are only 6 characters involved, and they each interact with each other several times throughout the story and with specific items multiple times. One possible future analysis would be to see if the rate of these false alarms is different for the different relational encoding groups. Because the Continuation task was difficult, this limited the number of correct trials to use in the linguistic analyses. Only 44% of all items were accurately recalled (see Table 3), creating constraints (e.g., low power) on the mixed effects models used to analyze pronoun/zero use and connector use. Because of these limitations, it's possible that the predicted effects do exist, but are not strong enough to detect in the current data set.

Thematic role did affect Continuation task accuracy, suggesting that goal continuations were inherently more related than source continuations. This can be taken as evidence that the natural connectivity of goal continuation stories is reflected in memory representations, as predicted in my proposed model (Figure 3). This has been shown to be the case: Rosa & Arnold (2017) reported that goal continuation events were perceived as more related to the context than the source continuation events in a relatedness rating study using the same picture story paradigm. Here, thematic role did affect reduced form use in Experiment 1, replicating previous findings. This paradigm was originally designed to test effects of thematic role on reference production, which may have created additional noise when examining other effects (such as the induced relational encoding manipulation).

My original question for this study was to see whether the strength of event coherence in memory can influence measurable changes in language production, specifically coherence marker use. While these two experiments did not find direct evidence supporting this, there may be other methods worth investigating. I think that a new paradigm that does not manipulate thematic role differences would be good to test, because we already know that the inherent relatedness in goal continuation events affects reference form choice. Eliminating this and instead directly testing the effect of relational encoding on coherence marker production may produce the predicted results. Perhaps the sentence structures should be less constrained, so that there is little inherent relatedness between event pairs; for example, “Ryan went to the library with John. Ryan/He checked out a new novel.” Further development of a new paradigm is needed to fully investigate the intended manipulation. Another potential change to the current design would be to make the Continuation task easier, so that more data is available for analysis. The manipulation in Experiment 2 has already been shown to be successful in that it affects memory accuracy in the expected ways, so this type of memory test is not strictly necessary in future iterations of this research question. (However, it was not an overwhelmingly big effect, so it was likely dampened by other aspects of this particular design.) Instead, both events could be presented during the Continuation task (as in Experiment 1), to ease memory load and increase the rate of correctly recalled descriptions. This task could also require spoken descriptions instead of typed, so that language production is as naturalistic as possible.

In sum, the current studies did not find direct support for the prediction that the strength of event encoding affects how people subsequently talk about the events. Inducing differences in memory encoding strength proved to be a difficult procedure as shown in Experiment 1, but ultimately was a successful manipulation in Experiment 2. However, this did not significantly affect language production, specifically coherence marker use, in the predicted ways. Future research may be needed to determine whether relational encoding does in fact affect language production in other types of situations.

APPENDIX: EXPERIMENT 1 AND EXPERIMENT 2 MODEL OUTPUT

Experiment 1 model output

Model of Relational test accuracy:

Effect	**Estimate**	**SE**	**DF**	**t-value**	**p-value**
Encoding condition	-0.00034	0.00138	950	-0.24	0.808
Thematic role	-0.00009	0.00138	950	-0.07	0.947
Encoding condition*Thematic role	0.000067	0.00277	950	0.02	0.981

Reference condition for Encoding condition is Side-by-Side. Reference condition for Thematic role is goal. All predictors were grand-mean centered. No random intercepts, because Relational test accuracy is a by-participant measure.

Model of Item test accuracy:

Effect	**Estimate**	**SE**	**DF**	**t-value**	**p-value**
Encoding condition	0.00309	0.00244	950	1.27	0.205
Thematic role	-0.00043	0.00244	950	-0.17	0.862
Encoding condition*Thematic role	-0.00084	0.00488	950	-0.17	0.864

Reference condition for Encoding condition is Side-by-Side. Reference condition for Thematic role is goal. All predictors were grand-mean centered. No random intercepts, because Item test accuracy is a by-participant measure.

Model of test difficulty:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	-0.00706	0.07381	952	-0.1	0.924

Reference condition for Encoding condition is Side-by-Side. All predictors were grand-mean centered. No random intercepts, because test difficulty is a by-participant measure.

Model of pronouns/zeros use:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	-0.5698	0.6131	20	-0.93	0.3637
Thematic role	0.8862	0.4341	35	2.04	0.0488
Encoding condition*Thematic role	-0.2566	0.3686	791	-0.7	0.4866
Connector use	2.3236	0.3975	41	5.85	<.0001

Reference condition for Encoding condition is Side-by-Side. Reference condition for Thematic role is goal. All predictors were grand-mean centered. Random intercepts of participant and of item. Random slopes of connector and thematic role by participant and of encoding condition by item were all estimated to be zero.

