# Tracking the Common Ground in Dialogues: Cultural and Genre Effects

by

Keely Morrow

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Linguistics
University of Alberta

## Abstract

A socio-cognitive approach to language assumes language is multimodal, embodied in general cognition, and modulated by contextual cues (van Dijk, 2014). Research on situation models confirms that language is processed multimodally and experiences top-down influence from pre-existing knowledge in memory (Kurby & Zacks, 2015; Therriault & Rinck, 2012; Zwaan, 2014; 2015). However, most of this research has been done using written narratives and therefore does not account for contextual cues that may arise in spoken dialogue contexts. Furthermore, context and memory have been shown to modulate how participants interact with the common ground, which is always present in spoken dialogues (Colston in Kecskes & Mey, 2008; Gibbs & Colston in Giora & Haugh, 2017; Duran et al., 2011; Gann & Barr, 2014; Fukumura, 2014). The first goal of the present study was to investigate the contextual influence of spoken dialogue and genre on event-indexing factors (protagonists/objects, time, setting, cause/effect relationships, goals/plans of the protagonists) and to see if new factors were elicited in spoken dialogue. As participants incorporate these factors into their situation models, difference in individual pre-existing knowledge would result in unique representations for each participant and therefore modulate overlaps in shared knowledge. Thus, the second goal was to investigate how differences in pre-existing cultural knowledge affected participants' ability to establish common ground with each other. This was done by manipulating the linguistic and cultural background of the participants included in the study.

The present study analysed videos of 10 minute of spoken dialogues between 72 undergraduate students at the University of Alberta (36 pairs). Participants were either Canadian native English speakers (NS) or non-native international students (NN). They were divided into pairs of NS-NS, NS-NN, and NN-NN. Furthermore, participants were divided into either gossip

or summarizing task conditions. In the gossip conditions participants received the prompt "Discuss Donald Trump". In the summarizing condition they read two different articles about Donald Trump and summarized them to each other. The 6 resulting conditions (e.g. Gossip: NS-NS) were counterbalanced by genre and pairing, with 6 pairs in each condition. Videos were coded for instances of event-indexing factors or other factors produced with the same frequency; feedback cues indicating successful or successful establishment of common ground; and pragmatic structures or sequences indicating a reliance on the common ground. Analysis was done using logistic regression models. The results suggest speaker attitude is as important as previously established event-indexing variables in dialogues. Speaker-attitude, setting, and cause/effect relationships occurred significantly more in gossip conditions, whereas goals/plans occurred significantly more in summarising conditions. Both NN and NS gave more feedback indicating understanding when speaking to a NS than NN. NS were more successful at establishing referents in conversations. Congruent pairs (NN with NN-NN; NS with NS-NS) were more successful at reducing the amount of linguistic information required to convey a concept. In conclusion, this study shows the genre of discourse modulates the information tracked and cultural background of interlocutors modulates how that information is understood in a discourse.

## Preface

This thesis is an original work by Keely Morrow. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name “Collaboration in Gossip and Summarizing”, No. Pro00065764  September 22, 2016.

## Dedication

For my dad, who first kindled my scientific curiosity.

## Acknowledgements

I would like to thank my supervisors, Dr. Juhani Järvikivi and Dr. Herbert Colston for their enthusiasm and support when I chose this topic. And for the countless hours of feedback and guidance they have so selflessly provided. I would also like to thank Lauren Rudat and Vincent Poretta for their help arranging the equipment and space needed to run this experiment.

Table of Contents

## List of Figures

# 1. Introduction

A wealth of psycholinguistic and neurolinguistic research has shown that people create situation models of a discourse when processing it (Kurby & Zacks, 2015; Therriault & Rinck, 2012; Zwaan, 2014; 2015). A situation model is a mental simulation of both the implied and stated events described by a discourse. To build a situation model, people combine bottom-up linguistic processing with top-down information from memory, constantly updating and integrating incoming information with existing schemas (van Dijk, 2014; Kurby & Zacks, 2015; Zwaan 2015;). However, a caveat in the current literature is that most research has focused only on written narratives in discourse processing. Spoken dialogues and spontaneous speech have been virtually unexamined. Thus, it seems hasty to assume currently known aspects of situation models extend to all forms of discourse processing. Take, for example, one theory of situation model building: the event-indexing model. The event-indexing model assumes five major factors of an event are tracked by people and integrated into their mental representations: participants (sometimes split into protagonists and objects), goals of the participants, setting, time, and cause/effect relationships (Zwaan, 2014; 2015). These have been confirmed as important in written narrative processing (Kurby & Zacks, 2015; Therriault & Rinck, 2012; Zwaan, 2015). However, it seems that other forms of dialogue may weight these variables differently. For example, the role of time may be less important in technical writing than in news reporting.. Furthermore, in scenarios of face-to-face dialogue, where interlocutors are aware of whom they are speaking to, pre-existing common ground may alleviate the need to establish some event-indexing factors in the discourse, such as setting. The goal of the present study is thus to investigate if interlocutors in a spoken dialogue rely on the same information that narrative readers use to build a situation model of a discourse, and if event-indexing factors are equally important across different genres of dialogues, specifically, gossip vs. summarizing.

By focusing on spoken dialogues, it is important to consider certain aspects of this genre that separate it from narratives and have the potential to influence situation models. First, if the goal of dialogue is assumed to be is collaboration (Brennan, Galati, & Kuhlen, 2010), then monitoring the common ground may become equally important to creating a situation model. Interlocutors could begin to integrate feedback cues from each other into their understanding of a discourse and subsequently adjust how they add events in order to ensure mutual intelligibility. Furthermore, speaker attitude (i.e. the emotional affect an interlocutor adds to the discourse) may

become an influencing factor in the situation model and be tracked and integrated similarly to event-indexing factors.

A second aspect of dialogue involves participants having different prior knowledge when entering the situation. Their situation models will therefore have different top-down influences (van Dijk, 2014; Zwaan, 2014). This means two participants could have different understandings of the same discourse, impacting collaboration and the agreed upon discourse model established in the common ground (Brennan & Hanna, 2009; van Dijk, 2014; Mills, 2014). Therefore, subsequent goals of this study are to monitor which pragmatic cues are important for understanding a discourse and to describe how participants overcome differences in prior knowledge in order to collaborate. The following sections describe in more detail the previous literature on situation models and collaboration. They begin by discussing the influence of memory on an individual and move into how these influences are realized in instances of dialogue.

## 1.1 Mental Models

The sociocognitive approach, as described by van Dijk (2014), suggests different types of knowledge are represented in several different types of memory models, such as cultural, societal, contextual, and situation models. Here, knowledge is assumed to take the form of mental representations of people's perceptual experiences in the world, whether those experiences are their own or someone else's experience relayed through discourse. These memory models are built by abstracting schemas from multiple experiences and discourses. These abstractions are influenced by common knowledge accepted within an epistemic community and by personal opinions and first-hand experience, making each of these abstractions subjective and unique. They are embodied in general cognition and are constantly updated as people experience new situations and when people learn about others' experiences through discourse (van Dijk, 2014; Zwaan, 2015; see also Brown-Schmidt, 2012 for similarities to common ground updating). Since both common knowledge and individual knowledge are integrated into memory models, these models will be overlapping but unique across individuals (van Dijk, 2014).

One example of a model that is similar within a community, and which may thus represent common knowledge, is a cultural model. A cultural model reflects the cultural community, its expectations about the world and interpretations of knowledge, and in turn the expectations and interpretations derived from this community. Therefore, overlap in participants'

common knowledge can vary based on whether they share a cultural community or not. This study will pair participants together based on their linguistic and ethnographic backgrounds in order to manipulate the overlap in cultural models across interlocutors. Even within the same larger culture, smaller socio-cultural groups may still exist. Therefore, the present study assumes participants from the same community simply have more overlap in their cultural models than people from different communities, it does not assume members of the same community have the same cultural models.

### 1.1.1 Situation Models

Situation models are a type of mental model participants build when information is abstracted from a specific discourse (van Dijk, 2014, Zwaan, 2014; Zwaan, 2015). A situation model is the subjective representation of a discourse built on the explicit linguistic information participants read or hear as well and the implied propositions the text or speaker makes (Kurby & Zacks, 2015; Therriault & Rinck, 2012; Zwaan, 2014; Zwaan, 2015). These propositions represent an underlying knowledge, which listeners abstract a multimodal representation of the discourse from (van Dijk, 2014; Kurby & Zacks, 2015; Zwaan, 2015). The construction of mental models is embodied in general cognition (van Dijk, 2014; Zwaan, 2014). Thus, situation models are constrained by top-down influences from previous knowledge found in memory models (van Dijk, 2014; and discussed above in the 'Mental models' section). Situation models serve to give the listener a globally coherent understanding of the discourse (Zwaan, 2015).

A key component of situation models is their multimodal quality (Kurby & Zacks, 2015; Therriault & Rinck, 2012; Zwaan 2014; 2015). Such models are not just a symbolic/semantic representation of the text is activated during construction. Rather, the models include sensorimotor, visual, auditory, and other aspects of a concept, to make a model that accurately reflects the knowledge of that situation or experience. This is what is meant by embodied cognition and is supported by several neurological studies showing areas of the sensorimotor, visual, and auditory cortices are activated during reading comprehension (Kurby & Zacks, 2015; Therriault & Rinck, 2012).

Zwaan (2014; 2015) and other supporters of embodied cognition (see van Dijk, 2014) suggest that both grounded systems (i.e. memory and cognitive systems) and symbolic systems (i.e. linguistic systems) are simultaneously active when building situation models, however the degree to which one system is relied upon varies based on the context. For example, in situations

where events are extensively described by the symbolic system a person does not need to lean heavily on their sensorimotor system, however if they are given sparse details regarding an event they can use their active grounded systems to fill in the gaps. This theory helps explain why people are able to comprehend (or "get by" with) discourses that lack a large amount of information. With regards to dialogue, common ground may have a top-down influence on a situation model. Thus, modulating how much linguistic information is needed to convey meaning. This is discussed in more detail below in the 'Common ground' section.

1.1.2 Event-Indexing

Previous research shows certain aspects of linguistic input are more important to track than others when establishing meaning in a discourse. These factors are the participants (i.e. protagonists and objects) in the dialogue, goals/plans of the participants, setting, time, and cause/effect relations (Zwaan, 2014; 2015). Updating and incorporating changes in only these five categories allows situation models to be built incrementally. In other words, an entire model does not need to be built every time something changes, the existing one can simply be updated. For example, when phrase B is uttered in example (1.1), a listener does not need to update "setting" since it already exists in the situation model thanks to phrase A. A listener hearing phrase B will update "participants" to include Monica but will not need to expend additional cognitive effort on Carrie and Rob since they are already established. Thus incremental processing reduces cognitive effort.

(1.1) A. Last Saturday Carrie and Rob had a picnic in the park.
B. Monica joined them for a game of Frisbee later in the afternoon

While previous evidence shows all five event-indexing dimensions are tracked and updated, studies show certain dimensions are tracked more closely than others, specifically protagonists and time (see Thierrault & Rinck, 2012 for a review). While this suggests all give event-indexing dimensions are not all equal, the previous studies continued to use narratives as their discourse medium. It remains to be seen how the weights of different event-indexing dimensions could vary across different genres of discourse.

Furthermore, unlike narratives, dialogue may elicit new dimensions that are important for comprehension. For example, contextual cues such as, speaker and listener attitudes, have been

shown to be integrated into the understanding of the discourse (van Dijk, 2014; Mills, 2014). Also, the task of establishing and monitoring the common ground is necessary for mutual understanding across interlocutors (Brown-Schmidt, 2012; van Dijk, 2014; Thierrault & Rinck, 2012).

## 1.2 Common Ground

Since aspects of the common ground will be the observed variables in this study, I will now take some time to 1) define the common ground, 2) discuss the types of contributions that can be made to the common ground, i.e. egocentric or other-centric, and 3) discuss how the common ground and its associated context is monitored.

### 1.2.1 Defining Common Ground

Generally speaking, common ground is the shared knowledge that is available to both interlocutors of a conversation, which interlocutors both know they share (Clark, 1985). Many past analyses of common ground have been centered on Gricean assumptions (Kecskés & Mey, 2008). These studies assumed 1) the goal of communication is to share knowledge, 2) that knowledge is understood by a listener who is actively trying to comprehend the intention of a speaker, and 3) the listener does this by making pragmatic inferences (Grice, 1957; 1975). Therefore the study of common ground focuses on listeners' derived meaning of the speakers' intention and the pragmatic inferences that got them there.

More recently pragmatic theories have shifted their focus on intended meaning from the listener's interpretation to the actual shared meaning between both speakers. Various theories account for this shared meaning in different ways. A contextualist approach proposes intention is the inferred meaning between an ideal speaker and an ideal listener who are both conscious of the other's pragmatic inferences (Jaszczolt in Kecskés & Mey, 2008). Motor theory and proponents of mirror neurons suggests that language meaning is shared between speakers by default since the same neurological structures that govern articulators are activated in the process of a speaker's production and a listener's comprehension (Zeevat in Kecskés & Mey, 2008). However, the intention of shared knowledge between interlocutors may not be so important in conversation. Rather, further experimental evidence suggests speakers and listeners often behave egocentrically.

Knutsen and le Bigot (2012) found that after shared and privileged knowledge had been established on a task, interlocutors were more likely to explicitly restate privileged knowledge as

opposed to shared knowledge. Thus, the conversation continued in what seemed to be a very egocentric way. They suggest that egocentric behaviour serves to lessen the cognitive load that comes form monitoring a conversational partner's understanding of the common ground. Furthermore, socio-cognitive researchers do not assume language is necessarily about sharing knowledge, but that its communicative goal may vary from situation to situation, fulfilling different functions. For example language may be used to show an affiliation with a social or cultural group (Enfield in Kecskés & Mey, 2008) or to simply encourage another person to continue speaking (Colston in Kecskés & Mey, 2008).

Another common assumption among the aforementioned theories is common ground is a somewhat static and *a priori* knowledge interlocutors bring to a conversation (Kecskés & Mey, 2008). However, socio-cognitive studies challenge this assumption. They claim common ground is both malleable and updated post factum (Brown-Schmidt, 2012; van Dijk, 2014; Kecskés & Mey, 2008). Brown-Schmidt (2012) argues common ground is not limited to shared knowledge that is distinct from privileged knowledge. Rather, common ground falls along a gradient that is constantly updated based on how much one interlocutor thinks their partner is paying attention to the common ground. For example, if Interlocutor A explicitly asks for information or explicitly confirms recently shared information, Interlocutor B will attend more to the common ground since they assume their partner is focused on the same.

Brown-Schmidt further claims that the *a priori* knowledge interlocutors bring to a conversation may not necessarily be static. Rather, when two participants update common ground with each other in a particular environment they both encode that scenario. The next time those interlocutors are in a similar context they can retrieve their previously updated common ground. An example of this would be two interlocutors who complete a task where they have to move game pieces across a map but only one of them actually knows the map. The next time the same interlocutors encounter a similar task, the strategies they used in the first task will now be part of their *a priori* knowledge. This process is consistent with the process of abstraction used to form memory models (van Dijk, 2014).

Thus, common ground is likely a cognitive tool speakers use to facilitate communication (i.e. collaborate). Common ground is therefore not only an effort to build shared knowledge (Brown-Schmidt, 2012; Knusten & le Bigot, 2012). It is also heavily dependent on memory, such that when interlocutors enter a social situation, previously learned knowledge about the social

context, conversational partner, cultural norms, discourse style, etc. are all retrieved from memory (Brown-Schmidt, 2012; van Dijk, 2014; Gann & Barr, 2014). Smooth communication occurs when two interlocutors' knowledge overlaps (Brown-Schmidt, 2012; Knusten & le Bigot, 2012). Egocentricity serves to lessen cognitive load and likely stems from activated memories, and collaborative speech events (and external cues) update these memories. Thus the context and feedback between interlocutors will influence how much the interlocutors attend to common ground.

### 1.2.2 Contributing to the Common Ground

Audience design occurs when speakers adapt their speech for the sake of an interlocutor; this is other-centric speech. Egocentric speech is when production or interpretation of speech reflects an interlocutor's privileged knowledge, i.e. personal knowledge unknown to other interlocutors. Previous research on the subject has provided mixed results by both confirming instances of other-centric speech but also showing a large amount of egocentric behaviour from speakers. Egocentric speech even occurs when speakers are aware that their interlocutors do not share the same perspective. I will now discuss some of the cues that modulate the balance between other-centric and egocentric behaviour, both in production and comprehension.

#### *1.2.2.1 Audience Design in Comprehension*

Duran, Dale, and Kruez (2011) found both interlocutors are not always equally responsible for contributing to the common ground. They had participants look at a scene, such as an office scenario, that imposed a perspective on the participant and a simulated interlocutor. A direction then followed from the simulated interlocutor such as "move the folder to the left", where the goal of the movement was ambiguous and could only be resolved by taking an other-centric (i.e. the spatial perspective of the interlocutor) or egocentric point of view (i.e. the participant's spatial perspective). Their findings show that when the interlocutor was simulated as a real person (i.e. someone who shared responsibility for clear communication) more participants adopted an ego-centric perspective, whereas when participants was told a hypothetical interlocutor was unaware of the participant's spatial perspective, more participants adopted an other-centric point of view. Thus the perceived knowledge an interlocutor has of a situation can affect how likely people are to adopt other-centric perspectives, even when they know the interlocutor is only hypothetical, despite processing costs. Duran, Dale and Kruez (2011) concluded that as participants increasingly perceive the responsibility of their partner to

contribute to clear communication, participants are more likely to (or perhaps, they have more freedom to) adopt an egocentric perspective.

*1.2.2.2 Audience Design in Production*

Gann and Barr (2014) suggest that speakers have competing egocentric and other-centric representations that are modulated by memory. They suggest that memory cues allow speakers to determine how they need to tailor a description of a referent to an interlocutor. If a referent has been repeatedly established in the memory, it will have a strong memory signal and thus the description of that referent is likely well established and requires little effort for planning. Furthermore, these memories are obligatory and contribute to a cascading effect of production, where having an initial description strongly activated may make it easier to expand on the description of that referent, but it would take cognitive effort to remove information from the description, thus offering one possible explanation for the seemingly egocentric, disproportionate amount of over-specificity in speech.

As Gann and Barr (2014) explain, egocentricity is not only affected by speaker feedback and responsibility (Duran, Dale, & Kruez, 2011), but can also be due to intrinsic factors such as memory retrieval and expert knowledge. Furthermore, Fukumura (2014) demonstrated interlocutors also use extrinsic factors—in particular visual information (and linguistic information)— when describing referents to interlocutors. In fact, visual information may actually ease the need for perspective taking when mapping to referents. Beyond the task of referent mapping, contextual factors (such as power relations or setting) may also come into play (van Dijk, 2014). Common ground is also reflected in the pragmatic structure of a dialogue.

Mills (2014) showed that partners "hypothesis-test" together to not only agree on semantic references, but also the pragmatic structure of an interaction. Hypothesis testing occurs when two participants try out new communicative strategies before agreeing on what works. For example, participants may refer to a referent as "the black shoe", "the shoe below the bag", etc. until agreeing upon "the men's loafer". Agreement is observed when participants no longer suggest new strategies for identifying the referent and continue forward using only the agreed upon strategies. An example of pragmatic hypothesis testing can be seen in a scenario where at the onset of a task participants may elicit information through phrases such as "tell me yours and when I say go…" but be able to achieve the same communicative effect with just "yours?" or by only gazing in their interlocutor's direction by the end of the task (Mills 2014). What makes these

end results collaborative is that they are agreed upon and used by both interlocutors. In addition, more semantic and pragmatic information is often expressed by fewer words, alluding to a more established common ground.

There does not appear to be a default between ego- or other-centric perspectives in production (Gann & Barr, 2014; Fukumura, 2014) or comprehension (Duran et al., 2011). Rather audience design is the result of a constantly changing interface where intrinsic and extrinsic cues modulate perspective taking (Gibbs & Colston in Giora & Haugh, 2017). Thus, contributions to the common ground, or lack thereof, appear to be a reasonable variable to measure when examining contextual cues on dialogue.

#### 1.2.3 Tracking the Common Ground

It is also important to consider how participants track contextual cues that contribute to common ground establishment. Van Dijk (2014) postulates a cognitive mechanism called the K-device that monitors the common ground and surrounding social context. The K-device is constantly updating the ongoing contextual model—different from an situation model since the K-device monitors the *context* of a dialogue, not the *content*—by keeping track of dimensions similar to event-indexing models: setting, identity, role and relation to other interlocutors, and goals of communicative intent.

Previous research as shown factors such as the task at hand, social norms or relationships, and visual and motor-sensory cues can all inform a participants interaction with the common ground (Gann & Barr, 2014; Gibbs & Colston in Giora & Haugh, 2017; Fukumura, 2014). Furthermore, as mentioned previously, these cues not only inform the common ground, they also manipulate how a participant interacts with the common ground, i.e. if they become more or less egocentric.

### 1.3. Dialogue

#### 1.3.1 Defining Dialogue

There are several ways to define dialogue, with each definition making its own assumptions about the processes that underscore language. Part of this definition includes determining the purpose of dialogue. The language-as-product view claims language is a cognitive by-product that arises solely from an individual speaker and that language use should look the same between generic native speakers of the same language (summarized in Brennan, Galati, & Kuhlen, 2010). This contrasts with language-as-action, which assumes language has

been developed via and for use in physical or communicative contexts. Evidence from shifting responsibility for common ground contributions between speakers (Brown-Schmidt, 2012; Duran, Dale, & Kruez, 2011; Gann & Barr, 2014) would suggest the later is more likely.

Dialogue must also be distinguished from monologue, or other language uses. One view is to assume the mere presence of at least one other person, even if they never speak, is enough to make a situation dialogue (summarized in Brennan, Galati & Kuhlen, 2010). This means any collaborative effort is made solely by the speaker. A similar view is to say if a partner only interacts once, for example to give an initial direction, then the whole interaction is still a dialogue. Here, at least one instance of collaboration between the speakers occurs. Both these approaches assume a unilateral use of language, i.e. where one person speaks and another listens. However, many interactions between speakers in both labs and natural use show all interlocutors are constantly contributing to the conversation, either by speaking or by giving feedback (either verbal or non-verbal). Furthermore, interlocutors' utterances often overlap with each other and their conversations demonstrate flexibility in their language use as they adapt to the cues given by their interlocutors and try to establish a shared meaning. So both participants appear to be active contributors to dialogue, even if they are not speaking, therefore any study of dialogue should consider all interlocutors as contributors.

To define a specific model of dialogue, Brennan, Galati, and Kuhlen (2010) claim a model must account for the cognitive processes observed to occur and possibly influence each other in a dialogue: planning, parsing, interpretation, articulation, and monitoring of language. I will now briefly review three possible models of dialogue, concluding with the grounding model, which is the model assumed in the present study.

### 1.3.2 Models of Dialogue

The message model or autonomous transmission model is derived from information theory and defines communication as the transmission and reception of information (see Brennan, Galati, & Kuhlen, 2010 for a review). A speaker encodes an acoustic signal with the information they want to share, a listener receives and decodes this signal and will correctly retrieve the information from it as long as speaker and listener share a common set of coding rules (i.e. a language). If miscommunication occurs, interlocutors will defer to feedback will regulate the information flow. This model assumes partners do not need to recognize an intention to communicate in order to share information, making dialogue similar to monologue, and that as

long as interlocutors share a language, they will follow the same coding rules. However, the latter assumption would predict different partners would establish similar referring expressions but actual data shows large variability in partner's referring expressions (e.g. "the men's loafer" vs. "the black boxy shoe").

A second group of models are two-stage models. In these models, partner-specific information is not taken into account in the first stages of processing, rather it occurs in a separate second stage. One of these models is the interactive alignment model (Pickering & Garrod, 2004). It claims dialogue is fundamentally different from monologue since dialogue simultaneously activates the production and processing systems. This model claims interlocutors begin to prime their production at the same time they begin to process the same dialogue, resulting in shared mental representations, which can appear to look like collaboration, despite being egocentric in nature.

A second two-stage model is the monitoring and adjustment or perspective adjustment theory (Keysar, Barr, & Horton, 1998). Here language processing begins as a fast, automatic, egocentric process. Later in processing, there is a slower, inferential second stage that is capable of integrating partner-specific information, but it requires greater cognitive effort and therefore may be optional or only used for instances of repair. For the majority of a dialogue partners will simply choose to do what is easiest for themselves, but these choices will likely converge due to the shared context. Like the interactive alignment model, alignment (i.e. collaboration) is really an egocentric effect that only appears to be partner orientated due to partners receiving the same contextual cues. However, critiques of these two-stage theories argue that since interlocutors are not exact copies of each other—and they are also likely being primed by other personal biases, not just their shared context—their mental representations can never be exact copies of each other.

A final model to consider, and the one assumed in this study, is the grounding model, as described by Brennan, Galati, and Kuhlen (2010). This model also assumes dialogue is different from monologue, but this is because dialogue is, at its core, a collaboration between interlocutors. In order for collaboration to occur three things must happen: 1) the intent to communicate must be understood from the beginning of a dialogue, 2) the purpose of linguistic signals must be to convey meaning to the listener, and 3) this meaning must be coordinated through feedback and adjustments (i.e. grounding) in order to establish a common ground. Unlike the previous models,

grounding theory claims feedback cues play a major role in shaping a dialogue. Cues can be 1) linguistic, such as a verbal request to clarify information or an "mmhmm" to signal understanding, or 2) paralinguistic, such as a head nod or eye contact. The criteria for these cues is 1) they must encode information, 2) this information must be salvageable by the receiver, and 3) be modified by the cue-givers intentions to collaborate, even if only unconsciously. It should be noted, some cues do not elicit modification from the other interlocutor but may still inform their understanding of the conversation. Furthermore, these cues may be overridden by other factors, such as participants' emotional states or previous experiences.

### 1.3.3 Memory Models in Dialogue

Importantly, participants do not show up to conversations as blank slates, ready to interact within the context of a conversation alone. Rather, their pre-existing models of culture, personal knowledge, former experience with a similar context, etc. will be retrieved during dialogue and can compete with collaborative cues (Brown-Schmidt, 2012; van Dijk, 2014; Thierrault & Rinck, 2012). If these individual influences override collaborative cues, an interlocutor's utterance or interpretation may appear to be more egocentric in nature (Brown-Schmidt, 2012; Brennan & Hanna, 2009; Gann & Barr, 2014; Knutsen & le Bigot, 2012). Furthermore, interlocutors do not only bring their own baggage to a dialogue but they also have expectations about their partners due to factors like familiarity with a partner, pragmatic expectations for a type of dialogue, stereotypical assumptions regarding a partner, or the social status or power-relationship between an interlocutor and their partner (Brennan & Hanna, 2009; van Dijk, 2014; Duran, Dale, & Kreuz, 2011; Gann & Barr, 2014; Mills, 2014). Even if one interlocutor is able to supress these signals in favour for a collaborative interaction, their conversational partner could still fail to act collaboratively, thus the first interlocutor may have their collaborative expectations violated and be forced to adjust.