Model of connector use:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	-1.5467	0.6728	20	-2.3	0.0324
Thematic role	0.1529	0.3376	39	0.45	0.6532
Encoding condition*Thematic role	-0.1366	0.4512	824	-0.3	0.7621

Reference condition for Encoding condition is Side-by-Side. Reference condition for Thematic role is goal. All predictors were grand-mean centered. Random intercepts of participant and of item. Random slope of thematic role by participant was included, but random slope of encoding condition by item was estimated to be zero.

Experiment 2 model output

Model of Encoding task accuracy:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	-0.3794	0.05048	38.5	-7.52	<.0001
Thematic role	-0.01636	0.06697	36.6	-0.24	0.8084
Encoding condition*Thematic role	0.05905	0.1077	42.5	0.55	0.5863

Reference condition for Encoding condition is Paired. Reference condition for Thematic role is goal. All predictors were grand-mean centered. Random intercepts of participant and of item. Random slopes of thematic role by participant and of encoding condition by item were included.

Model of Encoding task difficulty:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	0.0147	0.06944	895	0.21	0.8324

Reference condition for Encoding condition is Paired. All predictors were grand-mean centered. No random intercepts, because Encoding task difficulty is a by-participant measure.

Model of Continuation task accuracy:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	0.5832	0.1968	41.23	2.96	0.005
Thematic role	2.3679	0.4794	19.47	4.94	<.0001
Encoding condition*Thematic role	-0.6174	0.3548	44.09	-1.74	0.0888

Reference condition for Encoding condition is Paired. Reference condition for Thematic role is goal. All predictors were grand-mean centered. Random intercepts of participant and of item. Random slope of thematic role by participant was included, but random slope of encoding condition by item was estimated to be zero.

Model of Continuation task difficulty:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	-0.5047	0.05535	895	-9.12	<.0001

Reference condition for Encoding condition is Paired. All predictors were grand-mean centered. No random intercepts, because Continuation task difficulty is a by-participant measure.

Model of Change Detection task accuracy:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	0.1102	0.275	29.82	0.4	0.6914
Thematic role	0.989	0.6855	20.41	1.44	0.1643
Encoding condition*Thematic role	0.08089	0.4631	24.16	0.17	0.8628

Reference condition for Encoding condition is Paired. Reference condition for Thematic role is goal. All predictors were grand-mean centered. Random intercepts of participant and of item. Random slopes of thematic role by participant and of encoding condition by item were included.

Model of Change Detection test accuracy by type of change:

Effect	Estimate	SE	DF	t-value	p-value
Change type: none (reference)	1.0689	0.4507	18.87	2.37	0.0285
Change type: color	-1.0773	0.8707	18.69	-1.24	0.2313
Change type: deletion	-2.0156	0.8642	18.23	-2.33	0.0313
Change type: shift	-2.1101	0.7902	17.61	-2.67	0.0158

Random intercepts of participant and of item. Random slope of change type by participant was included.

Model of Change Detection task difficulty:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	0.2145	0.04664	895	4.6	<.0001

Reference condition for Encoding condition is Paired. All predictors were grand-mean centered.

No random intercepts, because Change Detection task difficulty is a by-participant measure.

Model of pronouns/zeros use:

Effect	Estimate	SE	DF	t-value	p-value
Encoding condition	-1.5652	0.7843	39.22	-2	0.0529
Thematic role	0.5343	0.5476	23.66	0.98	0.3391
Encoding condition*Thematic role	0.6981	0.8325	388	0.84	0.4023
Connector use	2.0243	0.5587	164.9	3.62	0.0004

Reference condition for Encoding condition is Paired. Reference condition for Thematic role is goal. All predictors were grand-mean centered. Random intercepts of participant and of item. Random slopes of connector and thematic role by participant and of encoding condition by item were all estimated to be zero.

Model of connector use:

Effect	**Estimate**	**SE**	**DF**	**t-value**	**p-value**
Encoding condition	0.13	1.1765	36.93	0.11	0.9126
Thematic role	0.3432	0.6027	22.09	0.57	0.5748
Encoding condition*Thematic role	-0.02474	1.0466	44.8	-0.02	0.9812

Reference condition for Encoding condition is Paired. Reference condition for Thematic role is goal. All predictors were grand-mean centered. Random intercepts of participant and of item. Random slope of thematic role by participant was included, but random slope of encoding condition by item was estimated to be zero.