Interlocutors are shown to be sensitive to partner-specific cues in experiments where participants adapt their behaviour when their partner is switched during the experiment (Brennan, Galati, & Kuhlen, 2010; Brennan & Hanna, 2009), For example, with a previous partner, speakers may have agreed upon a specific referring expression for an object in a context (e.g. "the men's loafer"), but if their interlocutor is changed, a speaker tends to discard the specific expression for a more general one (e.g. "the black shoe") and re-do the process of grounding with their new partner, even if the context remains the same. This is strong evidence that alignment is

dependent on collaboration between interlocutors, and not merely an egocentric effect that only appears partner-specific due to a shared context. Although some evidence shows participants do not adapt to a new partner (see Brennan, Galati, & Kuhlen, 2010 for a review). However, the latter may be due to competing signals to complete a task efficiently that override collaborative cues (Brennan & Hanna, 2009; Fukumura, 2014; Gann & Barr, 2014).

Counter arguments to grounding theory claim interlocutors do behave in egocentric ways, which is counterintuitive if their intention is collaboration. Brennan & Hanna (2009) suggest that collaboration, while paramount to dialogue, is ultimately only one of many constraints interlocutors are influenced by. Despite the importance of collaboration for communication, it may be fallible at times. They suggest egocentric cues like personal biases or mood, or even other external cues such as environmental cues, cognitive load, or task-related biases, all compete with collaboration in a probabilistic, constraint-based model of language processing and production. Thus, occasionally these cues may outweigh collaborative effects and give rise to egocentric behaviour. The variability of findings regarding egocentric and collaborative behaviour in dialogues is further evidence for a model of language processing that is flexible, fallible, and undergoing constant adaptation.

While these experiments suggest multiple factors may influence collaboration, explicit research has not yet been done to probe the pre-existing memory models participants bring from their "outside world" into the lab. Additionally, previous experiments often focus on only one type of dialogue. It is often a dialogue that requires participants to establish an agreed upon reference in order to complete a task (see, for example, Brown-Schmidt, 2012; Brennan & Hanna, 2014; Gann & Barr 2014; Fukumura 2014). If all dialogue is collaborative, as the grounding model assumes, interlocutors should make at least some type of effort to modify common ground in any linguistic situation with two or more people, including open-ended dialogue. However common ground efforts may also be affected by speakers' personal biases. I propose to look into two of these factors, pre-existing participant biases and dialogue genre, in the following naturalistic study.

### 1.4 Current Study

The current research has shown that certain aspects of a conversation, i.e. event-indexing factors, are incorporated into a situation model during comprehension and therefore carry relevant information for understanding. Furthermore, each person's situation model is modulated

by top-down influence from pre-existing knowledge, meaning each person represents the information in a discourse in a unique way. In the context of spoken dialogue these the amount of overlap in mental representations across participants may modulate their ability to establish common ground with each other.

Furthermore, research on dialogue also has shown that context affects the ways in which participants collaborate, their personal interpretations of the discourse, how they contribute to the discourse, and possibly, what aspects of a conversation are relevant to understanding. It has also been demonstrated that while an extensive amount of literature exists on written narrative comprehension, and common ground in referent-resolving tasks, the more natural context of spontaneous, spoken speech remains to be examined. While spontaneous speech is more difficult to control, I have suggested two manipulations with the potential to elicit observed effects: overlap in cultural models and the genre of dialogue as manipulated by a task.

To measure the effects of these variables, I will complete a naturalistic study where participants are given a prompt and their resulting speech is recorded in both video and audio format. The prompts only serve to elicit a particular type of discourse (gossip or summarizing) and are left open-ended so as to not interfere too much with the production of spontaneous speech. The genres here are defined based on information structure. Summarizing is meant to be more structured and similar to a written narrative where the information relevant to a discourse is determined by the narrators or authors of the articles. Gossip is defined as an unrestrictive prompt where speakers can bring their own personal knowledge about one common topic into a conversation.

I will then do an offline measure of my anticipated dependent variables, i.e. the time it takes to establish common ground, the effectiveness of common ground, the communicative cues/feedback given by interlocutors, and the dimensions tracked by interlocutors. To measure these I will code the videos for successful and unsuccessful hypothesis testing (i.e. when participants move on in a conversation) (Mills, 2014); explicit linguistic cues, such as the use of agreed upon referents (particularly referents that refer to any event-indexing dimension); and linguistic and non-linguistic cues of feedback indicating collaboration (e.g. a head nod or repetition of a previous sentence). I predict that 1) the more pre-existing cultural overlap participants have, the faster they will be to establish common ground; 2) participants with more open-ended discussion topics will supplement more of their own personal knowledge into the

common ground and will make a larger effort to collaborate and 3) the type of task will effect which event-indexing dimensions of a dialogue of are tracked and important to the common ground.

This experiment has the potential to shed light on how situation models are established across different genres of discourse. It can also offer a more direct insight into how different pre-existing models, specifically cultural models, influence common ground establishment and collaboration. On a practical level, the footage recorded in this experiment can offer baseline observations and be a resource for materials to be used in future, more controlled experimental studies on discourse models, dialogue, spontaneous speech and collaboration.

## 2. Methods and Procedures

### 2.1 Participants

72 undergraduate students (36 Native; 36 Non-native speakers) enrolled in a first-year linguistic course at the University of Alberta signed up using the Sona Linguistics Sign-Up System and were given course-credit as compensation for their participation. Participants completed a language background survey to be classified as either a native-speaker (NS) or non-native speaker (NN) prior to the experiment.

Native-speakers must have become fluent in English before the age of 5, use English as their primary language on a daily basis, and list English as their dominant language. To control for cultural background they must have lived in Canada for the majority of their life, with no more than three years spent abroad. Those three years abroad can also not have occurred before the age of seven, since peer-interactions sharply increase during preschool years and at the beginning of grade school. Heritage speakers, i.e. participants who spoke their parents' dominant language as a child before learning English, were included if they rated their heritage language as non-native like proficiency, became fluent in English by the age of 5, and rated English as their primary and dominant language with native-like proficiency.

Non-native speakers must speak any language other than English as their dominant and first language and cannot have lived in Canada for more than five years. Five-years is the approximate time it would take someone to complete an undergraduate degree at the University of Alberta, thus this restriction maximizes the potential to recruit the available NN participants while restricting the time spent living in a Canadian context. Multilingual speakers were included if they did not include English in their primary languages and they did not rate their English proficiency as native-like. The final distribution of NN by first language was: Arabic: 1, Cantonese: 1, Korean: 1, Ndebele: 1, Spanish: 1, Vietnamese: 2, Japanese: 3, Mandarin: 26.

### 2.2 Materials and Procedure

Participants were run in pairs and told they were being filmed and recorded as part of a naturalistic language study that will be used in future research. Before beginning the experiment, participants filled out a language background questionnaire to identify them as a NS or NNS of English, as per the qualifications listed above. Later on, a brief political view survey was added to the pre-experiment tests. Each pair participated in either a gossip or summarizing condition. The final distribution was Gossip NS- NS: $N = 6$ pairs, Summarizing NS-NS: $N = 6$ pairs, Gossip

NS- NNS: $N = 6$ pairs, Summarizing NS-NNS: $N = 6$ pairs, Gossip NNS- NNS: $N = 6$ pairs, Summarizing NNS-NNS: $N = 6$ pairs.

Prior to the dialogue participants filled a political ideology questionnaire (adapted from Grenier, 1998, this test was not used in the analysis for this research, but an explanation of its function can be found in Appendix A). Participants in the summarizing condition also read an edited New York Times article: one read an article about Donald Trump's failure in the 2012 presidential race with negative bias, the other read an article about Donald Trump's success in the 2016 presidential race with a positive bias. Donald Trump was chosen in this study because he is an internationally known, highly publicised, and a controversial figure, thus it was anticipated that participants would not withhold contributions to the dialogue due to a lack of knowledge about the topic.

For the recording session, participants sat at a table across from each other in a quiet room at the Centre for Comparative Psycholinguistics at the University of Alberta. Their interaction and audio was recorded using a Noldus Observer video-recording system with two cameras, one focused on each participant. The participant's head and torso were visible in the frame. In the structured summarizing condition, participants received prompt (2.1). The gossip condition received prompt (2.2).

(2.1) "Summarize and discuss the article you read with your partner."

(2.2) "Discuss Donald Trump with your partner."

All interactions were saved and coded for analysis using Noldus Observer and Coder software.

After the recording session, participants independently completed a subjective rating of common ground establishment, their affective perception of their partner, their perceived accentedness in English of themselves and their partner, and their overall affective rating of the experience. In addition, a more in-depth language and cultural background questionnaire completed by participants—the Language Experience and Proficiency Questionnaire (Marian, 2007)—was further used to classify NS and NN speakers. Participants completed four measures of emotional intelligence: Reading the Mind in the Eyes (Baron-Cohen, et al., 2001), Empathy Quotient Questionnaire (Baron-Cohen & Wheelwright, 2004), Big Five Index (John, 2008), and

Interpersonal Reactivity Index (Davis, 1980). At first glance, the results gathered from these tests showed a strong bias against non-native speakers, i.e. non-native speakers scored unusually low on most tests despite not disclosing any reading, learning, or social disabilities. This may have been due to the use of figurative language or cultural assumptions in the test. More data could be collected from native speakers to see how emotional intelligence and personality plays a role in spoken dialogue but the present study cannot address there variables. A more detailed description of these tests and their function can be found in Appendix A.

Participants were then debriefed.

### 2.3 Analysis of Dependent Variables

Following completion of the experiment, I went through the videos and coded the following dependent variables. Due to time restrictions I was the only coder of these videos, a caveat of the current study.

#### 2.3.1 Event-Indexing Factors

There are 5 confirmed event-indexing factors: protagonists/objects, cause and effect relations, goals and plans, setting, and time. Protagonists/objects were considered to be any human, animate, or inanimate object mentioned in the discourse, with the exclusion of the speaker, listener, and source of information. Cause and effects were coded into the same category. Effects were coded if they were preceded by a "so", "and then", or other outcome marker that linked it to another proposition. Causes were coded if they were preceded by a "because", "if" or other causal marker, or if they were the proposition linked to an effect (e.g. "All he cares about is twitter [=cause], he can't run a country [=effect]"). Goals and plans were defined as any mention of a volitional proposition that was linguistically marked as not yet taking place (e.g. "He wants to build a wall"). Setting included physical settings, nations, and political parties when they were not used as protagonist or object (e.g. "there was a lot of in-fighting *within the Reform Party*" (= setting)). Time was defined as specific times such as dates, months, morning vs. evening, as well as relational time expressions such as "before" and "after".

Every proposition was coded for at least one event-indexing (EI) factor. If an utterance could not be clearly placed into one of the above categories, it was assigned a new category. Speaker attitude (affective or evaluative statements marked as being the speaker's own opinion), interlocutors (references to the speaker and/or listener), and source (where the speakers got their

information from, e.g. “I saw a video, I think it was from *Facebook* (=source)”) were the three new categories.

2.3.2 Common Ground Cues

Explicit feedback cues demonstrating understanding or confusion relating to the previous statement were coded. Instances of understanding were coded as every instance of head nods, affirmative expressions (e.g. “yeah”, “mhmm”, “okay”) and reiterations (Speaker A: “I’m Japanese-” Speaker B: “Japanese?” Speaker A: “Right”). These cues of understanding often overlap with agreement and discourse particles so instances were coded as either definitely a common ground marker, possibly a common ground marker, and definitely not a common ground marker, depending on how linked they seemed to be to the previous statement. Possible and definitely not common ground markers were excluded from the current analysis.

Instances of confusion were coded as every instance of eye contact breaks, gestures like brow furrows and frowns, expressions like “huh”, “umm”, and clarifications (e.g. Speaker A: “He’s even worse” Speaker B: “Donald Trump?” Speaker A: “No Mike Pence”). Like understanding cues, confusions cues also overlapped with behaviours unrelated to common ground such as disagreement or boredom. The same restriction was applied so that depending on how obvious the connection of the cue was to the previous statement instances were coded as either definitely a common ground marker, possibly a common ground marker, or definitely not a common ground marker. Again, possible and definitely not common ground markers were excluded from the current analysis.

Another measure of common ground establishment is agreed upon references. Many common ground studies use the names of referents as a measure of common ground establishment (Brown-Schmidt, 2012; Brennan & Hanna, 2014; Gann & Barr 2014; Fukumura 2014). Repeated referents and unique names for referents suggest that both participants have agreed upon how to refer to a specific antecedent within the context of the task. Repeated referents were coded as the first time a speaker used the same label for an antecedent as their partner (e.g. Speaker A: “Ted Cruz” Speaker B: “Ted Cruz”). Unique referents when a previously established label was changed and then accepted by the other participant either by them repeating it or giving a cue of comprehension (e.g. “Ted Cruz” → “Mister Cruz”).

2.3.3 Pragmatic Convergence

I coded instances of alignment and pragmatic cues that alluded to an established common ground. Alignment was defined as use of the same linguistic structure (e.g. Speaker A: “Yeah and I feel like, …” Speaker B: “Yeah, and I feel like…”). Another sign that common ground had been established were instances where previously specific statements became generalized or more ambiguous, demonstrating less linguistic information being used to represent more semantic content (e.g. Speaker A, earlier in conversation: “It was when he became the presidential candidate, for the party- the republican party candidate, in Indiana” vs. Speaker A, later in conversation “he became the nominee”).

# 3. Event Indexing

## 3.1 Results

### 3.1.1 Statistical Analysis

The statistical analysis of results was carried out in R (version 3.3.1, R Core Team, 2016). The response variable was the number of times a participant mentioned an event-indexing factor and manipulated predictor variables were Genre of dialogue: gossiping or summarizing an article, Nativeness: native English speaker (NS) or non-native English speaker (NN), and Pairing: NN-NN, NS-NN (or mixed), NS-NS. Descriptive statistics were found using the summarise() function of the "dplyr" package ("dplyr", Wickham & Francois, 2016). I used generalized linear regression analysis with Poisson family specified in the glm() function of the lme4 R package ("lme4", Bates, Maechler, Bolker, & Walker, 2015).

All models were fitted in the same way. The initial model started with the event-indexing factor under investing (protagonists/objects, cause/effects, goals/plans, time, setting, or speaker attitude) modeled as a function of the main effects of Genre, Pairing, and Nativeness with a 3-way interaction. I used backward elimination to remove non-significant variables if they did not contribute to model fit in order to find the simplest models. Model comparisons were done using Akaike Information Criterion (AIC): AIC(model1,model2). More complex models were only kept if their AIC score was two or more points lower than the simpler model's AIC score.

None of the models found a significant 3-way interaction or a significant interaction between pairing and nativeness, thus pairing and nativeness was the first interaction removed in the backward fitting. I further removed insignificant interactions between Genre and Pairing and/or Genre and Nativeness if the model comparison showed the model fit was improved by discarding these interactions. Finally I removed insignificant main effects if model comparisons showed this improved the fit of the model. The best fitting models for each variable and their results are described in detail in each of the following sub-sections.

### 3.1.2 Protagonists and Objects

Protagonists and objects were grouped into the same category. They were considered to be any human, animate, or inanimate object mentioned in the discourse, with the exclusion of the speaker, listener, and source of information (e.g. "my article [=source] said…" or "I saw on Twitter [=source] that…"). Updates to the protagonist or object were also included (e.g. "He [=Donald Trump] is a businessman").

Each participant received their own score for how many times they mentioned a protagonist or object ($M$ = 28.99, $SD$ = 13.28, $Mdn$ = 28, $min$ = 8, $max$ = 73). Figure 3.1 shows a boxplot illustrating the distribution of the instances of protagonists/objects in each genre. Although the median is slightly higher in the gossip condition than the summarizing condition, there is large overlap between the distributions of the two genres.

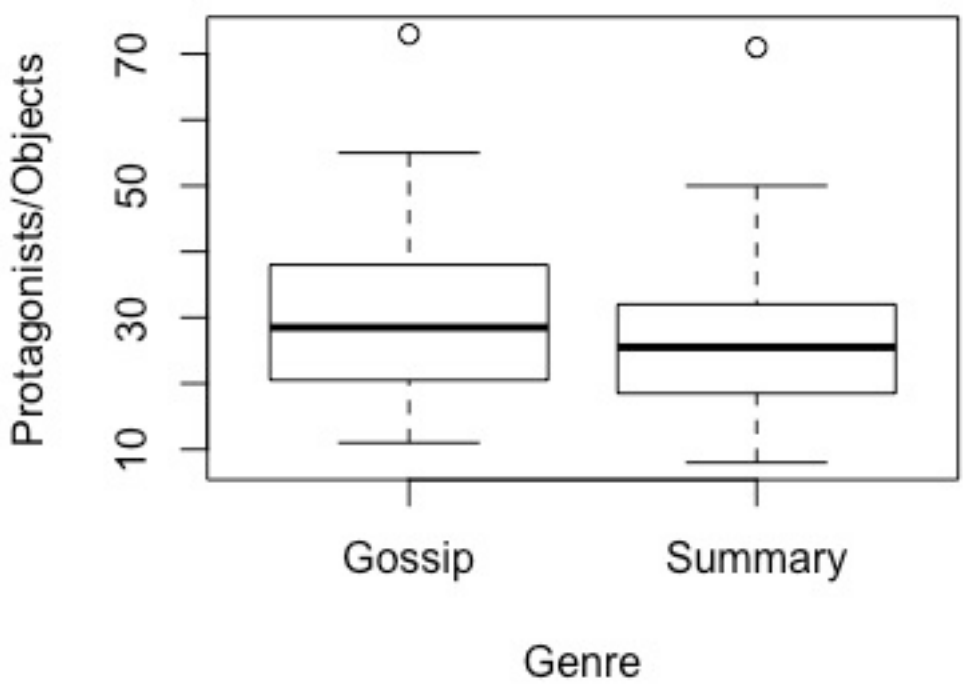


Figure 3.1. Distribution of the instances of protagonists and objects by genre.

The best fitting model was protagonists and objects as a function of the main effects genre (gossip or summarizing), nativeness (native speaker (NS) or non-native speaker (NN)) and pairing (NS-NS, NS-NN, NN-NN) with interactions between genre and pairing (*AIC* genre + pairing = 706.061; *AIC* genre * pairing = 702.653). The intercept of protagonists/objects was significantly different from zero (*estimate* = 3.258, $z$ = 57.550, $SE$ = 0.057, $p < 0.001$). The model showed no significant main effect for genre. Nativeness had a significant main effect such that instances of protagonists/objects are more likely to be produced by native English speakers than

non-native speakers (*estimate* = 0.575, *z* = 7.153, *SE* = 0.080, $p < 0.001$), as illustrated in Figure 3.2. All model effects were plotted with the "effects" package in R ("effects", Fox, 2003; Fox & Hong, 2009)

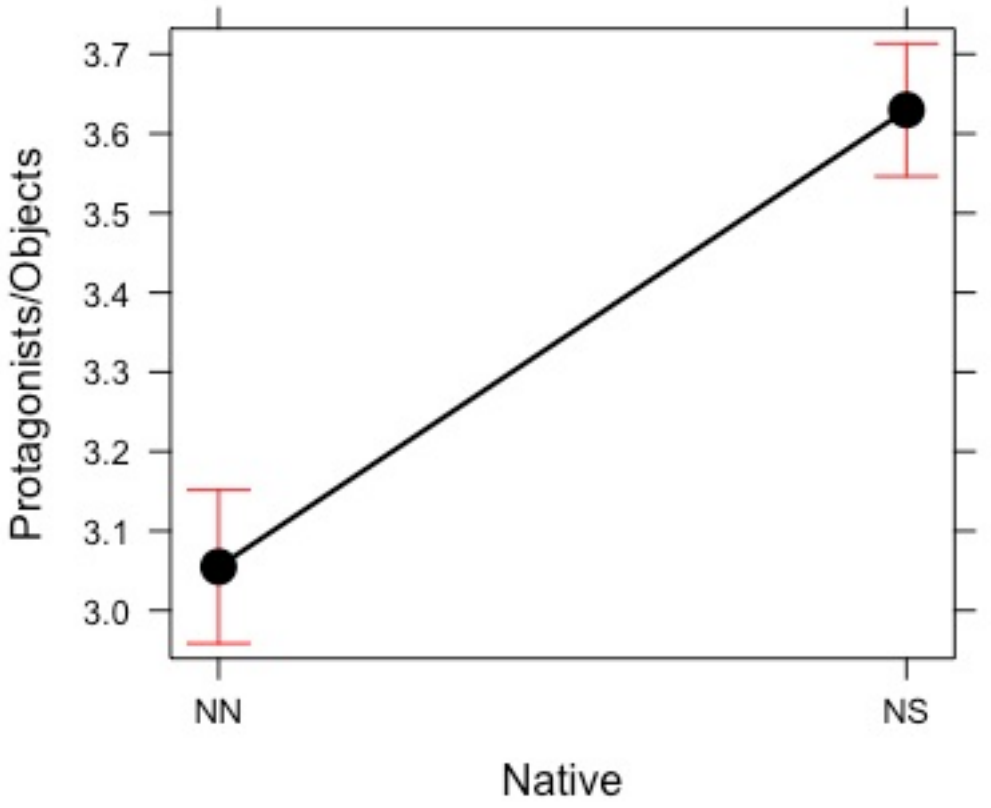


Figure 3.2. Main effect of nativeness on instances protagonists and objects. Error bars represent 95% confidence intervals.

However, pairing also had a main effect such that NN-NN pairs mentioned protagonists and objects significantly more often than NS-NS pairs (*estimate* = -0.294, *z* = -2.678, *SE* = 0.110, $p < 0.01$). A multiple comparison carried out using the pairs(lsmeans()) function from the "lsmeans" R package ("lsmeans", Lenth, 2016) confirmed NN-NN pairs had more instances of protagonists/objects than NS-NS pairs (*estimate* = 0.295, *z-ratio* = 3.048, *SE* = 0.097, $p < 0.01$) and that NN-NN pairs had more instances of protagonists/objects than NS-NN pairs (*estimate* = 0.241, *z-ratio* = 3.163, *SE* = 0.076, $p < 0.01$). There was no significant effect found between NS-NS and NS-NN pairs. The main effect of pairing is plotted in Figure 3.3.

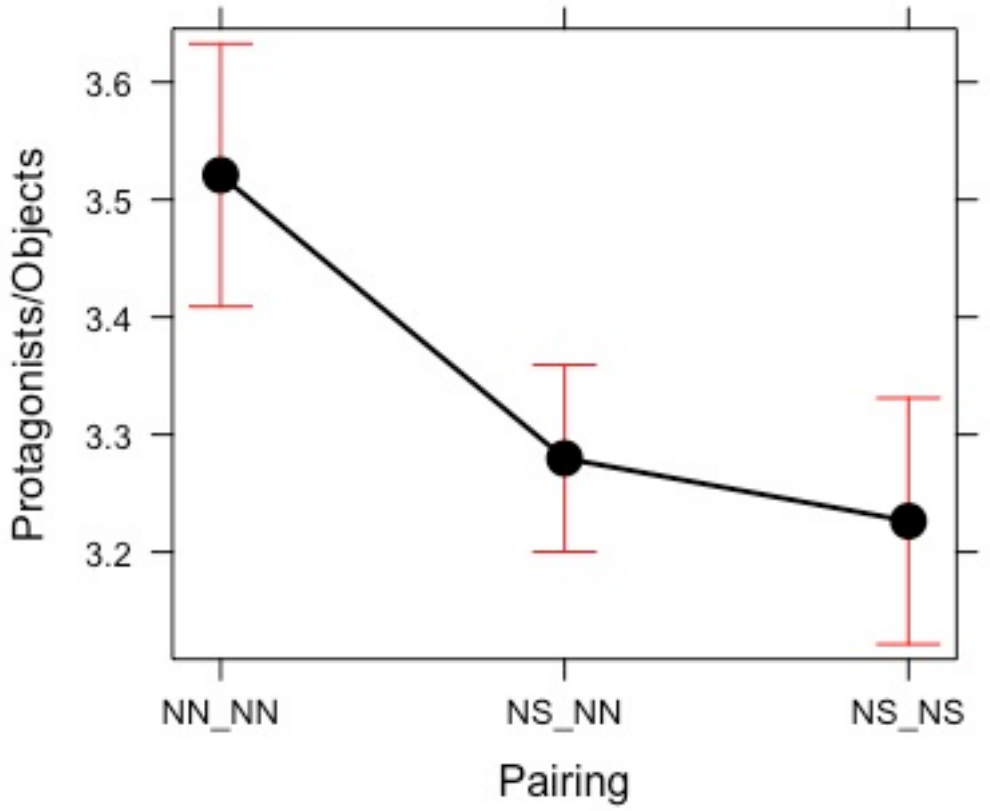


Figure 3.3. Main effect of pairing on number of mentions of protagonists/objects. Error bars represent 95% confidence intervals.

My model found a significant interaction between genre and pairing such that the number of instances of protagonists/objects significantly decreased for NS-NN pairs from gossip to summary conditions (*estimate* = -0.257, $z$ = -2.281, *SE* = 0.113, $p < 0.05$). This effect is shown in Figure 3.3 and was confirmed in a multiple comparison (*estimate* = 0.360, *z-ratio* = 3.919, *SE* = 0.078, $p < 0.01$).

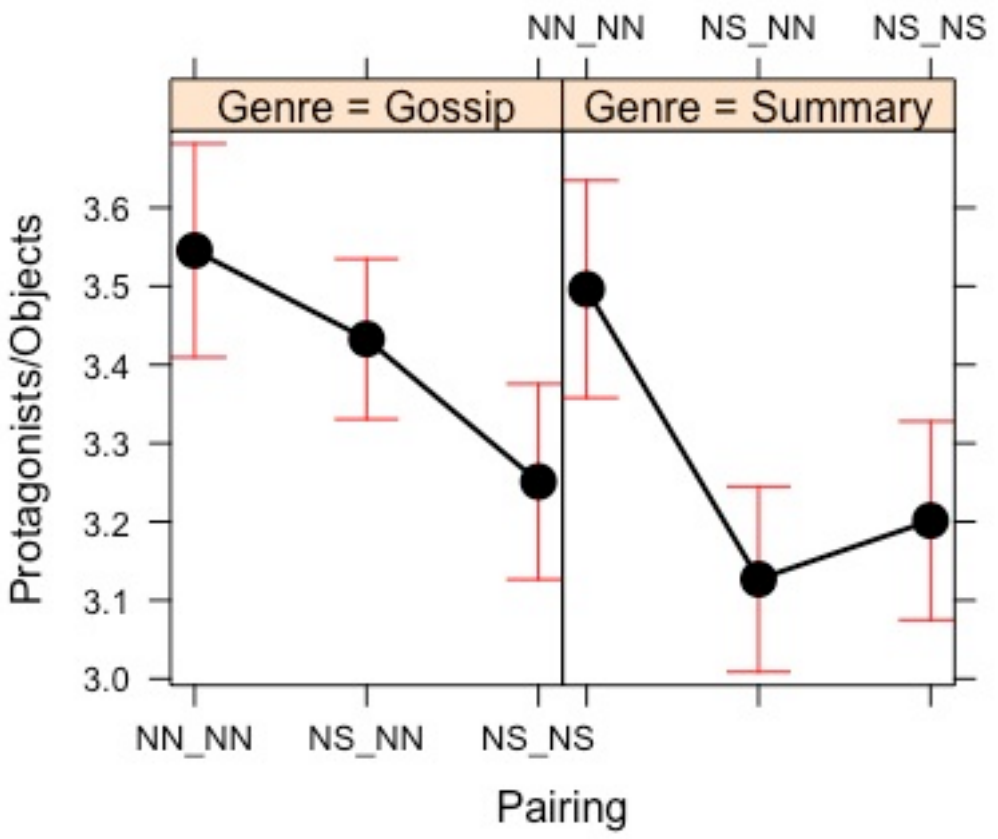


Figure 3.4. Instances of protagonists and objects as a function of the interaction between pairing and genre. Error bars represent 95% confidence intervals.

In summary, there was no significant main effect of genre on protagonists and objects. When genre interacted with pairing there was a significant effect on NS-NN pairs such that NS-NN pairs mentioned protagonists and objects more frequently in gossip conditions than summary conditions. Pairing also had a significant main effect where NN pairs mentioned protagonists and objects more than both mixed and NS pairs. Contrary to this there was a significant main effect of nativeness that showed native speakers had more instances of protagonists and objects overall than non-native speakers.

3.1.3 Cause and Effect Relationships

Effects were coded if they were preceded by a “so”, “and then”, or other outcome marker that linked it to another proposition. Causes were coded if they were preceded by a “because”, “if” or other causal marker, or if they were the proposition linked to an effect (e.g.

"All he cares about is twitter [=cause], he can't run a country [=effect]"). These relationships (cause and effect) were included in the same category.

The mean number of cause and effect relationships mentioned by a participant was 27.375 (*Med* = 25, *SD* = 13.680, *min* = 3, *max* = 71, *N* = 72). The distribution of these instances by genre is illustrated in Figure 3.4. Like protagonists and objects there appears to be slightly more occurrences in the gossip condition than the summarizing condition, although there is again a large overlap between the two conditions.

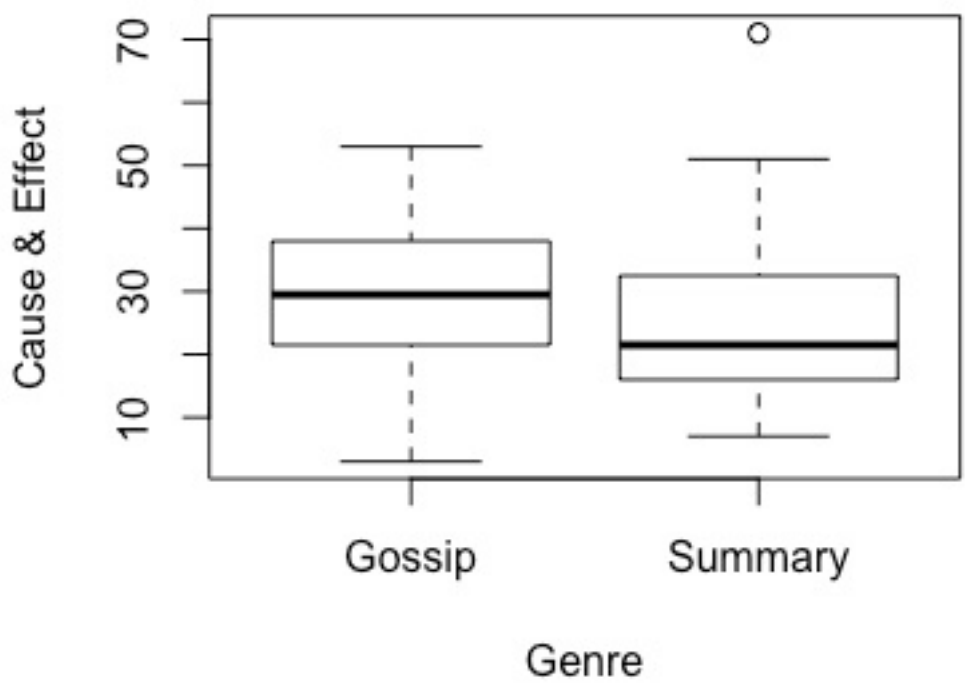


Figure 3.5. Distribution of the instances of cause and effect relationships in each genre.

The best fitting model included instances of cause and effect relationships as a function of the main effects of genre, nativeness, and pairing, with no interactions. The intercept of cause and effect relationships was significantly different from zero ($y = 3.165$, $z = 65.772$, $SE = 0.0481$, $p < 0.001$). Genre had a main effect such that there were fewer instances of cause and effect

relationships in summarizing conditions than there were in gossip conditions (*estimate* = -0.121, z = -2.679, $p < 0.01$) (Figure 3.5).

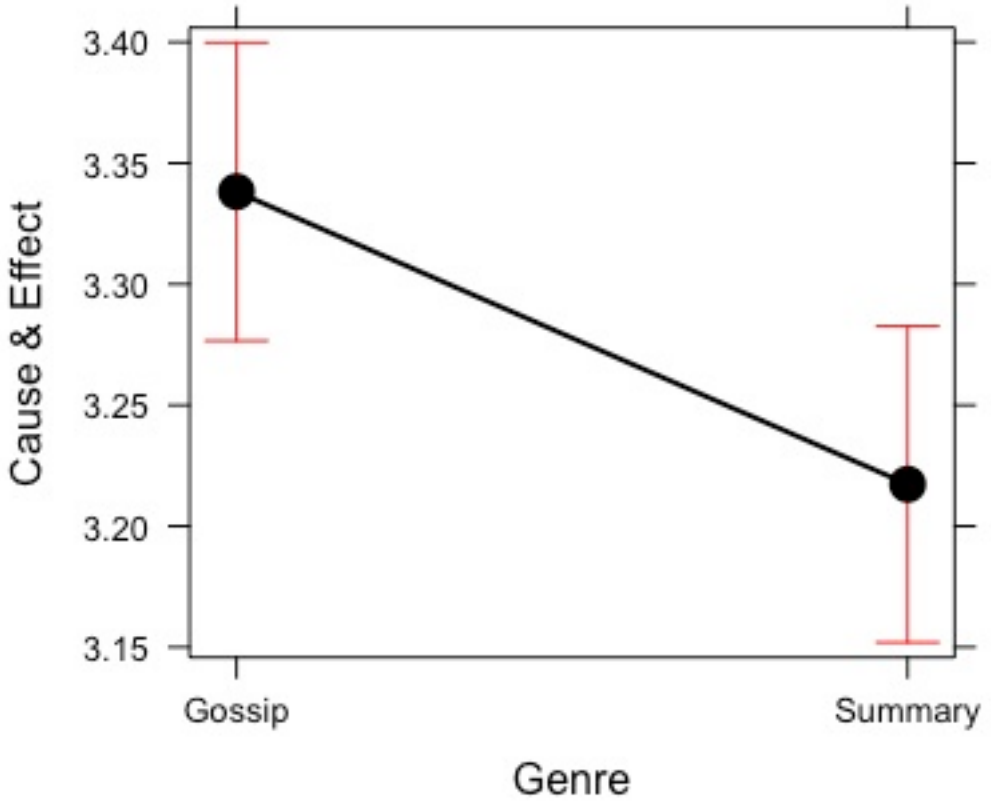


Figure 3.6. Instances of cause and effect as a function of genre. Error bars represent 95% confidence intervals.

Nativeness also a main effect on cause and effect, such that native English speakers had significantly more instances of cause and effect relations than non-native speakers (*estimate* = 0.684, $z = 8.329$, $SE = 0.082$, $p < 0.001$), as seen in Figure 3.6.

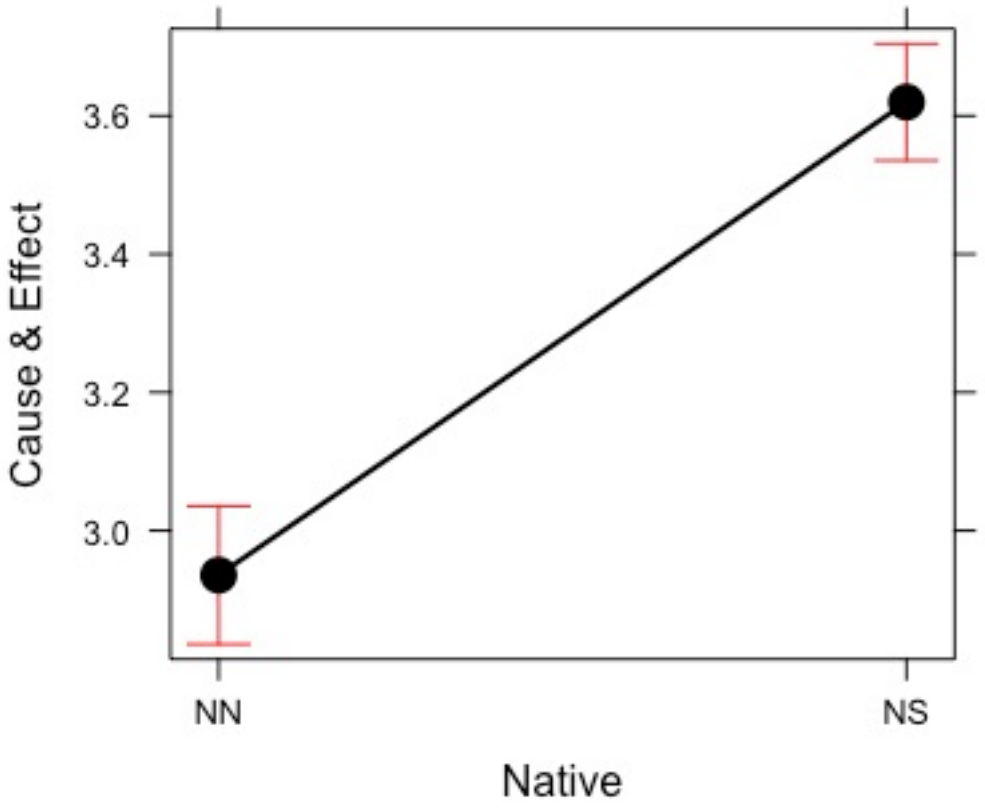


Figure 3.7. Instances of cause and effect as a function of nativeness. Error bars represent 95% confidence intervals.

Pairing also had a main effect on cause and effect relationships such that NN-NN pairs had significantly more instances of NS-NS (*estimate* = -0.32188, $z$ = -3.233, *SE* = 0.100, $p <$ 0.001) and NN-NS pairs (*estimate* = -0.184, $z$ = -2.307, *SE* = 0.080, $p < 0.05$), as seen in Figure 3.7 However a multiple comparison only confirmed a significant effect of NN-NN having more instances than NS-NS (*estimate* = 0.323 *z-ratio* = 3.233, *SE* = 0.100, $p < 0.01$), no significant effect were found for NN-NN compared to NS-NN or NS-NS compared to NS-NN.

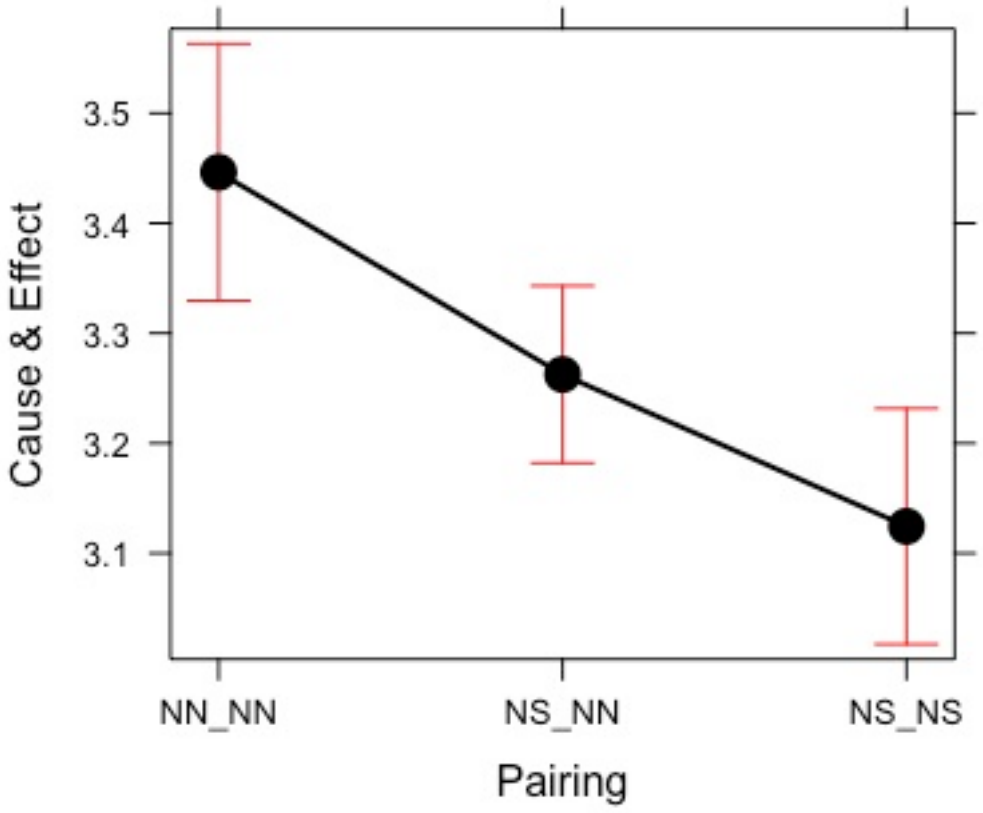


Figure 3.8. Instances of cause and effect as a function of pairing. Error bars represent 95% confidence intervals.

In summary, genre had a significant main effect on instances of cause and effect relations. Cause and effect relations were more likely to occur in gossip conditions than summarizing conditions. Like protagonists and objects, native speakers had a significantly higher number of instances of cause and effect than non-native speakers, but NN pairs had a significantly higher number of instances than mixed or NS pairs.

3.1.4 Goals and Plans

Goals and plans were defined as any mention of a volitional proposition that was linguistically marked as not yet taking place (e.g. “He wants to build a wall”). There were far fewer instances of goals and plans than the previous two variables, with a mean of 3.819 (*SD* = 3.105, *Mdn* = 4, *N* = 72), a minimum of 0 instances and a maximum of 13 instances. A look at

the distribution of these instances by genre shows the summary condition had more overall instances of goals and plans (Figure 3.8).

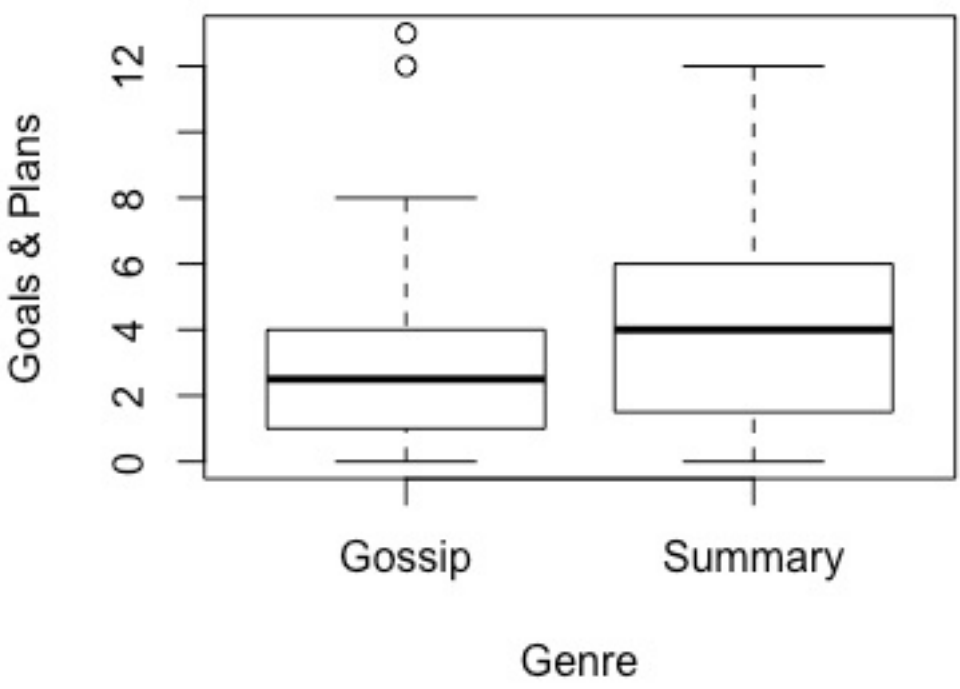


Figure 3.9. The distribution of the instances of goals and plans in each genre.

The best fitting model had instances of goals and plans as a function of the main effects of genre, pairing, and nativeness, with an interaction between genre and pairing (*AIC* genre + pairing = 375.632; *AIC* genre * pairing = 366.064). There is significant reason to believe the intercept is different from zero (*estimate* = 0.734, $z = 3.670$, *SE* = 0.200,, $p < 0.001$). Genre had a significant main effect such that instances of goals and plans were more likely to occur in summary conditions than gossip (*estimate* = 0.908, $z = 3.834$, *SE* = 0.237, $p < 0.001$). This effect is illustrated in Figure 3.9.

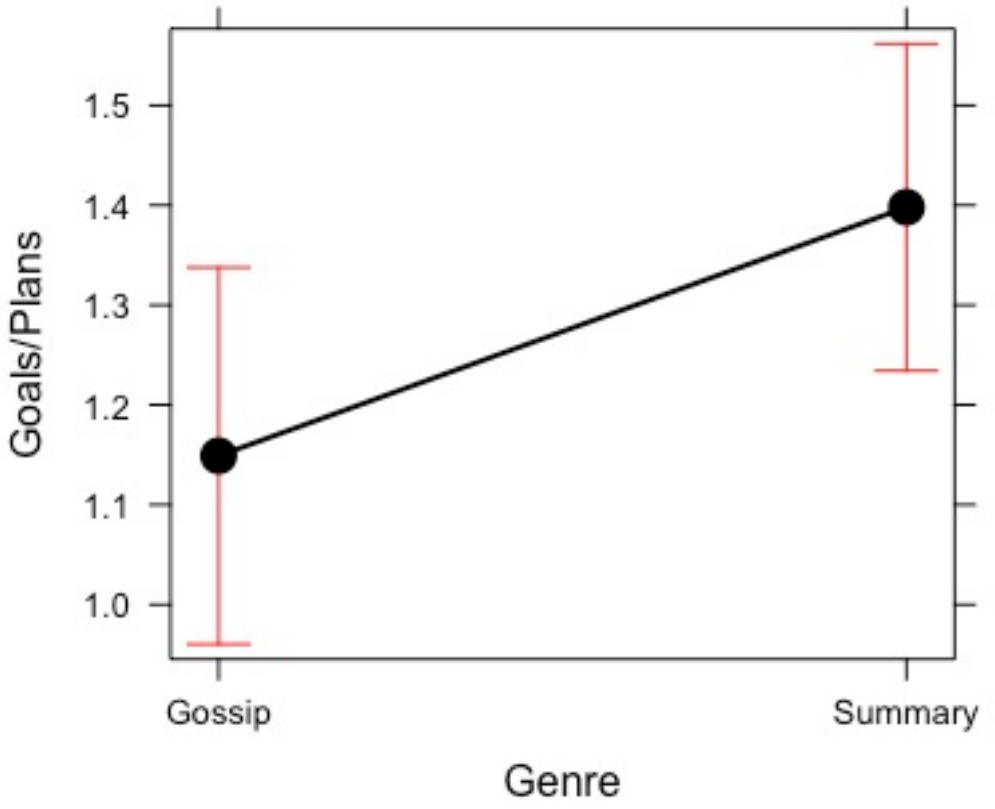


Figure 3.10. Instances of goals and plans as a function of genre. Error bars represent 95% confidence intervals.

The model showed a significant main effect of nativeness such that native speakers had more instances of goals and plans than non-native speakers (*estimate* = 0.778, $z = 3.754$, *SE* = 0.207, $p < 0.001$) (Figure 3.10).

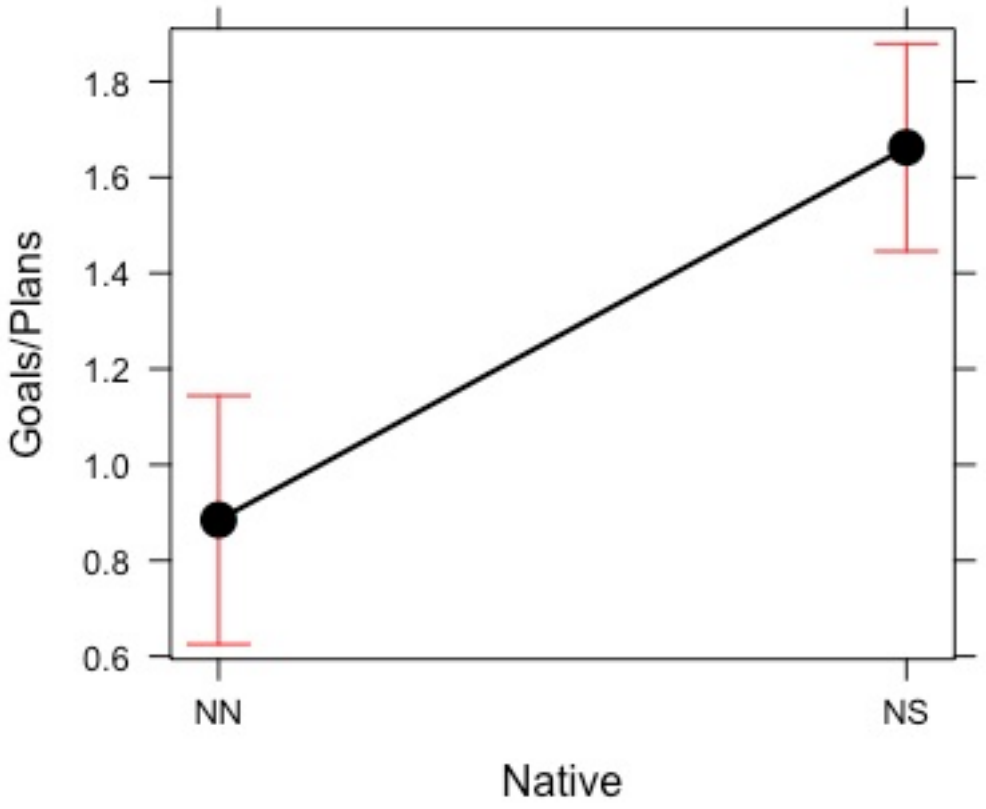


Figure 3.11. Instances of goals and plans as a function of nativeness. Error bars represent 95% confidence intervals.

There was no significant main effect of pairing on instances of goals/plans, but there was a significant interaction between pairing and genre (Figure 3.10). A multiple comparisons test showed that NN-NN pairs had significantly fewer instances of goals and plans in gossip conditions than summary conditions (*estimate* = -0.908, *z-ratio* = -3.834, *SE* = 0.237, $p < 0.01$). The same NN-NN gossip condition also had significantly lower instances of goals and plans than mixed pairs (NS-NN) in either gossip (*estimate* = -0.824, *z-ratio* = -3.436, *SE* = 0.240, $p < 0.01$) or summary conditions (*estimate* = -0.713, *z-ratio* = -2.920, *SE* = 0.244, $p < 0.05$).

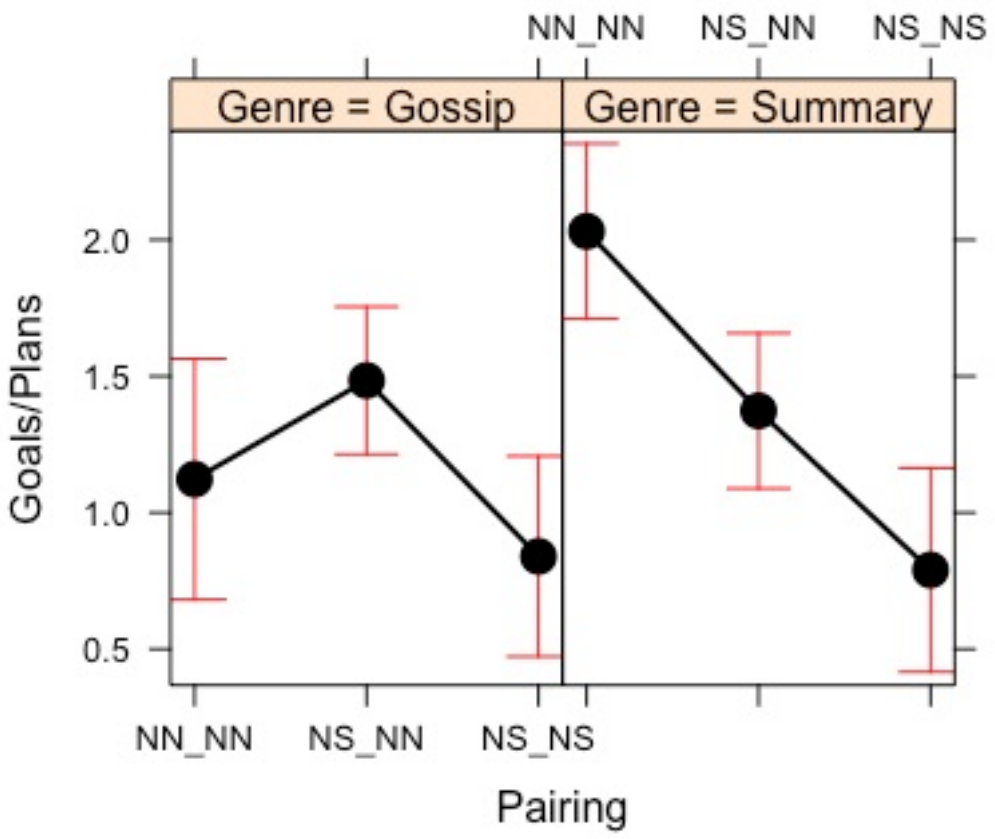


Figure 3.12. Instances of goals and plans as a function of the interaction between genre and pairing. Error bars represent 95% confidence intervals.

In summary, goals and plans appeared significantly more times in summary conditions than gossip conditions. The native speaker effect was again found with native English speakers having more instances of goals and plans than non-native English speakers. No significant main effect for pairing was found, however an interaction between pairing and genre showed NN pairs mentioned goals and plans significantly less in gossip conditions than in summary conditions.

3.1.5. Time

Time included specific times such as dates, months, morning vs. evening, as well as relational time expressions such as “before” and “after”. Time was mentioned less frequently overall out of all the event-indexing factors, with the mean number of mentions of time for any participant being 6.153 (*SD* = 4.313946, *Mdn* = 5, *min* = 0, *max* = 20, *N* = 72). A look at the

distribution of occurrences by genre shows more variation in the summarizing condition but similar medians in the time data (Figure 3.11).