REFERENCES

Addis, D. R., Giovanello, K. S., Vu, M. A., & Schacter, D. L. (2014). Age-related changes in prefrontal and hippocampal contributions to relational encoding. *NeuroImage*, *84*, 19-26.

Almor, A. (2000). Constraints and mechanisms in theories of anaphor processing. *Architectures and mechanisms for language processing*, 341-354.

Ariel, M. (1990). *Accessing Noun Phrase antecedents.* London: Routledge.

Ariel, M. (1996). Referring Expressions and the +/- coreference distinction. In T. Fretheim, & J.K. Gundel (Eds.), *Reference and Referent Accessibility*, *38*, 13.

Ariel, M. (2001). Accessibility theory: An overview. *Text representation: Linguistic and psycholinguistic aspects*, *8*, 29-87.

Arnold, J. E. (1998). *Reference form and discourse patterns* (Doctoral dissertation, Stanford University).

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and cognitive processes*, *23*(4), 495-527.

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187-203.

Arnold, J. E. (2016). Explicit and emergent mechanisms of information status. *Topics in cognitive science*, *8*(4), 737-760.

Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of memory and language*, *56*(4), 521-536.

Arnold, J. E., & Nozari, N. (2017). The effects of utterance timing and stimulation of left prefrontal cortex on the production of referential expressions. *Cognition*, *160*, 127-144.

Bell, B., Morgan, G. B., Schoeneberger, J. a, Loudermilk, B. L., Kromrey, J. D., & Ferron, J. M. (2010). Dancing the Sample-Size Limbo with Mixed Models: How Low Can You Go? *SAS Global Forum 2010*, 1–11.

Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive processes*, *10*(2), 137-167.

Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987, July). A centering approach to pronouns. In *Proceedings of the 25th annual meeting on Association for Computational Linguistics* (pp. 155-162). Association for Computational Linguistics.

Brown-Schmidt, S., Byron, D. K., & Tanenhaus, M. K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language*, *53*(2), 292-313.

Chafe,W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25?56). New York: Academic Press.

Chafe, W. L. (1994). *Discourse, consciousness, and time*. Chicago: Chicago University Press.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 761-772.

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the cognitive science society* (Vol. 30, No. 30).

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998-998.

Fraser, B. (1999). What are discourse markers?. *Journal of pragmatics*, *31*(7), 931-952.

Giovanello, K. S., Schnyer, D. M., & Verfaellie, M. (2004). A critical role of the anterior hippocampus in relational memory: Evidence from an fMRI study comparing associative and item recognition. *Hippocampus*, *14*(1), 5–8.

Giovanello, K. S., & Dew, I. T. (2015). Relational memory and its relevance to aging. *The Wiley handbook on the cognitive neuroscience of memory*, 371.

Givon, T. (1983). Topic continuity in discourse: An introduction. In T. Givon (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1-42). Amsterdam: John Benjamins Publishing.

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, *21*(2), 203-225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions. *Language, 69*, 274-307.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101-144.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar* (pp. 172-177). Stanford, CA: CSLI publications.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of semantics*, *25*(1), 1-44.

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and cognitive processes*, *16*(2-3), 287-308.

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1-38.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, *19*, 849.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, *126*(2), 313-318.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic Press.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42* (1-3), 107-142.

Rosa, E. C. (2015). *Semantic role predictability affects reference form* (Doctoral dissertation). Chapel Hill, NC: University of North Carolina.

Rosa, E. C., & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language*, *94*, 43-60.

Schmitt, B. M., Meyer, A. S., & Levelt, W. J. (1999). Lexical access in the production of pronouns. *Cognition*, *69*(3), 313-335.

Searle S., Cassella G. & McCullouch C. (1992). Variance components. New-York: John Wiley-Sons.

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex*, 19(5), 1198–1207.

Tager-Flusberg, H. (1999). A psychological approach to understanding the social and language impairments in autism. *International review of psychiatry*, *11*(4), 325-334.

Wang, W. C., & Giovanello, K. S. (2016). The role of medial temporal lobe regions in incidental and intentional retrieval of item and relational information in aging. *Hippocampus*, *26*(6), 693-699.

Zerkle, S. A., & Arnold, J. E. (2016). Discourse attention during utterance planning affects referential form choice. *Linguistics Vanguard*, *2*(s1).

Zerkle, S.A., Rosa, E.C., Arnold, J.E. (2015). Do addressee gestures influence the effects of predictability on spoken reference form? Presented as a poster at the CUNY 2015 Conference on Sentence Processing, USC, Los Angeles, CA. March 19-21, 2015.

Zerkle & Arnold (under review). Does pre-planning explain why predictability affects reference production?