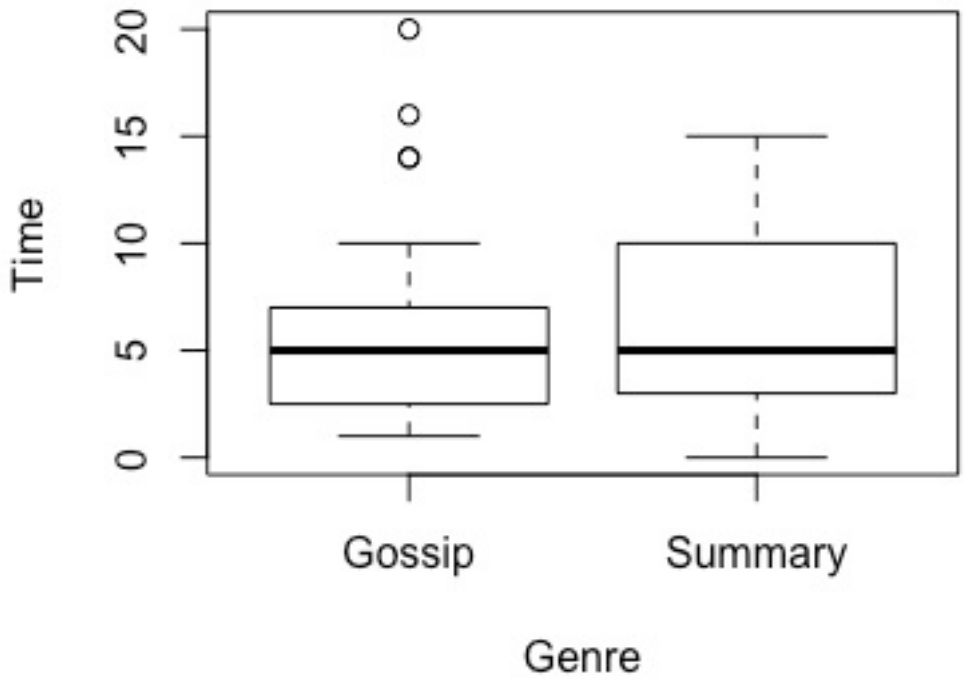


Figure 3.13. The distribution of the occurrences of time in each genre.

The best fitting model included instances of time as a function of the main effects of nativeness and pairing. The intercept was significantly different from zero ($y = 1.456$, $z = 13.452$, $SE = 0.108$, $p < 0.001$). Including Genre (*estimate* = 0.113, $z = 1.187$, $SE = 0.108$, $p > 0.1$) did not improve model fit and was dropped (*AIC* with genre = 388.601; *AIC* without genre= 388.0123). There was a significant main effect of nativeness on instances of time ($b = 0.707$, $z = 3.425$, $SE = 0.207$, $p < 0.001$), such that native English speakers mentioned time more frequently than non-native speakers (Figure 3.12).

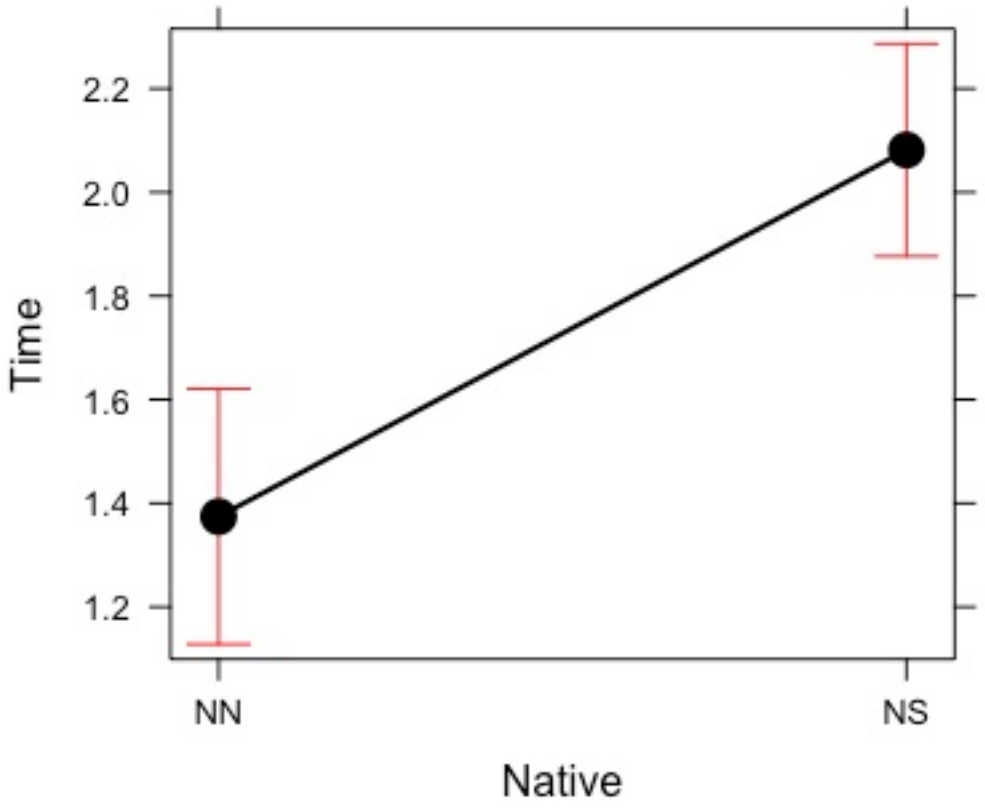


Figure 3.14. Instance of time as a function of nativeness. Error bars represent 95% confidence intervals.

Pairing had a significant effect such that NN pairs had significantly more instances of time than mixed pairs (*estimate* = -0.443, $z$ = -2.279, *SE* = 0.194, $p < 0.05$), however there was no significant difference between NN pairs and NS pairs. A plot of this effect confirms the prediction of our model and suggests NS pairs may also have significantly more instances of time than mixed pairs (Figure 3.13).. A multiple comparison confirms that NS pairs had significantly more instances of time than mixed pairs (*estimate* = -0.474, *z-Ratio* = -3.484, *SE* = 0.136, $p <$ 0.01) and NN pairs had a trending effect of more instances of time than mixed pairs (where mixed pairs are the intercept, *estimate* = 0.443, $z$-Ratio = 2.279, *SE* = 0.194, $p <$ 0.1) with no significant differences between the instances of time mentioned in NN pairs and NS pairs (*estimate* = -0.31, *z-Ratio* = -0.129, *SE* = 0.273, $p > 0.5$).

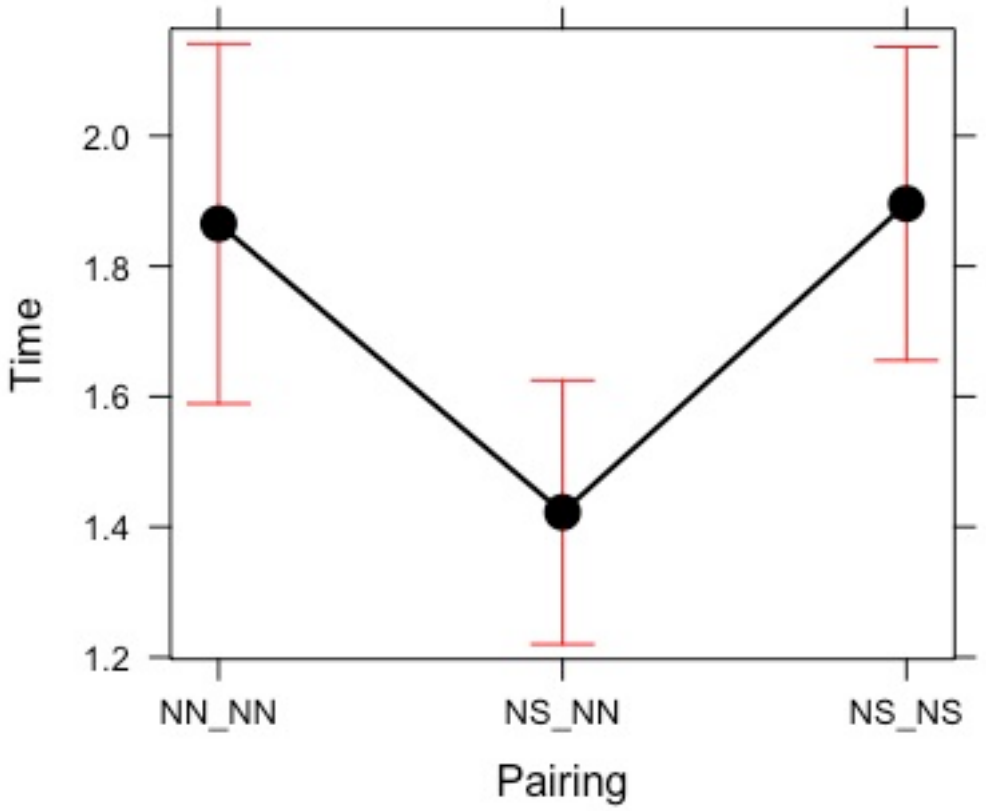


Figure 3.15. Time as a function of the main effect of pair. Error bars represent 95% confidence intervals.

In summary, genre did not predict the instances of time and was removed from the model. The native speaker main effect was found again for time with native English speaker mentioned time more often than non-native English speakers. Pairing had a significant main effect where both NN pairs and NS pair had significantly more instances of time than mixed pairs.

3.1.6 Setting

Instances of setting included physical settings, nations, and political parties when they were not used as protagonist or object (e.g. "there was a lot of in-fighting *within the Reform Party*" (= setting)). The mean number of instances of setting was 13.611 (*SD* = 7.082, *Mdn* = 14, *min* = 0, *max* = 30, *N* = 72). The distribution of setting is similar to protagonists and objects and

cause and effect relationships, with a higher median of occurrences in the gossip condition (Figure 3.14).

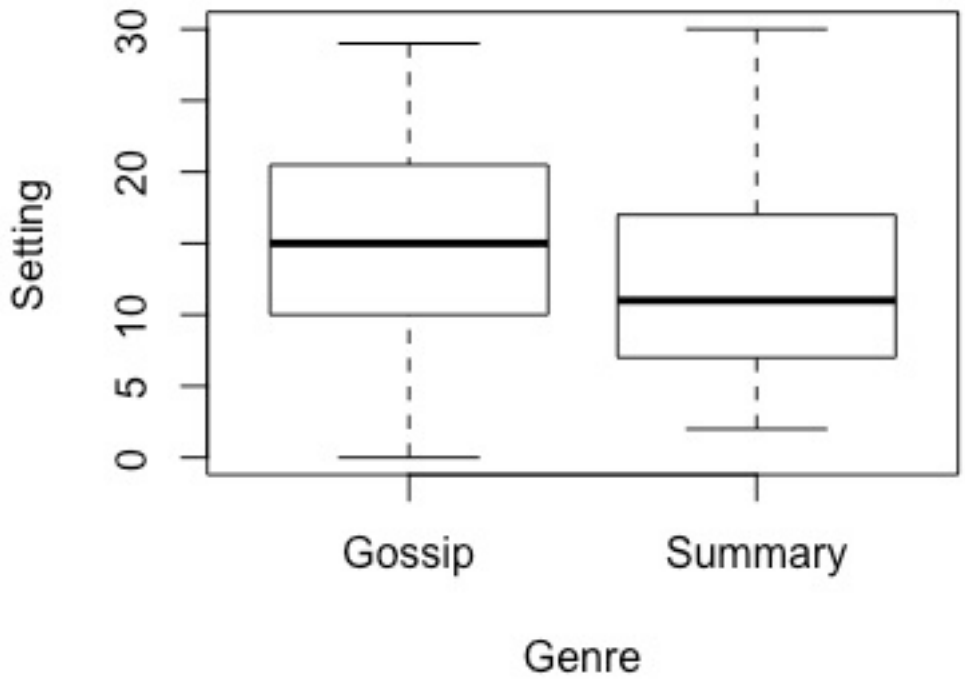


Figure 3.16. Distribution of the instances of setting by each condition.

The best fitting model included setting as a function of the main effects of genre, nativeness, and pairing, with an interaction between genre and pairing. (*AIC* with interaction = 570.071, *AIC* without interaction = 579.900). There is reason to believe the intercept is significantly different from zero ($y = 2.813$, $z = 39.788$, $SE = 0.071$, $p < 0.001$). The model shows instances of setting are significantly more likely to occur in gossip conditions than summary (*estimate* = -0.560, $z = -4.913$, $SE = 0.118$, $p < 0.001$). This effect is illustrated in Figure 3.15.

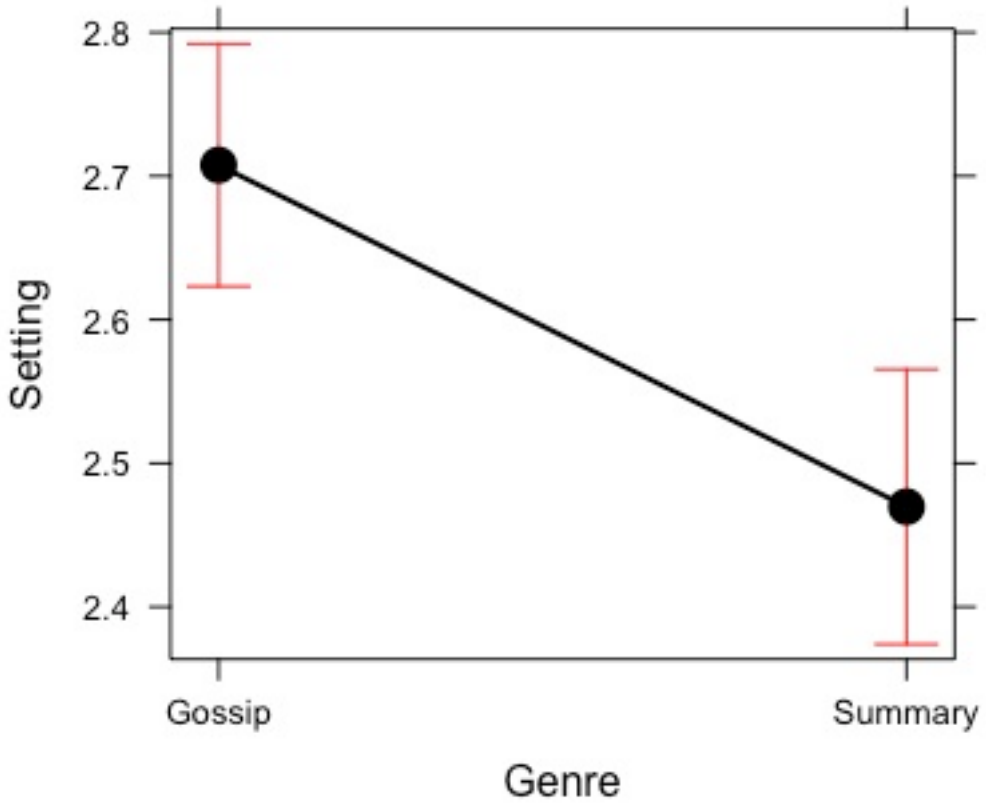


Figure 3.17. Instances of setting as a main effect of genre. Error bars represent 95% confidence intervals.

The model also predicted a main effect of nativeness such that native English speakers had more instances of setting than non-native speakers (*estimate* = 0.415, $z = 3.758$, $SE = 0.110$, $p < 0.001$), as seen Figure 3.16.

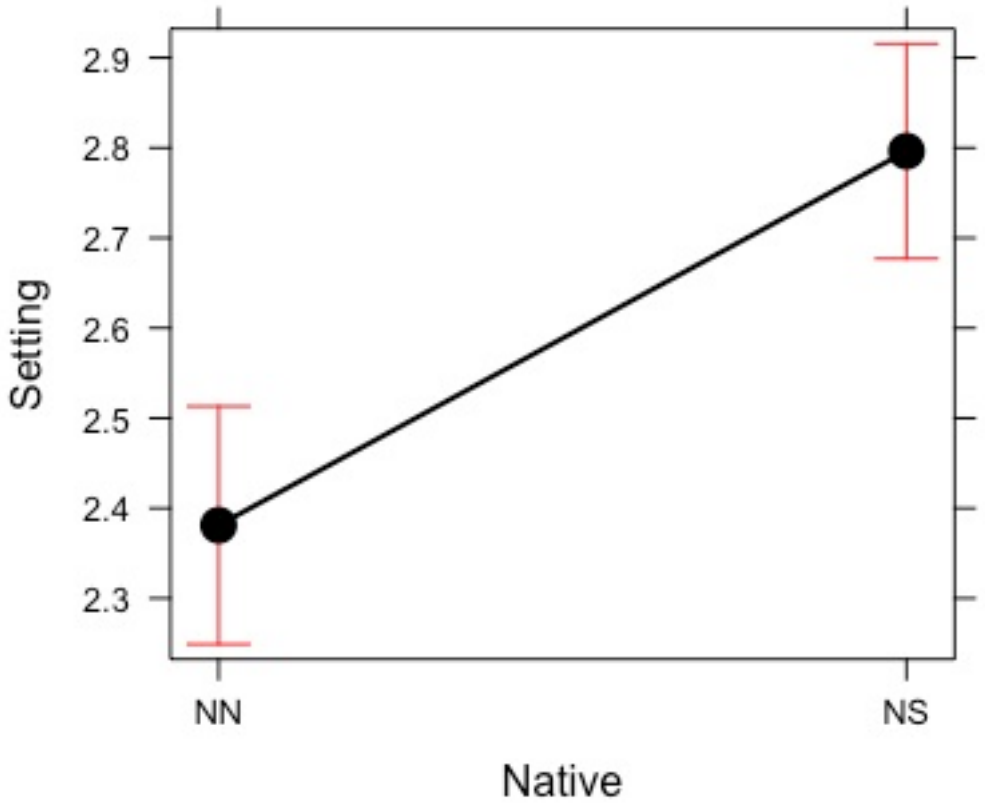


Figure 3.18. Instances of setting as a function of the main effect of nativeness. Error bars represent 95% confidence intervals.

There was a significant main effect of pairing on instances of setting between NN pairs and NS pairs (*estimate* = -0.566, $z$ = -3.731, *SE* = 0.151,, $p < 0.001$) and NN pairs and mixed pairs (*estimate* = -0.374, $z$ = -3.033, *SE* = 0.123,, $p < 0.01$), predicting NN pairs to have more instances of setting. A multiple comparisons test confirmed these results the effect between NN and NS pairs (*estimate* = 0.331, *z=ratio* = 2.419, *SE* = 0.137, $p < 0.05$), however it found no difference between NN pairs and mixed pairs (*estimate* = 0.096, *z-ratio* = 0.921, *SE* = 0.104 $p >$ 0.5). There was a significant effect found between NS pairs and mixed pairs that predicts mixed pairs to have more instances of setting than NS pairs (*estimate* = 0.236, *z-ratio* = 2.645, *SE* = 0.089, $p < 0.05$) This effect is plotted in Figure 3.17.

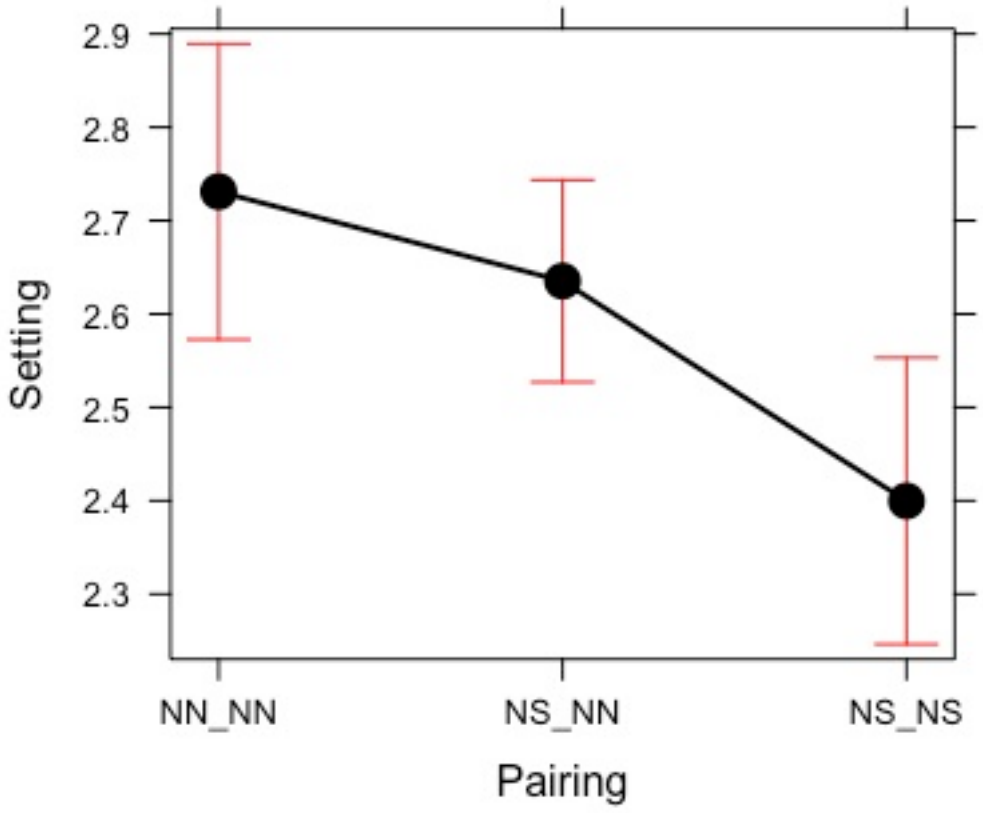


Figure 3.19. Instances of setting as a function of the main effect of pairing. Error bars represent 95% confidence intervals.

Pairing also had a significant interaction with genre, as seen in Figure 3.18. A multiple comparisons showed NN pairs had significantly more instances of setting in gossip conditions than summary conditions (*estimate* = 0.580, *z-ratio* = 4.913, *SE* = 0.118, $p < 0.0001$). NN pairs in summary conditions had significantly fewer instances of setting than mixed pairs in gossip conditions (*estimate* = -0.435, *z-ratio* = -3.585, *SE* = 0.121, $p < 0.01$) and summary conditions (*estimate* = -0.411, *z-ratio* = -3.377, *SE* = 0.131, $p < 0.01$). NN pairs in summary conditions also had significantly fewer instances of setting than NS pairs in gossip conditions (*estimate* = -0.429, *z-ratio* = -3.533, *SE* = 0.131, $p < 0.01$).

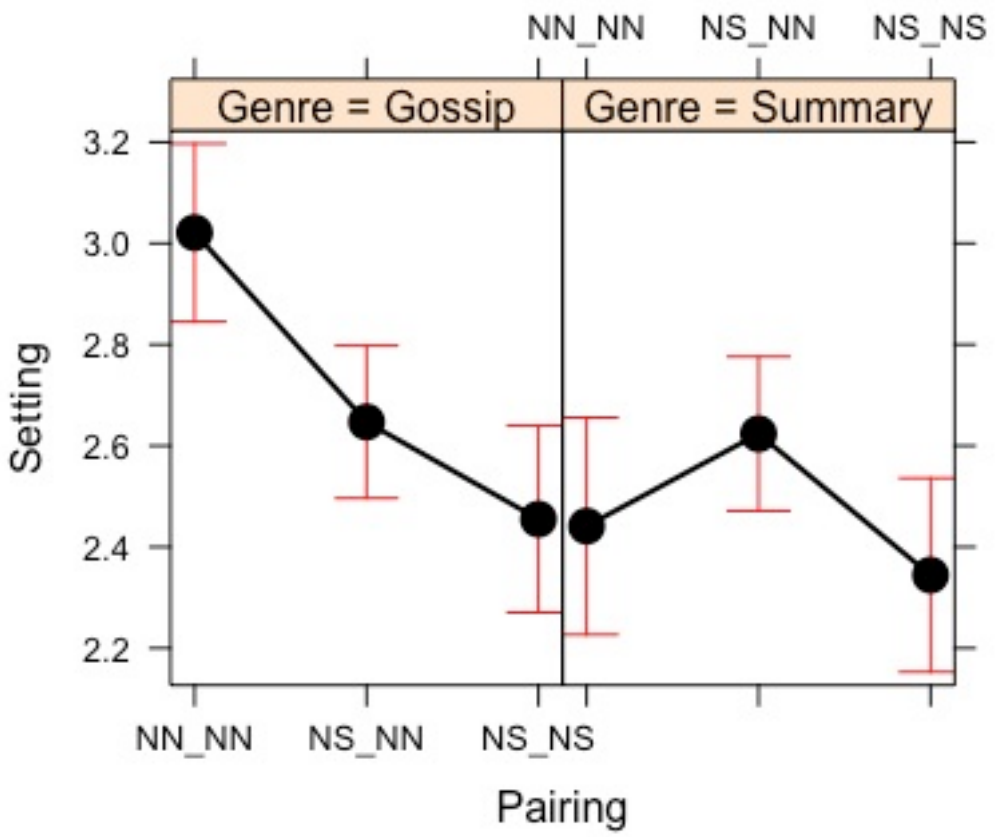


Figure 3.20. The instances of setting as a function of the interaction between genre and pairing. From this plot we see NN pairs appear to be the only pairing affected by genre.

In summary, the results for setting were very similar to the results for cause and effect relationships. Genre had a main effect such that setting occurred significantly more times in gossip conditions than summarizing conditions. Again, a significant main effect of nativeness showed native English speaker to have more instances of setting than non-native speakers. Pairing had a significant main effect such that NN pairs had significantly more instances of setting than NS pairs, however a multiple comparison did not confirm the model's prediction that NN also have more instances of setting than mixed pairs. Like cause and effect, there was a significant interaction for setting between genre and pairing, although this interaction showed the opposite relationship to cause and effect: NN pairs were significantly more likely to mention setting in gossip conditions than they were in summary conditions.

3.1.7 Other Factors of Interest

Three further factors were coded for in the event-indexing factor analysis. The factors that arose were interlocutors (the speaker and listener), source (where the interlocutors got their information from), and speaker attitude. These factors were coded for if an utterance did not clearly fit into any of the pre-existing event-indexing factor categories. The mean instances an interlocutor mentioned themselves or their partner (excluding their attitude) was 4.014 (*SD* = 6.326, *Mdn* = 2, *min* = 0, *max* = 30, *N* = 72). Source was mentioned a mean of 3.486 times (*SD* = 2.578, *Mdn* = 3, *min* = 0, *max* = 11, *N* = 72). Both interlocutors and source were mentioned fewer times than the least mentioned EI factor (goals and plans) so they were excluded from further analysis.

Speaker attitude was mentioned a mean of 12.139 times (*SD* = 8.404, *Mdn* = 10, *min* = 0, *max* = 38, *N* = 72). An utterance was included as speaker attitude if the speaker was making their own affective judgment about the information. A boxplot of the distribution by genre would suggest speaker attitude is more prominent in gossip (Figure 3.19).

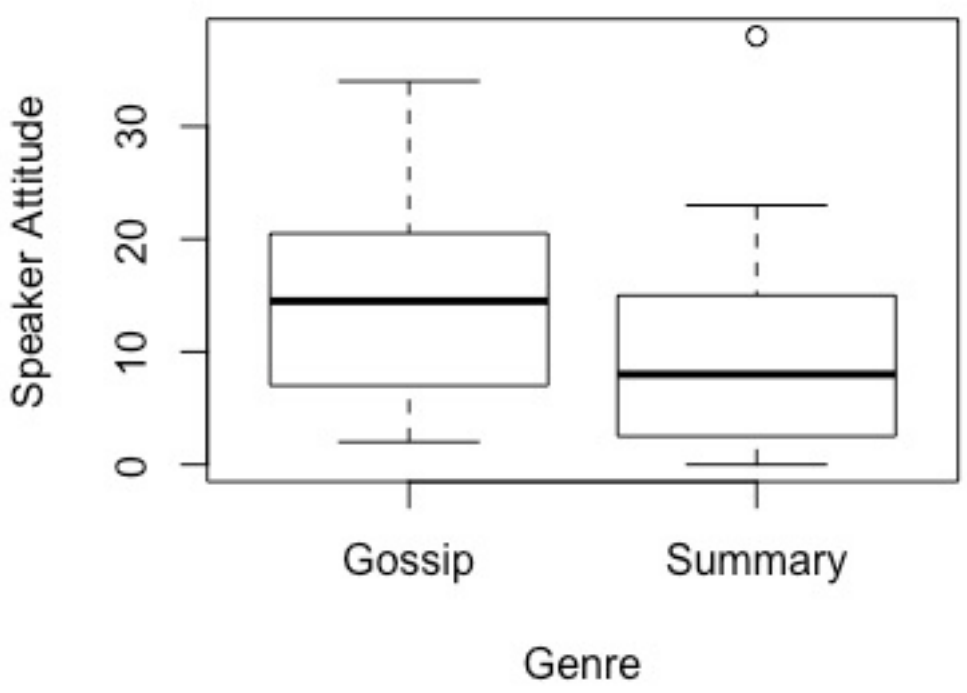


Figure 3.21. The distribution of the instances of speaker attitude by genre.

The best fitting model had speaker attitude as a function of the main effects of genre, nativeness, and pairing, and an interaction between pairing and genre (*AIC* with interaction= 569.245, *AIC* without interaction = 586.986), The intercept was significantly different from zero ($y = 2.464$, $z = 29.257$, $SE = 0.084$, $p < 0.001$). The main effect of genre confirmed the above observation that summary was less likely to have higher instances of speaker attitude (*estimate* = -1.037, $z = =6.299$, $SE = 0.165$, $p < 0.001$). This effect is illustrated in Figure 3.20.

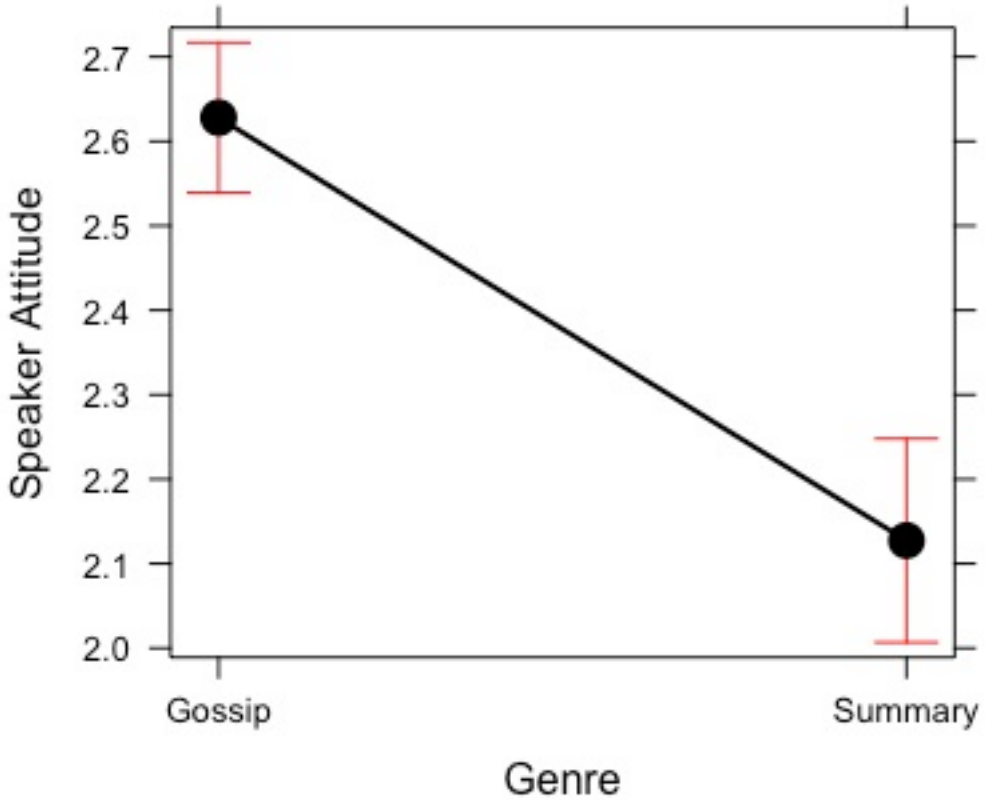


Figure 3.22. Instances of speaker attitude as a function of genre. Error bars represent 95% confidence intervals.

Like the previous EI factors, nativeness also had a significant effect such that the model predicted more instances of speaker attitude when the speaker was native speaker of English (*estimate* = 0.855, z = 6.722, *SE* = 0.127, $p < 0.001$) (Figure 3.21).

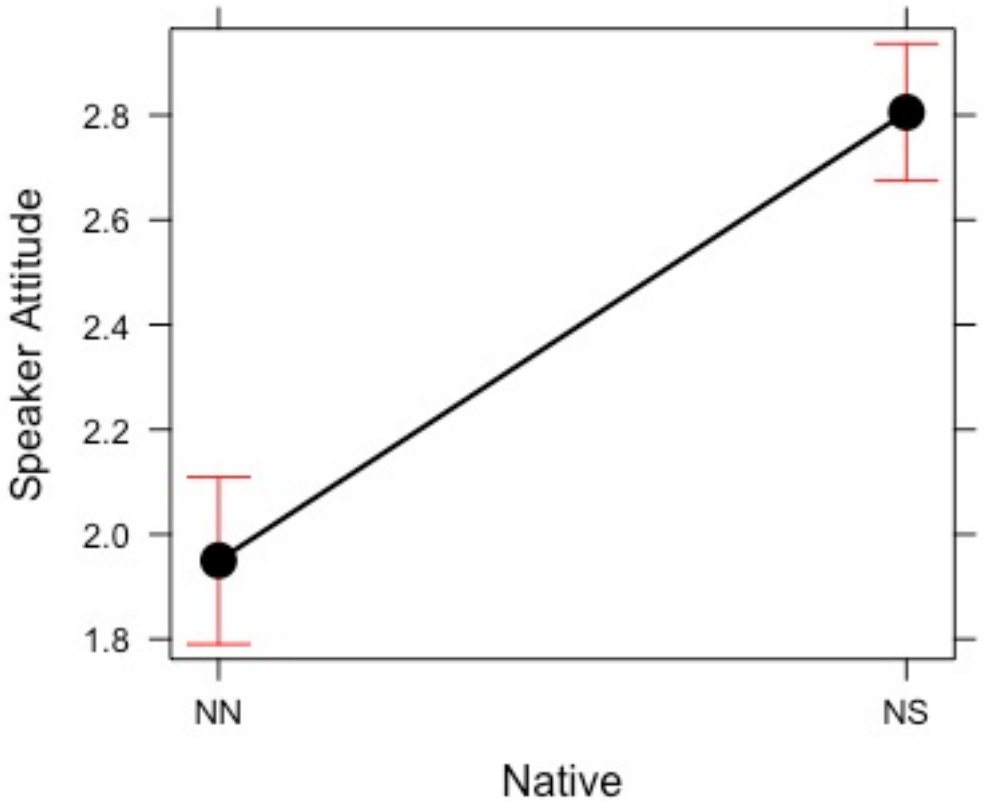


Figure 3.23. Instances of speaker attitude as a function of nativeness. Error bars represent 95% confidence intervals.

Pairing had a significant main effect with NN pairs having significantly less instances of setting than mixed pairs (*estimate* = -0.403, $z$ = -2.753, *SE* = 0.146, $p < 0.01$) or NS pairs (*estimate* = -0.388, $z$ = -2.330, *SE* = 0.167, $p < 0.05$), although a multiple comparison showed no significant interaction between any pairings. The effect predicted by the model is plotted in Figure 3.22.

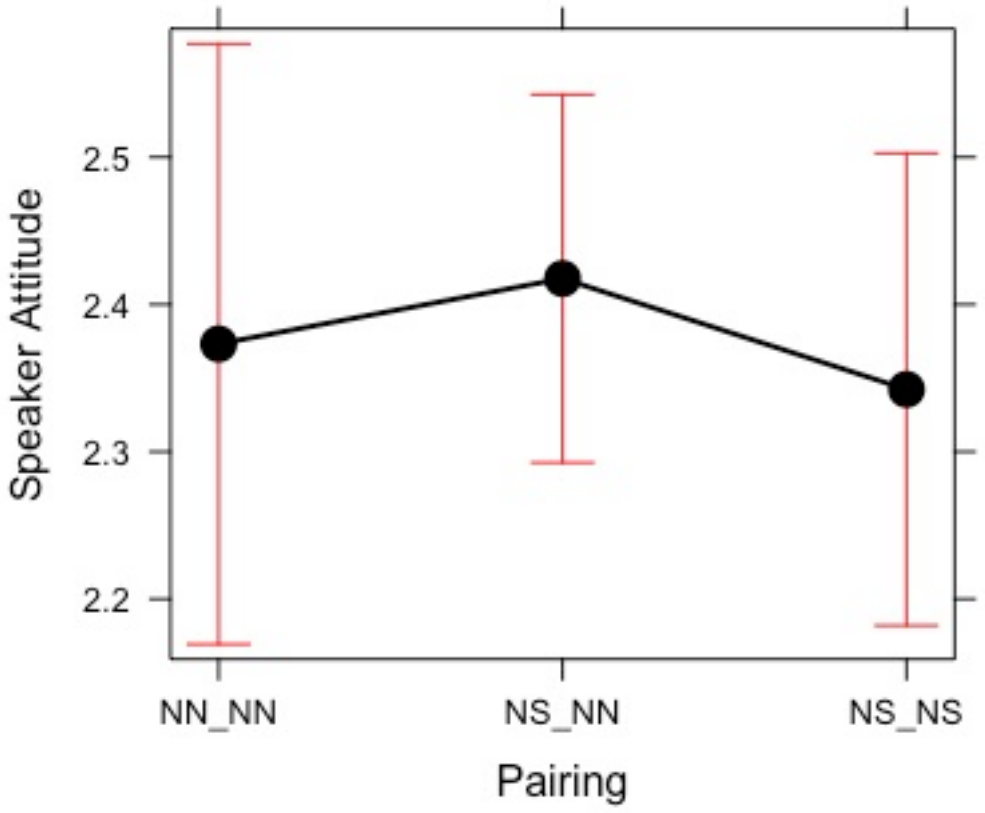


Figure 3.24. Instances of speaker attitude as a function of pairing. Error bars represent 95% confidence intervals.

Pairing did have a significant interaction with genre that withstood a multiple comparisons test (Figure 3.23). NN pairs in summary conditions have significantly less instances of speaker attitude than mixed pairs in gossip (estimate = -1.150, z-ratio = -7.091, SE = 0.162, p <.0001) and summary scenarios (estimate = -1.008, z-ratio = -6.101, SE = 0.165, p <.0001) and NS pairs in gossip (estimate = -1.504, z-ratio = -9.620, SE = 0.156, p <.0001) and summary scenarios (estimate = -1.182, z-ratio = -7.310, SE = 0.162, p <.0001). NN pairs in gossip have significantly more instances of speaker attitude than NN pairs in summary conditions (estimate = 1.037, z-ratio = 6.299, SE = 0.165, p <.0001) and less instances than NS pairs in gossip conditions (estimate = -0.467, z-ratio = -4.351, SE = 0.107, p <.001), although that effect is not clearly illustrated by the plot here. Lastly NS in gossip conditions have significantly more instances of speaker attitude than NS pairs in summary conditions ((estimate = 0.322, z-ratio =

3.134, SE = 0.103, p <.05)) or mixed pairs in summary conditions (with Gossip * NS-NS as the base level: estimate = -0.486, z-ratio = -4.578, SE = 0.108, p <.001)

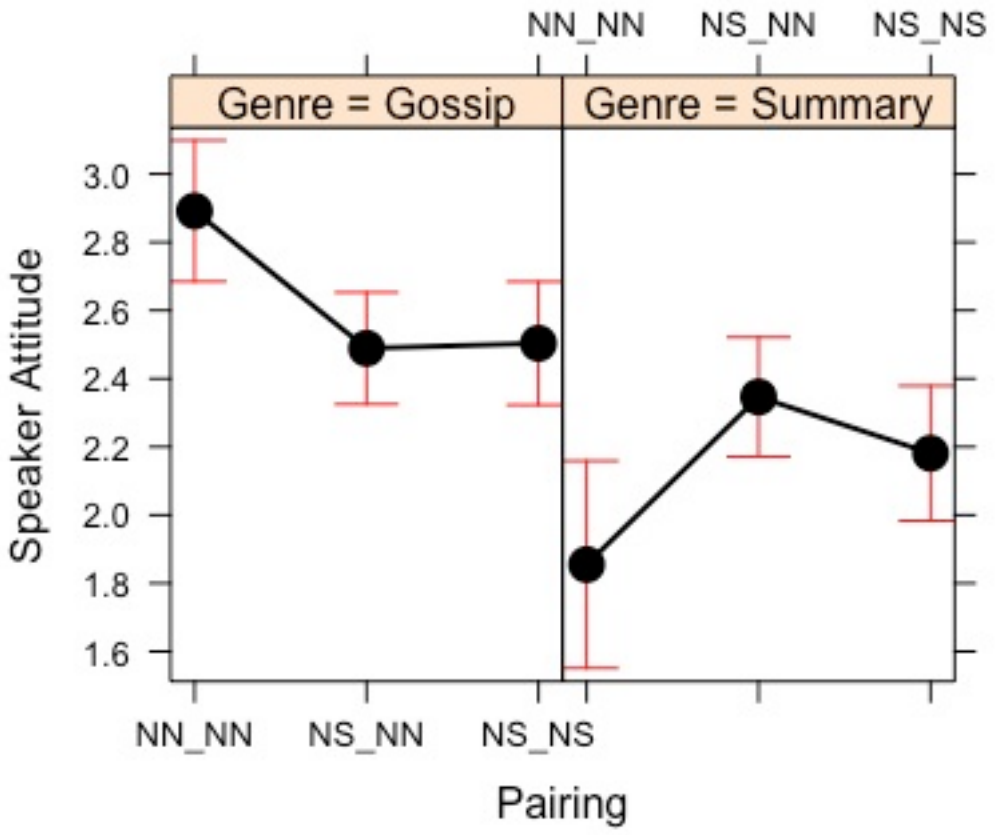


Figure 3.25. Instances of speaker attitude as a function of the interaction between genre and pairing. Error bars represent 95% confidence intervals.

In summary, utterances coded as an instance of source or interlocutor occurred less frequently than the least frequent confirmed event-indexing factor, time, and were therefore excluded from further analysis.

Instances coded as speaker attitude however occurred more frequently than time and were analyzed as a potential event-indexing factor. The results for speaker attitude are similar to cause and effect and setting. A main effect of genre showed speaker attitude occurred significantly more often in gossip conditions than summary conditions. As with every other event-indexing factor, a significant main effect of nativeness found native English speakers to

have more instances of speaker attitude than non-native speakers. There was a significant interaction between pairing and genre. The interesting interactions here are that speaker attitude is more likely to occur in gossip scenarios for both NN pairs and NS pairs than it is to occur in summarizing conditions for the same pairs.

3.2 Discussion of Event-Indexing Factors

The variable I was interested in was whether or not genre affected the prominence of event-indexing factors: protagonists/objects, cause/effects, goals/plans, time, setting; and whether or not dialogues or different genres would elicit new factors not previously accounted for in narrative studies.

Previous event-indexing studies on narratives have found protagonist/objects and time to be more closely tracked (Thierrault & Rinck, 2012). Protagonists/objects and time were the only two event-indexing factors that did not have a main effect of genre. It is possible this is because they are important to a dialogue regardless of genre. Furthermore, the production data gathered in this study shows protagonists and objects do have one of the highest occurrences in spoken dialogue, adding the evidence that they play a more significant role in situation model building. However, my data shows mentions of time occur less frequently in spoken dialogues than any other event-indexing. In contrast, cause and effect relations occur more frequently. A future comprehension study that uses a spoken dialogue like the ones collected in this study could shed light on the correlation between the number of times an event-indexing factor is produced by a speaker and how closely it is tracked by a comprehender.

With regard to spoken dialogue eliciting new factors, my results found that factors like interlocutors and source may occur in addition to event-indexing variables but they occur less than the less frequent EI variable (goals and plans) and thus do not seem as prominent as the other factors. This could suggest that not only are the five EI factors—protagonists/objects, cause/effect, goals/plan, setting, and time—important for narrative comprehension, they are also the most prominent factors in dialogue production. One new factor that did appear more often than the least frequent EI factor was speaker attitude. Like cause and effect relationships and setting, speaker attitude occurred significantly more frequently in gossip conditions; although it occurred in summary conditions as well suggesting this factor may be elicited by a spoken dialogue genre, not just gossip genres.

As for the effect of genre, as mentioned above, gossip elicited significantly more instances cause-effect relationships, setting, and speaker attitude. Summary elicited more occurrences of goals and plans. There was no significant difference between the number of instances of time or protagonists and objects in either gossip or summarizing conditions.

I can speculate reasons why gossip predicted higher instances of cause/effects, setting, and speaker attitude. Gossip often has an affective aspect to it (it can be "malicious", "juicy", etc.) and this may account for the increase in speaker attitude instances. For cause and effect relationships, a brief glimpse of the data shows participants explaining their attitudes towards Donald Trump (or other entities) are the result of his actions, thus possibly linking speaker attitude and cause and effect relationships. With regard to setting, many of the conversations shifted away from American politics and to speakers' own political views and beliefs. During this shift, the setting (political or national) was usually set up in length and sometimes co-occurred with cause and effect relationships to explain why the speaker felt a certain way. So it is possible that gossip scenarios allow speakers to discuss their own views more frequently and my participants often used a setting = cause, speaker attitude = effect structure when discussing their own political opinions. Further data collection and analysis would be needed to explore this phenomenon. Another possible route for research would be to investigate if using a less geo-political prompt would decrease the instances of setting in gossip conditions.

Regarding the effect of summarizing on instances of goals/plans, this may have something to do with the modality that speakers use when summarizing events within a specific time frame. Goals and plans are incredibly dependent on not being actualized. If they have been carried through they often take on a more cause and effect nature. Compare the following two quotes from participants:

"He said *he wasn't running to* promote his book" (goal/plan)
"I feel like *he was running as* joke and now he's stuck and doesn't know how to get out" (cause/effect)

In the summarizing condition, participants may be restricted by the timeframe of the article, and so even if events and actions have been actualized, participants may talk about them as if they are yet to happen and are still the planning phase in order to be consistent with the

timeframe of the article. In gossip conditions, participants are less restricted to a specific time frame and can talk about actions that have happened and justify their current beliefs, as suggested in the discussion on the high instances of cause and effect relationships in gossip conditions.

Ultimately the goal of this study was simply to see if genre affected the occurrences of certain event-indexing factors compared to others and this was confirmed in the present data. What specific factors motivate this difference is the topic for further study.

3.2.1 Other Predictors of Event-Indexing Factors

Although they were not the variables under investigation for this research question, it seemed appropriate to fit my models with the two other major manipulated variables included in this study: nativeness (English native speaker or not) and pairings (NS-NS, NN-NN, mix). Both these factors significantly improved the models for all of the event-indexing variables under interest with the exception of goals and plans (although this exception may be due to the small amount of data available for this factor). I will now discuss why nativeness and pairing may have been found to improve predictions for EI factors.

My models showed that for all EI factors, native English speakers were more likely to have a higher count of EI factors than non-native speakers. The fact that this effect stayed consistent for every EI variable would suggest it doesn't effect which EI variables are more prominent for different L1 speakers, rather that native English speakers had more instances of EI variables overall. I currently lack the data to further investigate this effect, but one possible explanation could be that native speakers have a faster rate of speaking in English than their non-native peers, allowing them to get more utterances and information, and therefore more instances of EI factors, into the ten minutes of dialogue.

As for pairings—two native speakers (NS-NS), mixed speakers (NS-NN), or two non-native speakers (NN-NN)—a positive effect for NN-NN pairs was found for protagonist, cause-effect relationships, and setting. This seems to contradict the previous positive effect for native speakers in our data and would require further investigation to tease apart. One route of investigation could be the preferred structures used by different speakers. For example, one NN-NN pair may have had higher counts for protagonists/objects because they used the following structure in their dialogue:

Speaker A: Donald Trump…he is a businessman.

Speaker B: He is very rich.
Speaker A: Yes. He is a celebrity.

Since every update to a protagonist was marked as a separate instance for this variable, this exchange would have 3 counts for protagonists. Compare this to a NS-NS exchange, which would only account for one count of protagonist and one of cause-effect:

Speaker A: He [=Donald Trump] is a very successful businessman. (protagonist)
Speaker B: Yeah but that doesn't mean you can run a country. (cause-effect)
Speaker A: True.

In the first exchange the speakers use a list-like structure to build a description of the protagonist together. Compare this to the second structure where the second speaker responds with a more generalized statement, rather than further specifying the protagonist. This could be one possible route for further research on the difference between NN-NN and NS-NS pairings.

Pairing did significantly interact with genre for several EI factors: protagonists/objects, goals/plans, setting, and speaker attitude. The only EI factors that showed a similar effect were setting and speaker attitude where NN-NN pairs were significantly more likely to mention these factors in gossip scenarios than summary conditions. These overlapping patterns may have something to do with the setting-cause, speaker attitude-effect described in the discussion on gossip as an effect in the preceding section. Another possible explanation for setting could be the limitations of summarizing an article compared to an un-restricted gossip prompt, a caveat discussed above with goals and plans. During the recorded conversations, several international students told their partners that they were unfamiliar with American politics during the dialogue. Some of them also mentioned they were unfamiliar with Indiana, the setting for one of the articles. The restriction to discuss an article set in an unfamiliar setting may have dissuaded them from using referencing setting as often in the summarizing condition. By contrast, in the gossip condition participants could pull information from wherever they wanted to, and many (native and non-native speakers) discussed their own political experiences in their native countries or the experiences of friends living abroad, possibly accounting for the significant increase in instances of setting. Further investigation into familiarity with a topic or the restrictiveness of a topic could

be one possible route for further investigation into this matter. Regardless, neither the type of pairing or English nativeness of the participants offer much of an explanation for the variance in event-indexing factor counts.

The only other potential pattern found in pair-genre interactions is that for the pairing affected, they often produced more of a given EI variable in gossip conditions: NS-NN had more instances of protagonists in gossip conditions than summarizing, NN-NN had more instances of setting and speaker attitude in gossip conditions than summarizing, and NS-NS pairs had more instances of speaker attitude in gossip than summarizing. The one exception to this pattern is NN-NN pairs produced more goal/plan instances in summarizing conditions than gossip. They may have been more sensitive to the main effect of genre on goals and plans described in the previous sub-section. However, the interaction between genre and pairing was unexpected and affected all types of pairs in an unsystematic way. It is possible that these interactions are a product of the independent pairs and participants included in our data set.

### 3.2.3 Summary of Event-Indexing Factors

In summary, my data shows speaker attitude occurs as frequently as, if not more than the five existing EI factors (protagonists/objects, cause/effects, goals/plans, setting, time). Speaker attitude also shows similar effects to setting and cause and effect. These two pieces of evidence may indicate that speaker attitude is just as important as other event-indexing factors in spoken dialogue. A future comprehension study that uses spoken dialogue could confirm or reject this.

Furthermore, the prominence of event indexing factors, specifically cause/effect, goals/plans, setting, and speaker attitude, could be predicted by genre. This suggests that the importance of these factors varies by genre. Protagonists/objects and time were not affected by genre. This could suggest they are important regardless of genre and lend support to previous studies that found protagonists/objects and time to be more closely tracked than other EI factors.

In this study an overall effect of nativeness was found such that native speakers had more instances of EI factors overall, although this could simply be due to their faster rate of speech. However this effect was somewhat contradicted by the main effect of pairing on protagonists/objects, cause/effect, and speaker attitude, where NS-NS pairs had significantly more instances of these EI factors than NS-NN or NS-NS pairs. Further investigation and data is required to tease apart this contradiction.

Lastly, pairing and genre often interacted and significantly affected how many instances of an EI factor were found for a particular pair in a specific genre. In this case, protagonists/objects, goals/plans, setting, and speaker attitude were all affected by this interaction. There did not appear to be a clear overlapping pattern across these factors as the pairs affected (NN-NN, NS-NN, or NS-NS) varied across factors and the resulting direction of change for EI instances in the genres also varied. This may have been due to the individual participants in each pair.

In the next chapter I will investigate more closely how different speakers bring different pre-existing knowledge to a conversation and how this in turn affects different pair types (NN-NN, NS-NN, NS-NS) as they navigate common ground and establish which information is important and shared in their conversation.

# 4. Common Ground and Alignment

## 4.1 Results

### 4.1.1 Statistical Analysis

The statistical analysis of results was carried out in R (version 3.3.1, R Core Team, 2016). The response variables were measures of common ground establishment and alignment and are defined in more detail in the following subsections. The manipulated predictor variables were Genre of dialogue: gossiping or summarizing an article, Nativeness: native English speaker (NS) or non-native English speaker (NN), and Pair Congruence: congruent pairings (NS with NS, NN with NN) or incongruent pairing (NS with NN, NN with NS). Due to high co-linearity between the two variables of interest, pairing and nativeness ($r = 0.82$, $p < 0.0001$), pairing was re-leveled as a factor with two levels: congruent pair (NN-NN or NS-NS) or incongruent pair (NS-NN), and renamed as congruence. A colinearity test between congruence and nativeness showed a correlation of 0 but with a *p*-value of 1. All measures of colinearity were taken using the pariscor.fnc() from the "languageR" R package ("languageR", Baayen, 2013).

Descriptive statistics were found using the summarise() function of the "dplyr" package ("dplyr", Wickham & Francois, 2016). I used generalized linear regression analysis with Poisson family specified in the glm() function of the lme4 R package ("lme4", Bates, Maechler, Bolker, & Walker, 2015). All models were fitted in the same way. The initial model started with the common ground or pragmatic factor under investigation modeled as a function of the three-way interaction between Congruence, Nativeness, and Genre.

The model fitting procedure was the same as it was described in the previous chapter.

### 4.1.2 Common Ground Feedback Cues

#### *4.1.2.1 Overall Feedback*

Common ground feedback was coded as explicit feedback cues made by a participant demonstrating understanding or confusion relating to the previous statement. The aspects of understanding and confusion are described in more detail in their respective subsections below. Here I look at the overall feedback cues (both understanding and confusion) given by the participants ($M = 69.639$, $SD = 26.228$, $Mdn = 64$, $min = 20$, $max = 133$, $N = 72$). The best model included overall common ground feedback as a function of a three-way interaction between congruence, nativeness, and genre (*AIC* congruence*nativeness*genre: 1014.823, *AIC* congruence*nativeness + genre: 1090.220, *AIC* congruence*genre + nativeness*genre:

1059.446). The intercept was significantly different from zero ($y$ = 4.200, $z$ = 118.7856, *SE* = 0.035, $p < 0.001$)

There was a significant main effect of genre: participants gave significantly less feedback overall in summary conditions than gossip conditions (*estimate* = -0.220, $z$ = -4.152, *SE* = 0.053, $p < 0.001$), as illustrated in Figure 4.1.

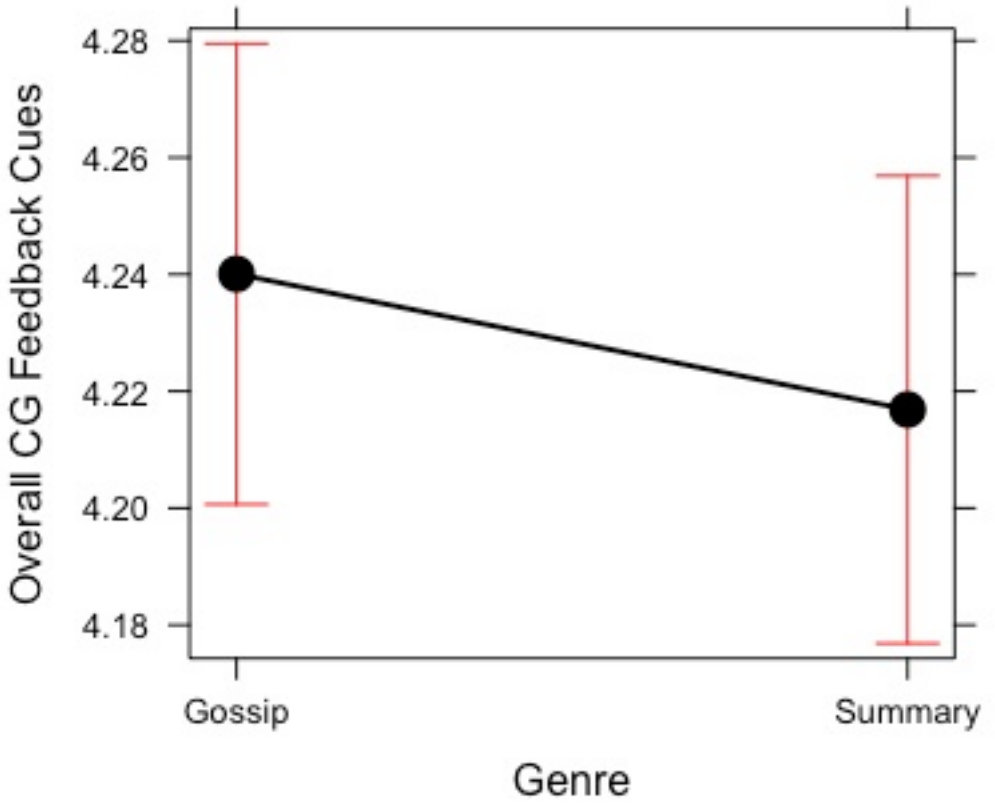


Figure 4.1. Overall common ground feedback cues as a function of the main effect of genre. Error bars represent 95% confidence intervals.

The model showed a significant interaction between nativeness and congruency (when compared to congruent NN, incongruent NN: *estimate* = 0.329, $z$ = 5.964, *SE* = 0.055, $p <$ 0.0001; congruent NS: *estimate* = -0.1123, $z$ =-2.183, *SE* = 0.051, $p < 0.05$, incongruent NS: *estimate* = 0.138, $z$ = 2.350, *SE* = 0.059, $p < 0.05$). Multiple comparison analysis confirmed NN in incongruent pairings (i.e. NS-NN) gave significantly more feedback than NN in congruent settings (NN-NN) (*estimate* = -0.334, *z-ratio* = -8.103, *SE* = 0.041, $p < .0001$). Contrast this to

NS who had no significant difference in the amount of feedback between congruent (NS-NS) and incongruent pairs (NS-NN) (*estimate* = 0, *z-ratio* = -0.005, *SE* = 0.042, *p* = 1.00). Within congruent pairs (NS-NS or NN-NN) NN gave significantly less feedback overall than NS (*estimate* = -0.166, *z-ratio* = -4.559, *SE* = 0.036, $p < 0.0001$), however they gave significantly more feedback overall than NS within incongruent pairs (NS-NN) (*estimate* = 0.169, *z-ratio* = 3.601, *SE* = 0.047, $p < 0.01$). This interaction is illustrated in Figure 4.2

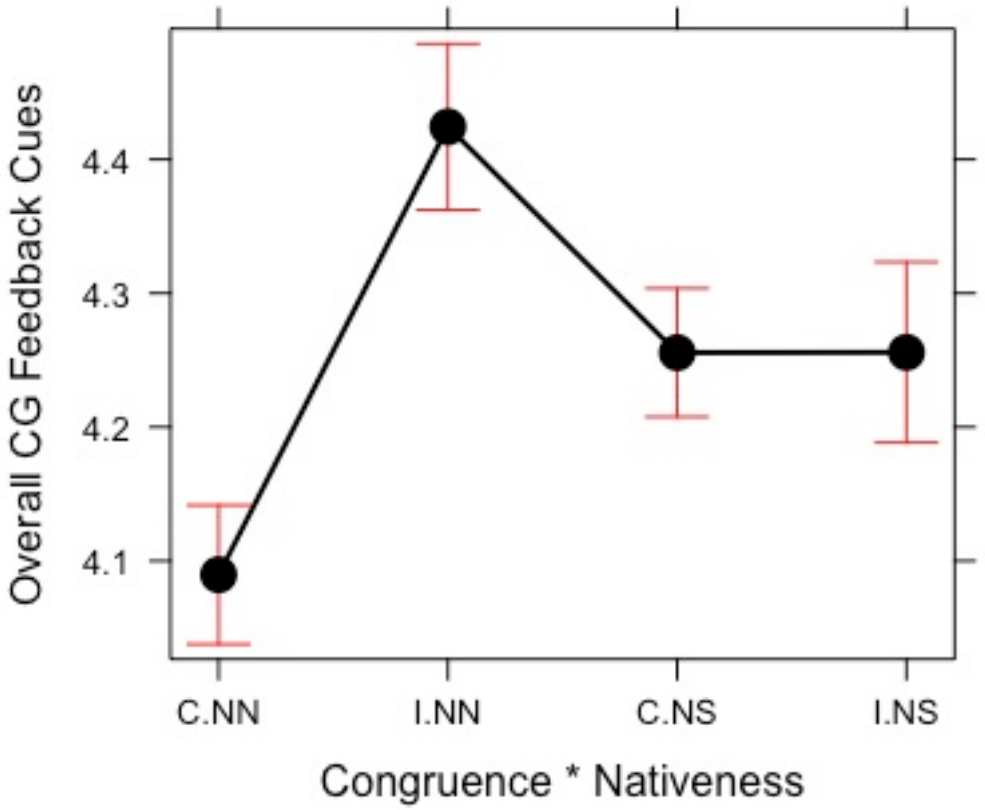


Figure 4.2. Overall common ground feedback cues as a function of the interaction between congruency and nativeness. "C" stands for congruent pair (NN-NN for non-native speakers, and NS-NS for native speakers), "I" stands for incongruent pair (NS-NN). The error bars represent 95% confidence intervals.

The model also predicted a significant three-way interaction between congruency, nativeness, and genre (see Figure 4.3) such that NS in congruent pairs gave significantly more feedback cues in summary conditions than they did in gossip conditions (*estimate* = 0.557, *z* =

7.713, *SE* = 0.072, $p < 0.0001$). A multiple comparison confirmed this effect congruent NS gave more cues in summary than gossip (with summary as the baseline: *estimate* = -0.336, *z-ratio* = -6.872, *SE* = 0.049, $p < 0.0001$). Furthermore, within summary conditions NS gave significantly more cues if they were in a congruent pair (NS-NS) than if they were in an incongruent pair (NS-NS) (*estimate* = 0.249, *z-ratio* = 4.179, *SE* = 0.060, $p < 0.001$), but within gossip conditions NS gave significantly less cues if they were in a congruent pair than if they were in an incongruent pair (*estimate* = -0.250, *z-ratio* = -4.178, *SE* = 0.060, $p < 0.001$).

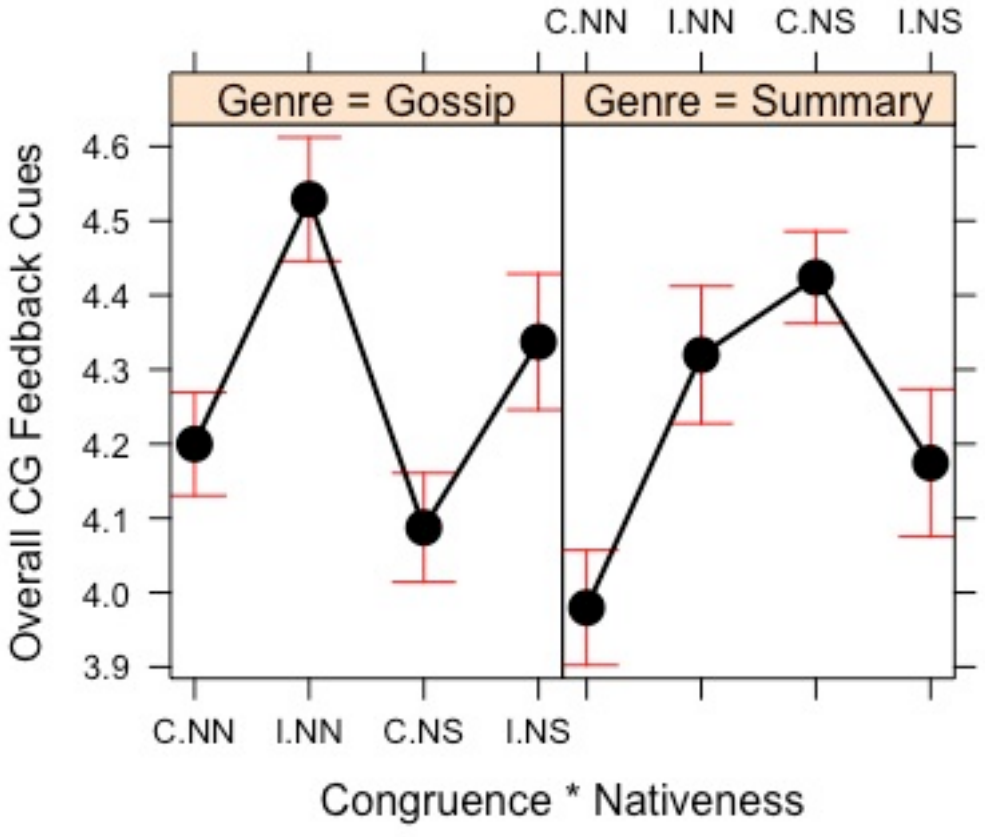


Figure 4.3. Overall common ground feedback cues as a function of a 3-way interaction between congruency, nativeness, and genre. "C" stands for congruent pair (NN-NN for non-native speakers, and NS-NS for native speakers), "I" stands for incongruent pair (NS-NN). The error bars represent 95% confidence intervals.

In summary, non-native English speakers give fewer overall common ground feedback cues than native English speakers when they are speaking to a non-native speaker, but more

overall CG cues when they are speaking to another non-native speaker. The amount of feedback cues native speakers gave in incongruent and congruent pairs was dependent on genre. A three-way interaction between congruency pair and genre suggests that when gossiping, native speakers give significantly fewer cues if they are in a congruent pair (NS-NS) than an incongruent pair (NS-NN), whereas in summary conditions, NS give significantly more cues if they are in a congruent pair than if they are in an incongruent pair. Genre also had a main effect such that participants gave significantly more overall CG cues in gossip conditions than they did in summarizing conditions. I will now provide analysis for these CG cues when they are divided into instances of successful common ground establishment (i.e. understanding) and instances of unsuccessful common ground establishment (i.e. confusion).

*4.1.2.1. Understanding*

Instances of feedback denoting understanding were defined as every instance of head nods, affirmative expressions (e.g. “yeah”, “mhmm”, “okay”) and reiterations (Speaker A: “I’m Japanese-” Speaker B: “Japanese?” Speaker A: “Right”). The mean number of feedback cues given by a participant was 50.167 ($SD$ = 25.511, $Mdn$ = 46.5, $min$ = 6, $max$ = 111, $N$ =72).

The best fitting model included cues of understanding as a function of a three-way interaction between congruency, nativeness, and genre (*AIC* Congruency * Nativeness * Genre: 1039.903, *AIC* Congruency * Nativeness + Genre: 1086.098, *AIC* = 1261.661). Genre had a main effect such that participants gave more cues of understanding in gossip conditions than summarizing conditions (*estimate* = -0.150, $z$ = -2.205, $SE$ = 0.068, $p < 0.05$).

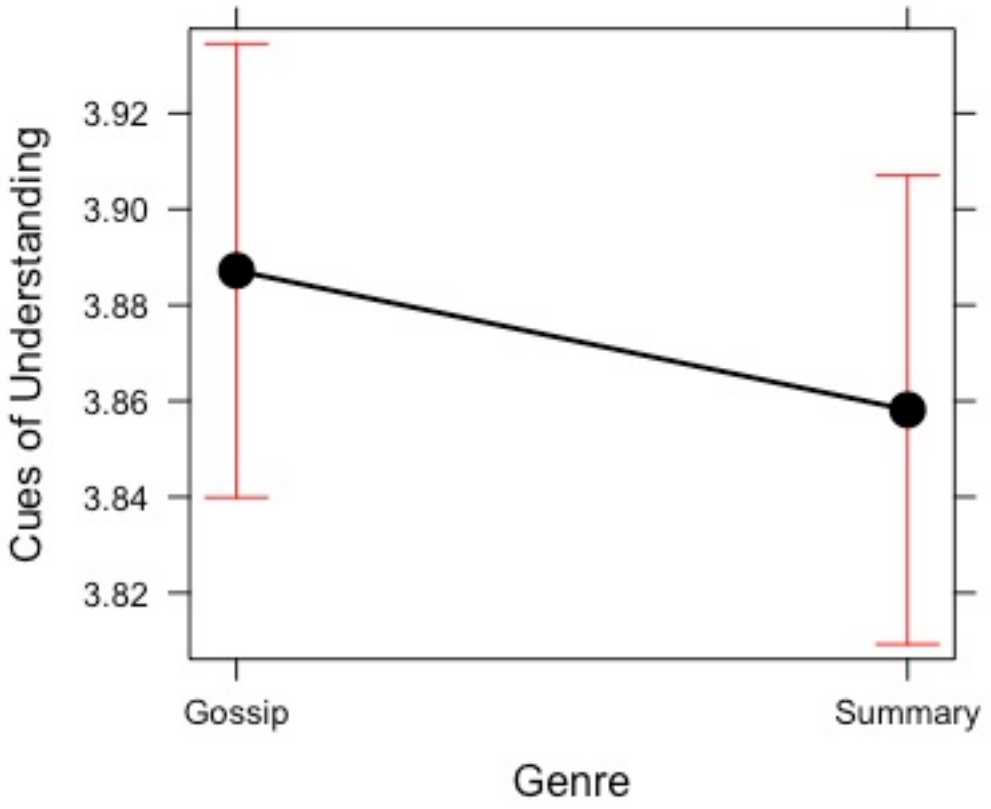


Figure 4.4. Feedback cues indicating understanding as a function of the main effect of genre. Error bars represent 95% confidence intervals.

The model showed a significant interaction between congruency and nativeness such that congruent NN gave more cues of understanding than congruent NN (*estimate* = 0.629, *z* = 9.444, *SE* = 0.067, $p < 0.0001$) or incongruent NS (*estimate* = 0.313, *z* = 5.138, *SE* = 0.061, $p < 0.0001$), but congruent NN were not significantly different from incongruent NS (*estimate* = 0.113, *z* = 1.468, *SE* = 0.077, $p > 0.1$). A multiple comparison confirmed that NN gave significantly less cues of understanding in congruent pairs (NN-NN) than in incongruent pairs (NS-NN) (*estimate* = -0.596, *z-ratio* = -12.057, *SE* = 0.049, $p < 0.0001$). NS had a reverse effect such that they gave significantly more cues of understanding in congruent pairs (NS-NS) than incongruent pairs (NS-NN) (*estimate* = 0.466, *z-ratio* = 8.707, *SE* = 0.054, $p < 0.0001$). Furthermore, was no significant difference in the feedback given by NS in incongruent pairs and NN in congruent pairs (*estimate* = -0.068, *z-ratio* = -1.174, *SE* = 0.058, $p > 0.5$), or in the

feedback given by NN in incongruent pairs and NS in congruent pairs (*estimate* = 0.062, *z-ratio* = 1.401, *SE* = 0.044, $p > 0.1$). This effect is illustrated in Figure 4.5.

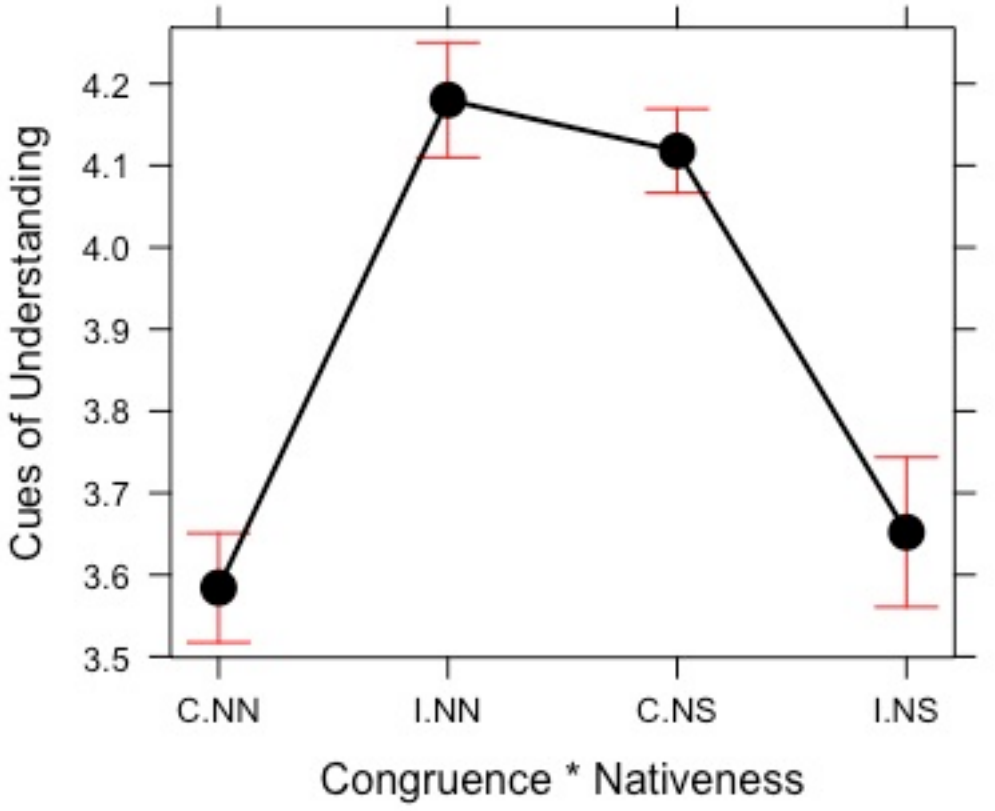


Figure 4.5. Feedback cues indicating understanding as a function of the interaction between congruency and nativeness. "C" stands for congruent pair (NN-NN for non-native speakers, and NS-NS for native speakers), "I" stands for incongruent pair (NS-NN). The error bars represent 95% confidence intervals.

The model also predicted a significant three-way interaction between congruency, nativeness, and genre, such that NS in congruent pairs gave significantly more cues of understanding in summary conditions than in gossip conditions (*estimate* = 0.442, *z* = 5.147, *SE* = 0.08592, $p < 0.0001$). A multiple comparison confirmed this effect (with gossip as the baseline: *estimate* = -0.336, *z-ratio* = -6.872, *SE* = 0.049, $p < 0.0001$) and found the reverse to be true of NN in incongruent pairs, i.e. NN gave significantly less cues of understanding in summary conditions than in gossip conditions (with gossip as the baseline: *estimate* = 0.209, *z-ratio* =

3.303, *SE* = 0.063, $p < 0.05$). As was also seen in the interaction between nativeness and congruency, NN participants in congruent pairs had no significant different from NS in incongruent pairs in gossip conditions (*estimate* =0.112, *z-ratio* = 2.183, *SE* = 0.051, $p > 0.1$), although congruent NN did have significantly less feedback cues than incongruent NS in summary conditions (*estimate* = -0.195, *z-ratio* = -3.033, *SE* = 0.064, $p < 0.05$). These effects are illustrated in Figure 4.6.

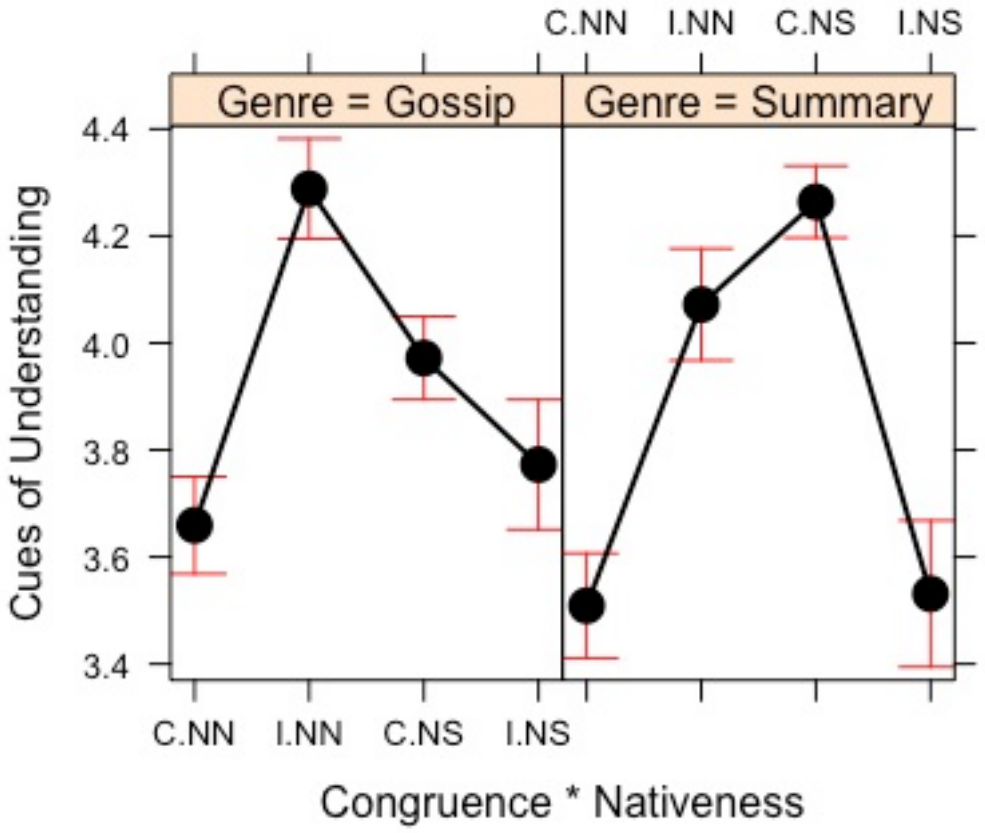


Figure 4.6. Feedback cues indicating understanding as a function of the interaction between nativeness, congruency, and genre. "C" stands for congruent pair (NN-NN for non-native speakers, and NS-NS for native speakers), "I" stands for incongruent pair (NS-NN). The error bars represent 95% confidence intervals.

In summary, NN participants gave more cues of understanding if they were in an incongruent pair than if they were in a congruent pairs. NS had a reverse effect of congruency and gave more feedback cues if they were in a congruent pair than in an incongruent pair. The

interaction between congruency and nativeness predicted no significant differences between congruent NN and incongruent NS or incongruent NN and congruent NS, although the former became significantly different in summarizing conditions in the three-way interaction. Furthermore, incongruent NN and congruent NS were oppositely affected by genre in the three-way interaction: incongruent NN gave significantly more cues of understanding gossip conditions than summarizing conditions and the opposite was true for congruent NS. Genre also had a main effect such that more cues of understanding were given in gossip conditions than summarizing conditions.

*4.1.2.1. Confusion*

Feedback cues indicating confusion were coded as every instance of eye contact breaks, gestures like brow furrows and frowns, expressions like "huh", "umm", and clarifications (e.g. Speaker A: "He's even worse" Speaker B: "Donald Trump?" Speaker A: "No Mike Pence"). The mean number of cues of confusion for a given participant was 19.472 ($SD$ = 14.238, $Mdn$ = 18, $min$ = 0, $max$ = 51, $N$ = 72). The best fitting model had cues of confusion as a function of the three-way interaction between congruence, nativeness, and genre *(AIC* Congruence * Native * Genre: 867.755, *AIC* Congruence * Native + Genre: 898.947, *AIC* Congruence * Genre + Native * Genre: 1060.3964). There is reason to believe the intercept is significantly different from zero ($y$ = 3.326, $z$ = 60.789, $SE$ = 0.055, $p < 0.0001$).

Genre had a main effect such that participants produced less cues of confusion in gossip conditions than summary conditions (*estimate* = -0.326, $z$ = -3.861, $SE$ = 0.085, $p < 0.001$), as illustrated in Figure 4.7.

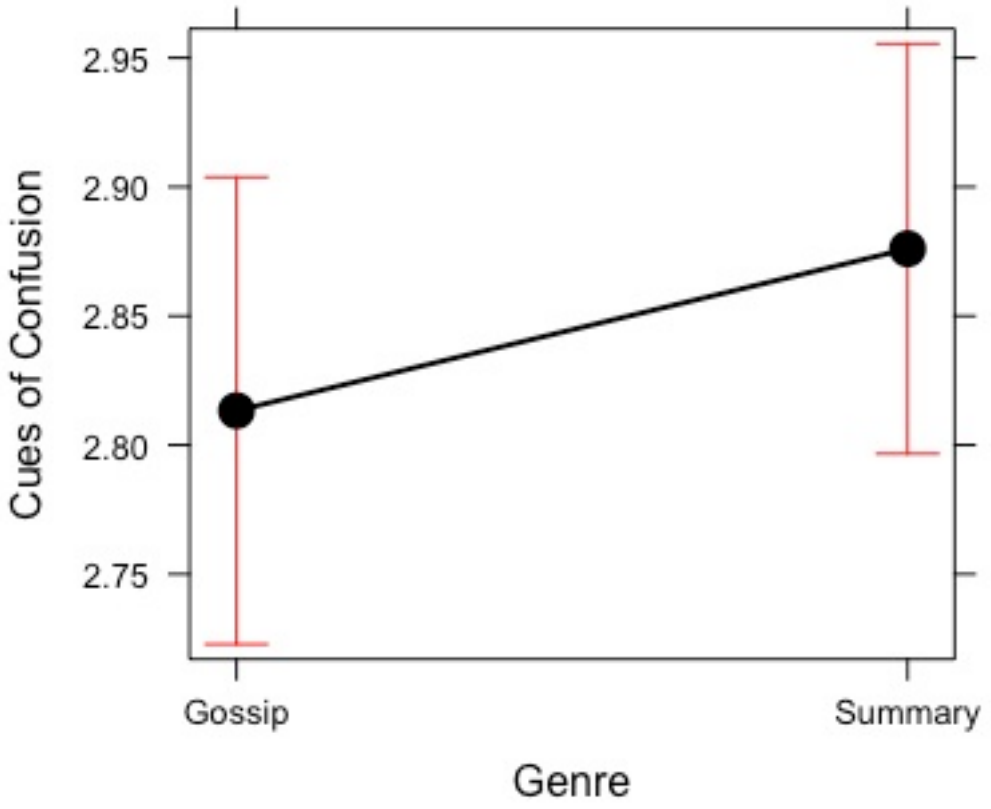


Figure 4.7. Feedback cues indicating confusion as a function of the main effect of genre. Error bars represent 95% confidence intervals.

The interaction between congruence and nativeness predicted NN in congruent pairs (NS-NN) to produce significantly more cues of confusion than NS in congruent pairs (*estimate* = -1.454, $z$ = -11.566, *SE* = 0.126, $p < 0.0001$) and NN speakers in incongruent pairs (*estimate* = -0.339, $z$ = -3.714, *SE* = 0.107, $p < 0.01$). A multiple comparison analysis confirmed these effects (congruent NN: incongruent NN: *estimate* = 0.268, *z-ratio* = 3.343, *SE* = 0.080, $p < 0.05$; congruent NN: congruent NS: *estimate* = 0.971, *z-ratio* = -11.880, *SE* = 0.971, $p < 0.0001$). It also showed NS in incongruent pairs had significantly more instances of confusion than NS in congruent pairs (with congruent NS as the baseline: *estimate* = -1.270, *z-ratio* = -14.662, *SE* = 0.087, $p < 0.0001$), NN in incongruent pairs (with incongruent NN as the baseline: *estimate* = -0.567, *z-ratio* = -6.666, *SE* = 0.567, $p < 0.0001$), and congruent NN (*estimate* = -0.299, *z-ratio* = -4.515, *SE* = 0.066, $p < 0.0001$). Lastly incongruent NN had significantly more instances of

confusion than congruent native pairs (*estimate* = 0.703, *z-ratio* = 7.208, *SE* = 0.098, $p < 0.0001$). The effect shown by the model is illustrated in Figure 4.8.

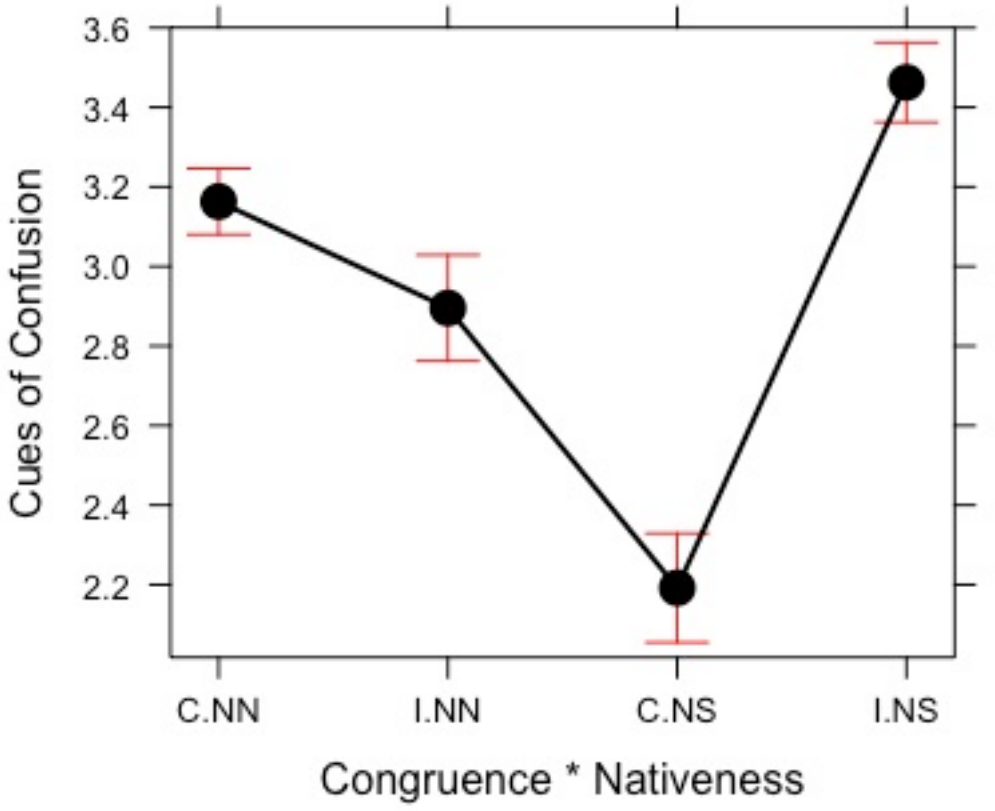


Figure 4.8. Feedback cues indicating confusion as a function of the interaction between congruence and nativeness. Error bars represent 95% confidence intervals.

The three-way interaction in the model predicted NS in congruent pairs would produce significantly more confusion cues in summary conditions than gossip conditions (*estimate* = 0.967, *z* = 5.915, *SE* = 0.163, $p < 0.0001$). A multiple comparisons analysis confirmed this effect (with summary as the baseline: *estimate* = -0.336, *z-ratio* = -6.872, *SE* = 0.049, $p < 0.0001$). Furthermore, NN in congruent pairs showed the opposite effect of congruent NS such that as congruent NN produced significantly less confusion cues in summary conditions than gossip conditions (with summary as baseline: *estimate* = 0.220, *z-ratio* = 4.152, *SE* = 0.053, $p < 0.001$). In turn, there is no significant difference between the amount of confusion cues produced by congruent NN participants and incongruent NS participants in gossip conditions (*estimate* = -

0.138, *z-ratio* = -2.350, *SE* = 0.059, $p > 0.1$). There was also no significant difference between incongruent NN participants and congruent NS participants in summary conditions (*estimate* = -0.104, *z-ratio* = -1.836, *SE* = 0.057, $p > 0.5$). The effect of the three-way interaction predicted by the model is illustrated in Figure 4.9.

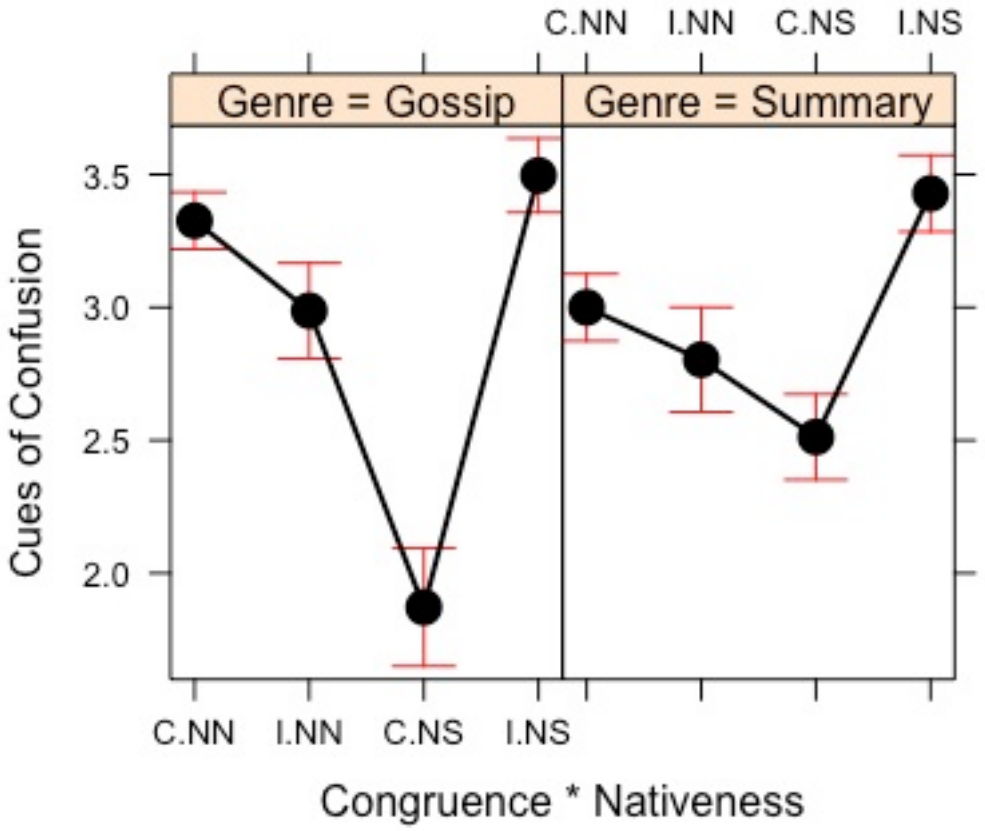


Figure 4.9 Feedback cues indicating confusion as a function of the three-way interaction between congruence, nativeness, and genre. Error bars represent 95% confidence intervals.

In summary, the interaction between congruence and nativeness predicted NS in incongruent pairs (NS-NN) to produce more cues of confusion than any other participant group and NS in congruent pairs (NS-NS) to produce the few cues of confusion than any other participant group, with no difference between NN in congruent or incongruent pairs. Adding genre to the interaction eliminated some of these effects: incongruent NS did not produce significantly more confusion cues than congruent NN in gossip conditions and congruent NS did not produce significantly less confusion cues than incongruent NN in summary conditions.

Furthermore, congruent NS pairs produced significantly more confusion cues in summary conditions than in gossip conditions. The model did show a main effect of genre such that summary conditions had more instances of confusion cues than gossip conditions.

4.1.3 Reference

The mean number of times a single participant introduced an antecedent that was referred to by both speakers more than was 2.319 (*SD* = 1.791, *Mdn* = 2, *min* = 0, *max* = 7, *N* = 72). The mean number of times any of those referents were then repeated by a single participant in the conversation was 2.5 (*SD* = 2.214, *Mdn* = 2, *min* = 0, *max* = 11, *N* = 72). The mean number of times any those referents became more unique was 1.431 (*SD* = 1.677, *Mdn* = 1, *min* = 0, *max* = 9, *N* = 72). When the unique reference and agreed upon reference were summed (total references) the number of times a single participant referred back to any antecedent was 3.931 (*SD* = 3.055, *Mdn* = 3, *min* = 0, *max* = 16, *N* = 72).

The best fitting model had total references as a function of the interaction between congruence and nativeness (*AIC* congruence * nativeness: 361.199, *AIC* congruence + nativeness: 364.058), with an intercept significantly different from zero (*estimate* = 1.166, *z* = 10.229, *SE* = 0.114, $p < 0.0001$). The model showed congruent NS had more instances of reference than congruent NN (*estimate* = 0.516, *z* = 3.583, *SE* = 0.144, $p < 0.001$). Multiple analysis showed confirmed that congruent NS had significantly more references than congruent NN (*estimate* = -0.516, *z-ratio* = -3.583, *SE* = 0.144, $p < 0.01$). Congruent NS also had more significant references than incongruent NN (where congruent NS is the baseline: *estimate* = -0.4778, *z-ratio* = -2.640, *SE* = 0.181, $p < 0.05$) and incongruent NS (*estimate* = 0.556, *z-ratio* = 2.980, *SE* = 0.186, $p < 0.05$). No other significant differences were found. The effect is illustrated in Figure 4.10.

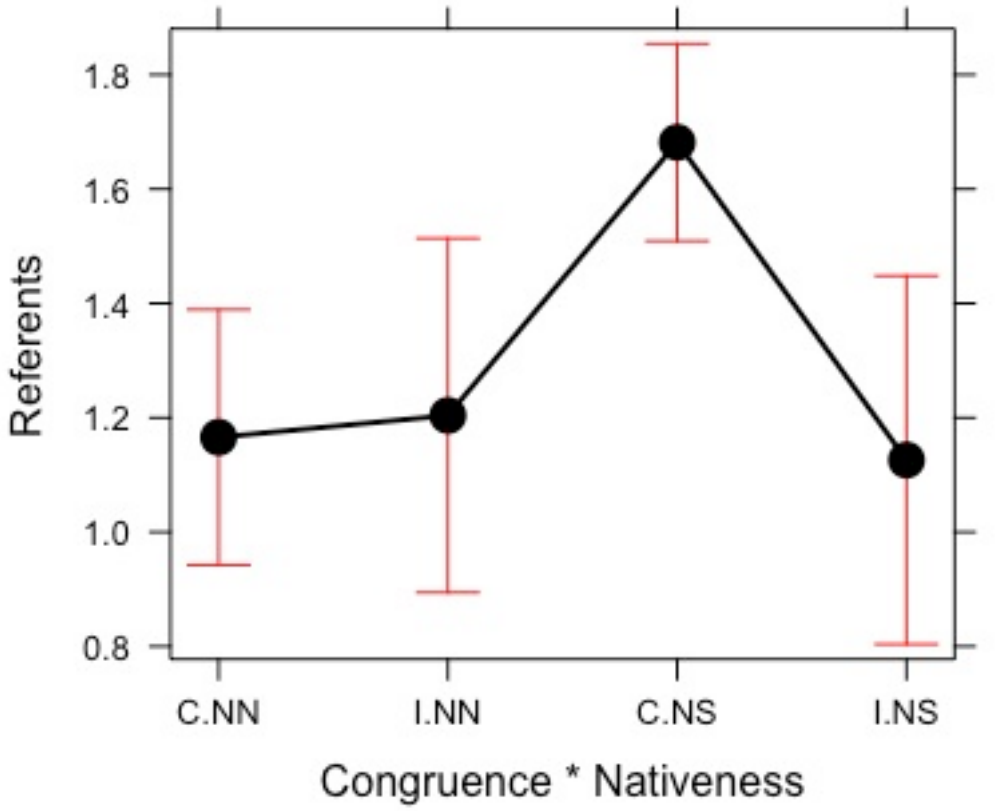


Figure 4.10. Unique and repeated references modelled as a function of the interaction between congruence and nativeness. Error bars represent 95% confidence intervals.

In summary, congruent native speakers produced significantly more unique and repeated references than any other condition.

4.1.4 Reduced Linguistic Information

Using less semantic information to represent more semantic information is another cue that participants are relying on a shared common ground. These instances were coded when an ambiguous phrase was understood by a listener (determined by a cue of understanding such as a head nod), but did not contain enough information to be understood as a stand-alone proposition. The mean amount of times a participant made a linguistically reduced statement was 2.528 (*SD* = 2.627, *Mdn* = 2, *min* = 0, *max* = 18, *N* = 72).

The best fitting model had reduced linguistic information as a function of the main effects of congruence, nativeness, and genre, with an interaction between congruence and genre (*AIC* congruence * genre: 300.948, *AIC* congruence + genre: 308.280).

Nativeness had a main effect such that native speakers produced significantly more reduced linguistic utterances than non-native speakers (*estimate* = 0.355, $z = 2.360$, *SE* = 0.151, $p < 0.05$)(see Figure 4.11)

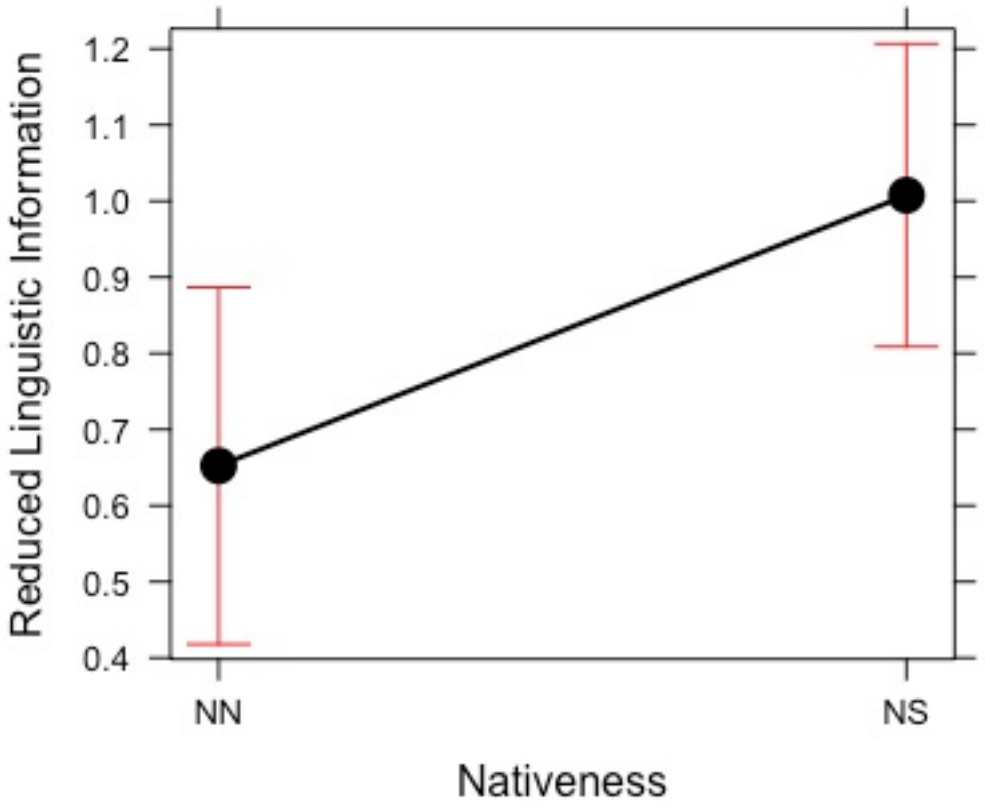


Figure 4.11. Reduced linguistic information as the function of the main effect of nativeness. Error bars represent 95% confidence intervals.

Congruence had a main effect such that reduced linguistic utterances occurred significantly more frequently in congruent conditions than incongruent (with congruent as the baseline: *estimate* = -1.142, $z = -4.108$, *SE* = 0.278, $p < 0.0001$) (see Figure 4.12).

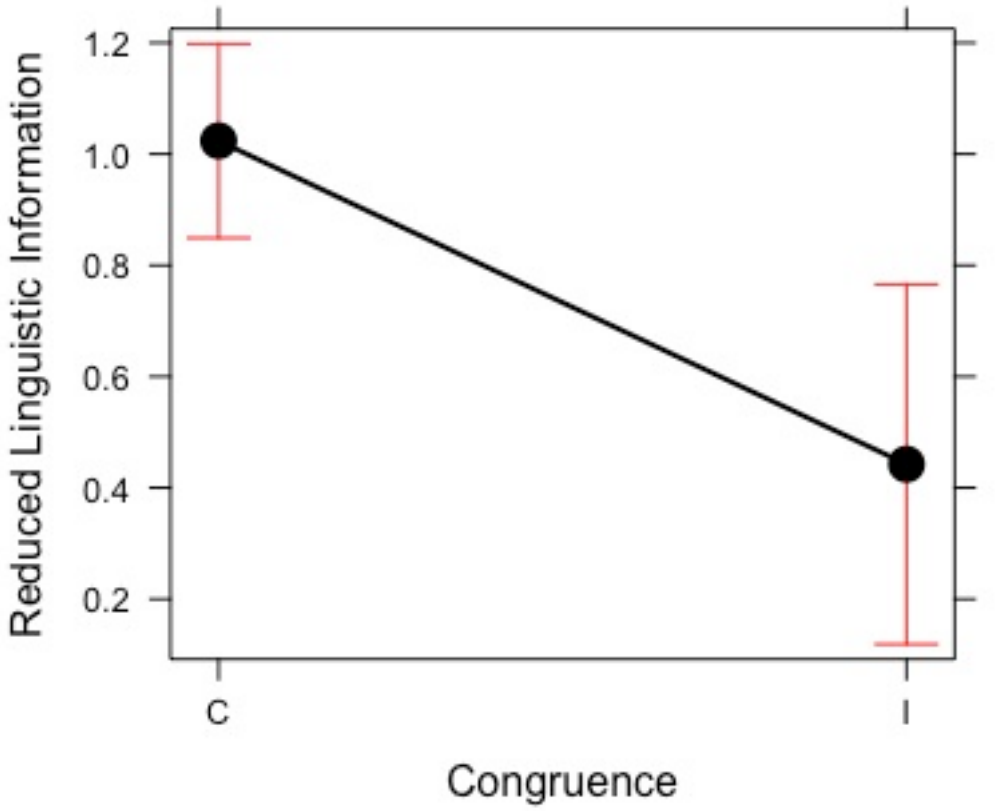


Figure 4.12. Linguistic utterances with reduced information as a function of congruence. "C" represents NN-NN or NS-NS pairs, "I" represents NS-NN pairs. Error bars represent 95% confidence intervals.

Genre had a main effect such that more instances of linguistically reduced utterances occurred in gossip conditions than summary conditions (with gossip as the baseline: *estimate* = -0.652, $z$ = -3.697, *SE* = 0.176 $p < 0.001$) (see Figure 4.13).

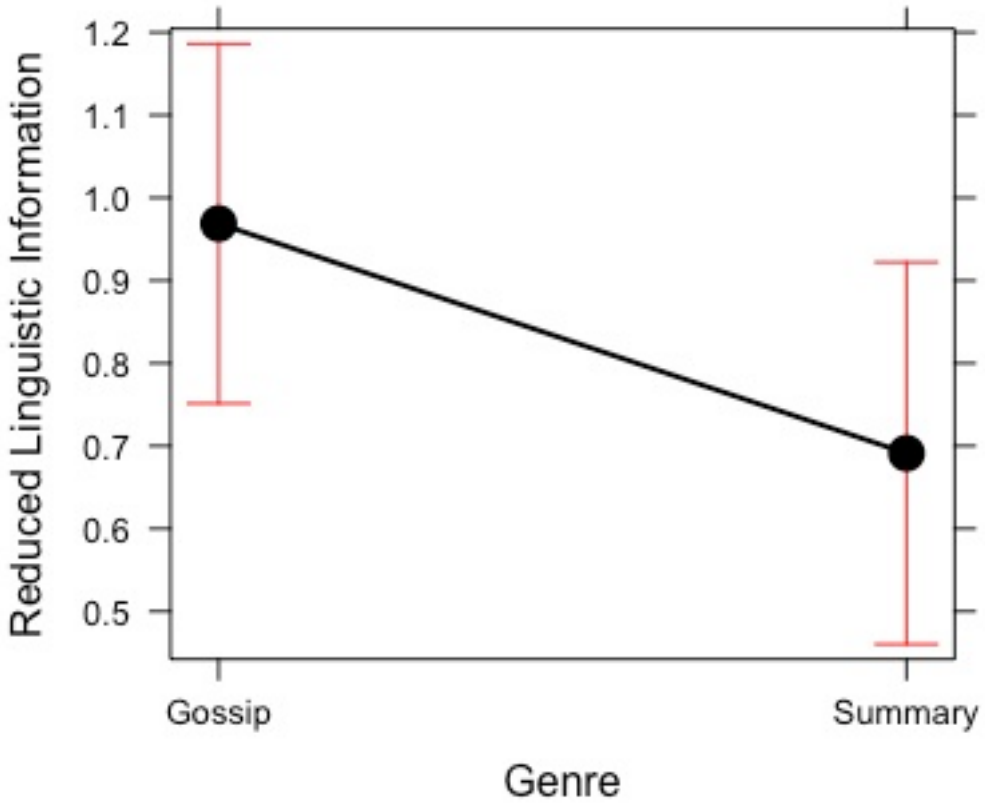


Figure 4.13. Linguistic utterances with reduced information as a function of genre. Error bars represent 95% confidence intervals.

There was a significant interaction between congruence and genre such that congruent gossip scenarios showed significantly more instances of reduced linguistic information than incongruent summary conditions (*estimate* = 1.122, $z = 3.004$, $SE = 0.373$, $p < 0.01$). A multiple comparison confirmed that congruent gossip scenarios have significantly more linguistically reduced utterances than incongruent summary conditions (*estimate* = 0.672, *z-ratio* = 2.939, $SE = 0.229$, $p < 0.05$). Congruent gossip conditions also have more linguistically reduced utterances than incongruent gossip conditions (*estimate* = 1.142, *z-ratio* = 4.108, $SE = 0.278$, $p < 0.001$) and congruent summary conditions (*estimate* = 0.651, *z-ratio* = 3.697, $SE = 0.176$, $p < 0.01$). No significant difference was found between congruent summary conditions and incongruent summary conditions (*estimate* = 0.0206, *z-ratio* = 0.083, $SE = 0.249$ $p > 0.5$). This interaction is illustrated in Figure 4.14.

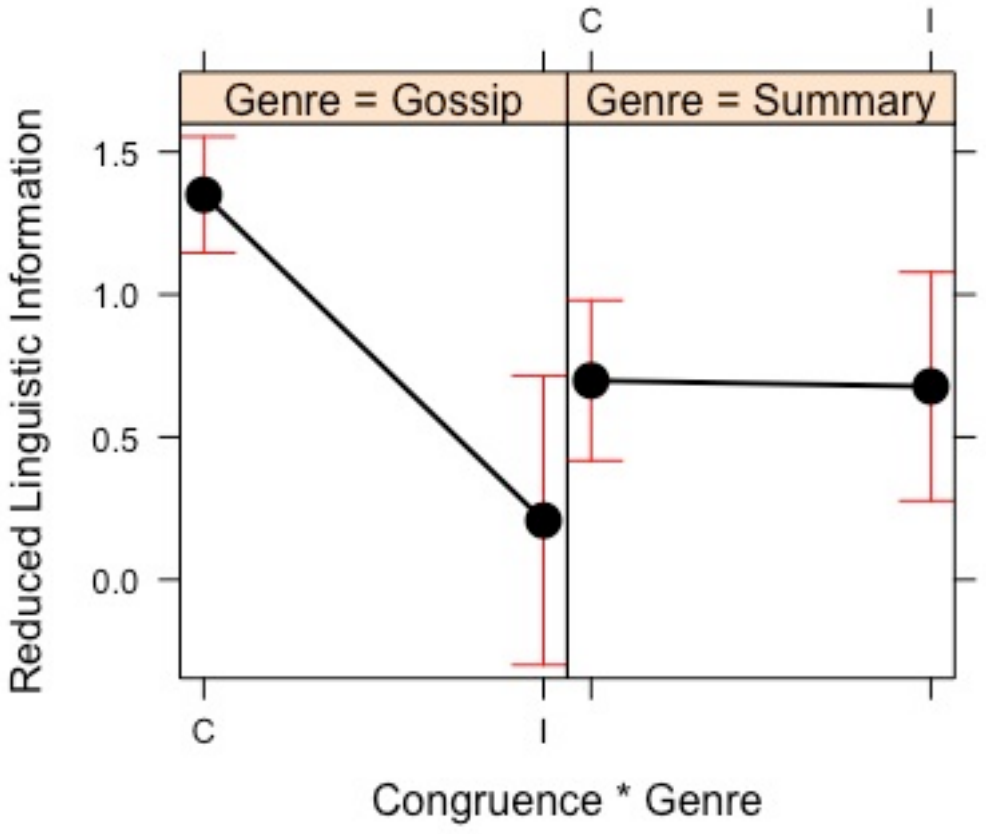


Figure 4.14. Linguistic utterances with reduced information as a function of the interaction between congruence and genre. Error bars represent 95% confidence intervals.

In summary, native English speakers produce significantly more linguistically reduced utterances (to express more semantic content) than non-native English speakers. These utterances occur significantly more frequently in gossip conditions and congruent conditions. However, the main effect of congruence disappears in summary conditions when congruence and genre interact.

4.1.5 Alignment

Alignment was coded as every time a specific linguistic structure co-occurred in both participants (e.g. Speaker A: "Yeah, and I feel like, …" Speaker B: "Yeah, and I feel like…"). I only coded the initial instance of the structure, not every time it was used by the participants. The mean number of occurrences of alignment for a participant was 2.472 (*SD* = 1.703, *Mdn* = 2, *min* = 0, *max* = 6, *N* = 72). The model with the best fit had alignment modelled as a function of the

main effects of congruence, nativeness, and genre, with an intercept significantly different from zero (*estimate* = 1.033, $z = 7.609$, $SE = 0.136$, $p < 0.0001$).

## 4.2 Discussion

### 4.2.1 Common Ground Factors

The variables of interest here was how the interaction between congruency and nativeness affected common ground establishment. The response variables I looked at here were explicit feedback cues indicating successful or unsuccessful common ground establishment, the overall amount of feedback given, and establishing agreed upon and unique labels for antecedents.

#### *4.2.1.1 Feedback Cues*

A goal of the present study was to investigate the differences, if any, in how participants with different linguistic and cultural backgrounds established common ground compared to participants with the same backgrounds. In this subsection I will discuss the overt feedback cues participants produced indicating successful common ground establishment (cues of understanding) or unsuccessful establishment (confusion cues). I will also look at the overall feedback cues produced as this may reveal differences in how much participants engaged in common ground establishment. The manipulated variables of interest here were whether or not the participant was a Canadian native English speaker (NS or NN) and whether a speaker was speaking to someone with the same background as them (i.e. NN in a NN-NN pair or NS in a NS-NS pair), referred to here as congruent, or whether they were in an incongruent pairing (NN or NS in NN-NS pair).

More overall feedback cues were produced in incongruent pairs for both NS and NN. A possible explanation of this is that interlocutors make fewer assumptions about someone with a different linguistic and cultural background, or interlocutors have less overlap in pre-existing knowledge, and therefore must put more effort into ensuring mutual intelligibility.

When feedback cues were divided into cues of understanding and confusion and analysed as a function between the interaction of nativeness and congruency, cues of confusion and understanding appeared to occur in complimentary distribution: incongruent NN and congruent NS produced significantly more instances of understanding and significantly less cues of confusion than congruent NN and incongruent NS (see Figure 4.17 for a side by side comparison of these effects).

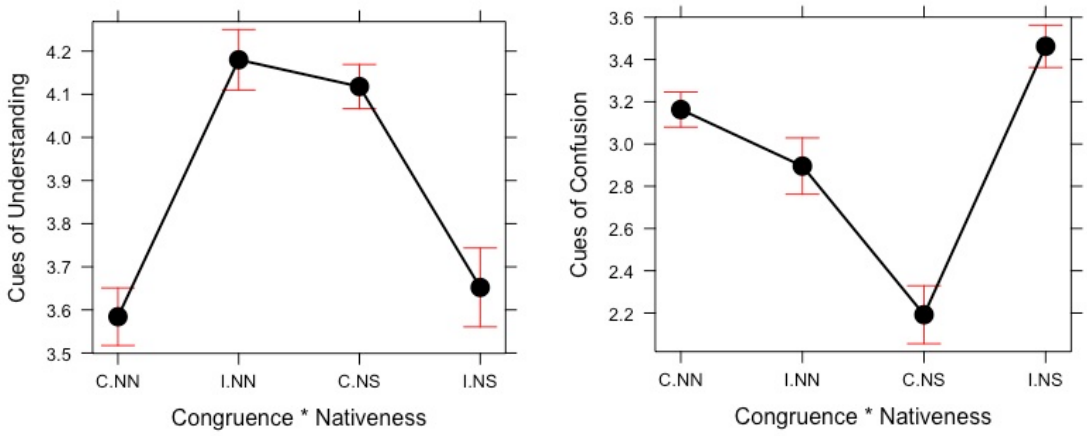


Figure 4.15. The distributions of cues of understanding and cues of confusion modelled as the function of the interaction between congruence and nativeness.

Furthermore, NN and NS showed opposite effects across congruency. If a NS spoke to another NS they showed less confusion and more comprehension. If a NS spoke to a NN they showed less comprehension and more confusion. NS however showed more comprehension when speaking to a NS and less confusion. By contrast when NN spoke to another NN they showed less comprehension and more confusion. Since all speakers were required to complete the task in English it is possible that accented speech added a level of difficulty for both NS and NN when speaking to NN. A second possible explanation of this could be that since both international and Canadian students have at least some experience interacting with Canadians. However, NN students did not have to share a cultural or linguistic background to be included in a NN-NN condition, therefore NN in NN-NN pairs and NS in NS-NN pairs could have had a similar lack of overlap in pre-existing knowledge from their fellow interlocutor.

Genre also had a main effect on feedback cues. Gossip conditions had significantly more cues of understanding and feedback cues in general than summary conditions and summary conditions had significantly more confusion cues than gossip conditions. This could be because summary conditions restricted the topic of discussion to the events described in the article. One of the articles focused on Donald Trump's presidential bid in 2000 and may participants expressed that they were unaware he had run for president before 2016. Thus, being forced to discuss a topic they did not have any pre-existing knowledge of may have made establishing common ground more difficult. Furthermore, if we look at NS in congruent pairs, who showed

significantly less confusion cues than any other group, an interaction between genre, nativeness, and congruence shows that they produce significantly more confusion cues in summary conditions than in gossip conditions, and that the number of confusion cues produced by congruent NS was no longer significantly different from the amount produced by incongruent NN. This could possible be because the summarizing task was more difficult than the gossip task, even for the group that showed the least confusion overall.

Lastly, a three-way interaction between congruency, nativeness, and genre showed that there were more cues for understanding in gossip conditions for NN if they were speaking to a NS-NN. This may be evidence that the unrestricted topic in gossip and the ease of speaking to a native English speaker in an English-based task had an additive effect that contributed to better common ground establishment.

*4.2.1.2 Reference*

When native English speakers spoke to another native English speaker they repeated and modified referents significantly more than any other speaker-congruency group. There was no significant difference in referent establishment between NS in incongruent conditions, NN in incongruent conditions, and NN in congruent conditions. No other effects for referent establishment were found. It is important to note that this variable only looks at referents brought up and established in the current study. It did not include ambiguous reference that may have referred to pre-existing or cultural knowledge. This suggests NS in congruent conditions either spent more time and effort on establishing referents, and therefore establishing the common ground, or they were more successful at establishing referents and common ground than other groups.

I would like to point our here that in the video coding stage of the present study it became apparent that participants did not consistently refer to the same referents throughout the ten minutes of conversation in this task, rather they might go though 1-2 exchanges regarding a referent before moving onto a new topic. This is reflected in the low number of initial mentions of antecedents given by any participant as it reflects the actual number of antecedents referred to in the conversation. Thus, the data reported here might be task dependent. A different study design where speakers are not allowed to change their topic of conversation may lead to more nuanced results regarding how linguistic and cultural background affect establishing referents.

4.2.2 Pragmatic Factors

Certain pragmatic behaviours also indicate common ground establishment. In this analysis I measured two of these variables: reduced linguistic information and alignment. Reduced linguistic information is when less linguistic information (i.e. more ambiguous or generalized utterances) are used to describe more semantic content. The type of alignment I looked at here was when both speakers in a dialogue began to use the same linguistic structures in their utterances. Again, the interaction between nativeness and congruency was the manipulated variable of interest here.

*4.2.2.1 Reduced Linguistic Information*

My analysis found that native speakers reduced linguistic information significantly more often than non-native speakers did. This could be due to the fact that speakers with similar linguistic and cultural backgrounds should have a greater overlap in pre-existing knowledge, therefore they may not need to explicitly communicate and establish this knowledge. Non-native speakers may have not experienced this same overlap since their only shared experience was that of being international students in Canada. They did not have to have the same heritage language or culture. However, congruent pairs (i.e. NS-NS pairs and NN-NN pairs) also reduced linguistic information significantly more often than incongruent pairs (NN-NN) so it is possible international students did benefit from some shared experience. No significant interaction was found between nativeness and congruence though so further research is required to tease apart these nuances.

Genre also had an effect on linguistic reduction. Speakers were more likely to reduce linguistic information in gossip conditions than they were in summary conditions. Furthermore, genre interacted with congruency such that in gossip scenarios congruent pairs and more instances of linguistic reduction than incongruent pairs. There was no difference between congruent and incongruent pairs in summary conditions. I would like to once again suggest that the restrictiveness of the summary of the condition compared to the gossip condition might have had something to do with this effect, since participants were forced to spend at least some time discussing the specific events described in their articles, regardless of how familiar they (or their partner) were with these events beforehand, and thus would have had to be more explicit in their statements if their partner had never heard of the event before. In gossip conditions participants had more flexibility to select topics related to Donald Trump that were more common in general knowledge or at least more familiar to them.

*4.2.2.2. Alignment*

My analysis did not show any significant effects for alignment. One possible explanation of this could be that alignment happens equally across all conditions and so no significant differences will be found. However, the descriptive statistics showed very few instances of alignment coded in my data ($M = 2.472$, $SD = 1.703$, $Mdn = 2$, $min = 0$, $max = 6$, $N = 72$) so I am inclined to believe that this is due a design flaw in my experiment. It could be possible that by only coding word-for-word overlaps in structure, for example when a pair of participants both started using "[Agreement], *and I also feel that* …" as a way to initiate their turn to speak and change the topic I coded it as alignment, but I did not code instance of [Agreement] + [Speaker Attitude]. Furthermore, I often relied on instances of alignment to co-occur with each other and therefore I may have missed instances of alignment that happened over the course of several exchanges. Using a transcript analysis rather than a video analysis may improve the measure of this variable in my data.

4.2.3 Summary of Common Ground and Alignment

Although I did not find a significant effect for alignment in my data, my variable of interest, the interaction between congruence and nativeness, did have a significant effect on the feedback cues given by participants and referent establishment in their dialogues. Furthermore, congruence and nativeness had main effects on linguistic reduction. This suggests that not only does linguistic and cultural background of a participant plays a role in how they establish common ground in a dialogue, the shared experiences of participants also affects the common ground. In the next chapter I will discuss how these results and my event-indexing results shed light on how dialogue is modulated by the "baggage" participants bring with them to a conversation.

# 5. General Discussion

This present study was based on the assumption that participants do not show up to dialogues as blank slates. Rather they bring baggage such as pre-existing cultural models and contextual knowledge (Brown-Schmidt, 2012; van Dijk, 2014; Thierrault & Rinck, 2012). This pre-existing knowledge combined with contextual factors such as social norms, power relations, and partner feedback will all have a top-down influence on an interlocutor's interpretation of incoming linguistic information (Brennan, Galati, & Kuhlen, 2010; Brennan & Hanna, 2009). Furthermore, participants monitor not only their own understanding of the conversation, but also the knowledge shared between them and their fellow interlocutor when establishing common ground (Keysar, Barr, & Horton, 1998; Brennan, Galati, & Kuhlen, 2010).

In this study, I examined how contextual factors like genre and dialogue (as opposed to written narrative comprehension) influenced the information tracked in a conversation by measuring the event-indexing factors produced in a dialogue. I also examined how differences in pre-existing knowledge affect how participants with different cultural and linguistic backgrounds modified the common ground. I will now summarize how my findings contribute to these topics.

Contextual factors like the genre of discourse could influence what information participants incorporate into their understanding of that discourse. Event-indexing (EI) models have shown five factors to be important in comprehending written narratives: protagonists/objects, cause/effect, goals/plans, time, and setting (Zwaan, 2014; 2015). I investigated whether these factors would be equally as important within the context of spoken dialogues, and if spoken dialogues would elicit new factors equally important to the conversation.

To do this I collected the frequency of utterances produced that provided new information regarding an EI factor. Utterances that modified information not accounted for in the five EI factors were coded as a new factor. I found utterances about speaker attitude emerged as frequently as the five existing EI factors. Furthermore, previous studies using written narrative comprehension have shown protagonists/objects and time are tracked more closely (Thierrault & Rinck, 2012). Speakers in my study produced instances of protagonists/objects, cause/effect and speaker attitude more frequently than other EI variables, and time was produced less frequently. I suggest this shows not only does spoken dialogue elicit a new factor not found in written narrative studies but the weights of event-indexing factors vary between spoken dialogues and written narratives. This suggests that contextual effects like the medium of discourse have a top-

down effect on which aspects of the incoming linguistic information are incorporated into understanding. Furthermore, I found the genre of dialogue had an even more nuanced effect on the distribution of EI factors.

My study included two conditions of spoken dialogue: gossiping about Donald Trump with a partner or summarizing articles about Donald Trump with a partner. I found that cause and effect, setting, and speaker attitude occurred significantly more frequently in gossip conditions. This may have been due to the geo-political nature of the topic since many participants used a "setting as cause, speaker attitude as effect" structure when discussing their political views. An example of this is listed in (5.1).

(5.1) "I grew up in Alberta so I was shocked when the NDP won"

By contrast, summarizing conditions had significantly more instances of goals/plans than gossip conditions. This may be because the events described in an article are frozen within that time frame. Thus relations that may have been discussed as cause and effects in gossip conditions had to be discussed in summarising conditions as potential goals/plans. This modality often shifted these events to being closer to a goal/plan (5.2) than cause/effect (5.3).

(5.2) "He said *he wasn't running to* promote his book"
(5.3) "I feel like *he was running as* joke and now he's stuck and doesn't know how to get out"

While further investigation is required to tease apart the effects of topic and modality within genres, the evidence in my study does show a main effect of genre on several EI factors. This is at the very least evidence that participants will behave differently in different genre conditions. I would like to suggest that genre has a top down influence that modulates which information is relevant to understanding. Therefore, the context and style of discourse would affect the mental representation a participant forms. Genre may even interact with other contextual cues.

In this study pairing and genre often interacted and significantly affected how many instances of an EI factor were found for a particular pair in a specific genre. Pairing had three conditions: non-native English speakers speaker with another non-native speaker (NN-NN),

native speakers speaking with another native speaker (NS-NS), and a mix of the two (NN-NS). In this case, protagonists/objects, goals/plans, setting, and speaker attitude were all affected by this interaction. NS-NS pairs had more instances of speaker attitude in gossip conditions than summarizing conditions. Mixed pairs had significantly more instances of protagonists/objects in gossip than summary conditions. NN-NN speakers were most affected by the interaction. They had significantly more instances of setting and speaker attitude in gossip conditions than they did in summarizing, and they showed significantly more instances of goals/plan in summarizing than they did in gossip.

No clear overlapping pattern across these pairs and genres so this may have been due to individual differences within our pairs. However, this could also suggest the pre-existing cultural knowledge speakers bring to a conversation, and the contextual overlap between that cultural knowledge amongst speakers may also have a top-down influence on the weight of EI variables. Further research is required to investigate the interaction between cultural background and genre has on EI factors. The present study does however address how these factors may have an influence on common ground establishment.

Many common ground studies use referents as a variable to measure common ground establishment (Brown-Schmidt, 2012; Brennan & Hanna, 2014; Gann & Barr 2014; Fukumura 2014). Participants will update the names of these variables and agree on them as they are entered into the common ground over the course of a discourse. My study found native speakers in a congruent pair (i.e. NS-NS) did this significantly more times than any other speaker-congruency condition. No significant differences between NS in incongruent pairs (i.e. when they spoke to a NN), NN in congruent pairs, and NN in incongruent pairs were found. This could suggest that cultural and linguistic overlap facilitate successful referent establishment, especially since congruent NN pairs did not necessarily have the same cultural background. However, in the present task many referents were only mentioned 1-2 times by both participants. Over the ten minutes of conversation, participants seemed to change topic quickly and no longer referred back to previously mentioned antecedents. Thus, a more traditional hidden map task with an added manipulation of congruence may shed further light on this effect.

The present study did find evidence suggesting contextual familiarity plays role on common ground establishment. Both NS and NN showed significantly more cues of understanding when speaking to a NS. I propose this may be because both are familiar with

interacting with Canadians on a day-to-day basis. Thus, their pre-existing expectations about their partner may have increased their reliance on assumption and possibly facilitated understanding. This is in line with previous common ground studies showing participants are affected by contextual familiarity (Brown-Schmidt, 2012; Gann & Barr, 2014) and partner specific cues (Brennan, Galati, & Kuhlen, 2010; Brennan & Hanna, 2009). Furthermore, NS in both congruent and incongruent pairs used less linguistic information to express greater semantic content significantly more times than NN in either congruency. Previous research claims this is a pragmatic effect that often indicates reliance on the common ground (Mills, 2014). I suggest that since the experiment took place in a Canadian context, Canadian students (i.e. NS) may have felt more comfortable relying on assumed pre-existing knowledge, thus allowing them to reduce the amount of explicitly stated linguistic information.

The contextual cue of genre may have also affected common ground. Participants in gossip conditions gave significantly more feedback cues of understanding than participants in summary conditions. Furthermore, participants in summary conditions gave significantly more feedback cues of confusion than participants in gossip conditions. This may be because summary conditions forced speakers to discuss a specific event, regardless of how much pre-existing knowledge they had of that event. By contrast, in gossip conditions speakers had more freedom to rely on pre-existing knowledge. It is possible speakers incorporated positive feedback cues to determine where their knowledge overlapped with their interlocutor, allowing them to determine their overlap in knowledge with their partner and facilitate understanding. Further research into the role feedback and pre-existing knowledge played in these situations is required.

However, a second result suggested a reliance on pre-existing knowledge as well. Congruence had a main effect where speakers in congruent pairs were able to reduce linguistic information more frequently than speakers in incongruent pairs. As explained above, this is an indication of speakers relying on common ground. Thus, since congruence pairs have more cultural overlap than mixed pairs, they may have had more overlap in pre-existing knowledge, allowing them to assume a certain amount of information in the common ground.

The study also found a significant interaction between genre and nativeness. NN participants showed significantly more feedback cues of understanding within gossip conditions if they were speaking with a NS. In the above paragraphs, I suggested gossip might allow for speakers to rely upon their overlap in world knowledge and having a NS as a partner might allow

for participants to rely on contextual knowledge of interacting with Canadians. If those speculations are correct, it is possible that when a NN speaks to NS in gossip conditions, there is an additive effect of contextual knowledge and overlaps in world knowledge, which facilitates common ground establishment.

In the present study I investigated how discourse context and pre-existing knowledge affect what information is tracked in a conversation and how successful participants will be at establishing this information in the common ground. I applied three broad manipulations to study these effects: 1) a cultural and linguistic manipulation between native English speaking Canadian students and non-native speaking international students; 2) whether or not participants speaking to a partner whose cultural/linguistic background was congruent with their own; and 3) whether or not speakers participated in an unrestrictive discourse style like gossip or a more restrictive genre like summarising. My study shows all three of these factors played significantly affected the EI factors participants mentioned in their discourses and the feedback they gave to one another regarding the common ground.

One caveat of my study is that it was a production study. Thus, I can only speculate that the frequency with which my responding variables were produced correlates with their importance in comprehension and establishing common ground. The dialogues collected in this study could be used as materials in future comprehension studies to confirm some of these assumptions.

Since my study used very generalized manipulations, the finer nuances of individual differences are not explored here. One such difference to explore in future studies would be to compare heritage Canadian speakers identity with monolingual Canadian speakers and non-native speakers to see if this modulates overlaps in cultural knowledge. Furthermore, individual factors such as personality may have also played a role in the effects found in this study. I initially attempted to measure these effects, but found my measures did not seem to be an effective measure for NS. If more data could be collected with just NS, I may be able to expand the scope of the present study.

However, the present study shows that participants bring baggage with them to a conversation that modulates their dialogues. My data suggests there are differences between NS and NN and suggests cultural overlap has an effect on establishing common ground. It also contributes new information on the role of EI factors in spoken dialogue, as opposed to the

common narrative comprehension studies. Finally, it provides rich data and recordings that may be useful for further investigations.

## Bibliography


Baayen, R.H. (2013). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". *R package* [Version 1.4.1]. Retrieved from https://CRAN.R-project.org/package=languageR

Baron-Cohen, S. & Wheelwright, S. (2004). The Empathy Quotient: An Investigation Of Adults With Asperger Syndrome Or High Functioning Autism, And Normal Sex Differences. *Journal of Autism and Development Disorders, 34*(2), 163-75.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger Syndrom or High-functioning Autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialogue: Coordinating speaking and understanding. *Psychology of Learning and Motivation, 53*, 301-344.

Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialogue. *Topics in Cognitive Science, 1*(2), 274-291.

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.

Clark, H. (1985). *Language use and language users.* In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3), 179-231. Harper and Row.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

van Dijk, T. A. (2014). *Discourse and knowledge: A sociocognitive approach.* Cambridge University Press.

Duran, N. D., Dale, R., & Kreuz, R. J. (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition, 121*(1), 22-40.

Fox, J. (2003). Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software, 8*(15), 1-27.

Fukumura, K. (2014). Interface of Linguistic and Visual Information During Audience Design. *Cognitive science.*

Gann, T. M., & Barr, D. J. (2014). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience, 29*(6), 744-760.

Gibbs, R.W. & Colston, H.L. (2017). The emergence of common ground. In R. Giora & M. Haugh (Eds.), *Applications of intercultural pragmatics: Cognitive and sociopragmatic perspectives on language use*. Walter de Gruyter GmbH & Co KG.

Grenier, A. (1998). Political Ideology Questionnaire in P. Popple & L. Leighninger (Eds.) *Social work, social welfare, and American society,* (Vol. 4). Ally and Bacon.

Grice, H. P. (1957) Meaning. *The Philosophical Review*, *66*(3), 377-388.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3). Academic Press.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research*. 114-158. New York, NY: Guilford Press.

Kecskés, I., & Mey, J. (Eds.). (2008). *Intention, common ground and the egocentric speaker-hearer* (Vol. 4). Walter de Gruyter.

Knutsen, D., & le Bigot, L. (2012). Managing dialogue: How information availability affects collaborative reference production. *Journal of Memory and Language, 67*, 326–341.

Kurby, C. A., & Zacks, J. M. (2015). Situation models in naturalistic comprehension. *Cognitive Neuroscience of Natural Language Use, 59*.

Lenth, R.V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software, 69*(1), 1-33.

Marian, V., Blumenfeld, H.K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research, 50* (4), 940-967.

Mills, G. J. (2014). Dialogue in joint activity: complementarity, convergence and conventionalization. *New Ideas in Psychology, 32*, 158-173.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 167–226.

R Core Team (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Retrieved from https://www.R-project.org/.

Thierrault, D. J., & Rinck, M. (2012). Multidimensional situation models. Higher level language processes in the brain: Inference and comprehension processes, 311.

Wickham, H & Francois, R (2016). dplyr: A Grammar of Data Manipulation. *R package* [Version 0.5.0]. Retrieved from https://CRAN.R-project.org/package=dplyr

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in cognitive sciences, 18*(5), 229-234.

Zwaan, R. A. (2015). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic bulletin & review*, 1-7.

## Appendix A: Additional Individual Measures

As mentioned in the Methods section of this paper, participants complete a political ideology questionnaire (adapted from Grenier, 1998) and four questionnaires measuring emotional intelligence and personality traits: Reading the Mind in the Eyes (Baron-Cohen, et al., 2001), Empathy Quotient Questionnaire (Baron-Cohen & Wheelwright, 2004), Big Five Index (John, 2008), and Interpersonal Reactivity Index (Davis, 1980). The reasons for initially including these measures and their eventual exclusion from the present analysis are briefly explained here.

The political ideology questionnaire was originally intended to be a control for political bias. Participants rate how much they do or don't support several political ideologies (e.g. "cuts to welfare programs") on a 7-point scale. These ratings place participants as being more right-leaning or left-leaning in their political beliefs. Since the prompts for discussion in this study—"Discuss Donald Trump" or summarizing articles about Donald Trump's two presidential bids—were politically charged, it was anticipated that overlap in political leaning could affect how participants interacted in the discussion portion of the study. For example, it is possible that differences in political views between two interlocutors may have negatively influenced participants' willingness to discuss the topic with their partner and decreased overall responses. However, the political questionnaire was introduced after data collection had begun and at the end of data collection there were not enough political questionnaires completed by native English speakers to warrant including it in the study. Furthermore, some of the collected questionnaires did not appear to be completed correctly (e.g. data missing or responding too all items with the same response), further decreasing the amount of usable data. Further data collection would benefit the investigation of political ideological effects in dialogues.

The four emotional intelligence and personality questionnaires were also intended to be controls. In the Reading the Mind in the Eyes task, participants choose one of four words that best describes what person is thinking or feeling when they are only shown a picture of that person's eyes. The results score indicates a participant's ability to detect another person's emotional state. The Empathy Quotient Questionnaire is a self-report questionnaire that measures the ability for a person to put themselves into someone else's shoes. The Big Five Index is a self-report questionnaire that scores participants on scales of extraversion, agreeableness, conscientiousness, openness, and neuroticism. The Interpersonal Reactivity Index is self-report

questionnaire that scores participants on scales of perspective-taking, fantasy, empathetic concern, and personal distress. It was anticipated that individual factors such as the ability to empathize with another person, or a high degree of openness or extraversion may affect participants' willingness to engage with their partner and establish common ground. However, the results from these tests were excluded from the final analysis as they appeared to have a strong cultural bias. For example, many non-native English speakers scored below average on the Empathy Quotient Questionnaire and Mind in the Eyes. Furthermore, for all four tests, many non-native speaker participants had to ask for clarification of terms and phrases more than native speakers. Therefore any differences in scores between native speakers and non-native speakers may actually represent unfamiliarity with vocabulary and certain figures of speech, not actual personality measures. Therefore this portion of the data collected was excluded from the analysis. To look at these individual factors and their effect on common ground in the future, more native English speaker data could be collected and the effects could be examined within a Canadian context. However, these test results do not appear to be accurate for non-native speakers.