# Temporal and spatial distance:

# An experimental study on pronoun resolution in Instant Messaging dialogue

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophischen Fakultät
der Universität zu Köln

im Fach Linguistik

vorgelegt von

Yuting Li

geb. am 01.07.1996

in Liaoning

Köln, den 10.09.2025

# Acknowledgements

I am deeply grateful to Katja Jasinskaja for her guidance, patience, and support throughout the years. Her advice has shaped this work in countless ways, and her encouragement carried me through the moments when I doubted that this dissertation would ever be completed. Looking back, I realize that I have known Katja for seven years, exactly a quarter of my life. I still vividly remember her semantics course, which I attended just a week after arriving in Germany for my master's studies. From that first encounter to the completion of this dissertation, her guidance has been truly formative, and I could not have wished for a better advisor.

I also owe special thanks to Anna Bonifazi, whom I first met on 10 October 2018 in her course *Multimodality in Language and Culture: Words, Music, and Tears*, where she used the song *While My Guitar Gently Weeps* for analysis. Her lecture sparked my initial interest in multimodality and inspired my later exploration of emoji throughout my MA and PhD research. Whenever I turned to her for help, she was always willing to lend a hand. Even in challenging times, she made herself available to discuss my work, for which I am deeply grateful.

I would further like to thank Petra Schumacher, Doris Mücke, and Marco García García. Their classes within the a.r.t.e.s. Graduate School broadened my understanding of prominence and offered valuable feedback on my experimental design. I am equally indebted to my fellow peers in the class who generously proofread parts of this work and provided constructive suggestions.

My sincere thanks go to Fahime Same, Job Schepens, and Maximilian Hörl for their help with data analysis, and to CRC 1252 for supporting my participation in the Statistics Summer School in Potsdam. I am also grateful to David Uerlings and Sabine Horstmann for their assistance in refining the experimental stimuli and for testing the experiments before they went online.

This work was made possible through the financial and infrastructural support of the Deutsche Forschungsgemeinschaft (Project-ID 281511265) and the Collaborative Research Centre 1252 "Prominence in Language." I am especially thankful to its leadership during the time this dissertation was developed: Klaus von Heusinger and Petra Schumacher, for building such a supportive and collaborative community.

Finally, I want to express my heartfelt thanks to my friends for their encouragement that carried me through the ups and downs of the past years. They are, in alphabetical order by last name: Qiaoxi Jia, Weijia Wang, Congshan Zhao, Yanshan Zhao, and Qilin Zhu . Above all, I owe my deepest gratitude to my grandparents Xiuyu Chu and Hongyun Li, my parents Wei Liu and Jun Li. Even though they may not fully understand the details of my research, they have always supported me unconditionally in whatever I chose to pursue. This dissertation is as much theirs as it is mine.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

**BC** Back Channel

**CA** Conversation Analysis

**CMC** Computer-mediated communication

**CR** clarification request

**ET** Editing Term

**H** Hypothesis

**I** Interruption

**IM** Instant Messaging

**MC** Main Channel

**NLP** Natural Language Processing

**NP** Noun Phrase

**OU** Original Utterance

**R** Repair

**RD** Referential Distance

**Ref** Referent

**RFC** Right Frontier Constraint

**RQ** Research Question

**RST** Rhetorical Structure Theory

**SD** Spatial Distance

**SDRT** Segmented Discourse Representation Theory

**TD** Temporal Distance

**TCU** Turn-Constructional Unit

**U** Utterance

**WPM** Words per Minute

# Chapter 1

# Introduction

Communication is at the heart of human interaction. Whether we are sharing information, building relationships, or solving problems together, language is the primary tool we use to connect with others. With the rise of digital technologies, the way we communicate has changed; some of our conversations no longer take place face-to-face or over the phone, but through written messages exchanged on screens. Among these, instant messaging (IM) has become one of the most widely used tools, shaping how we talk and what kinds of expressions we use.

IM is not just written conversation on a screen. It combines features of spoken and written language with unique capabilities: users can edit messages, react with emojis, recall messages, and engage in multiple chats at once. These application-based features influence not only the flow of dialogue but also how people interpret it. Messages do not always appear in strict chronological order, and feedback may be layered or delayed. This means that interlocutors must take into account both temporal sequence and visual layout when considering who or what is being referred to.

This dissertation investigates how people interpret references in IM dialogue. It focuses on how recency and different layers of IM communication affect which entities are seen as prominent. Through experimental studies and theoretical reflection, it seeks to bridge the gap between traditional discourse models and the realities of digital interaction.

## 1.1 Why Reference Matters in Instant Messaging Dialogue

In Frege's classic formulation, the reference (*Bedeutung*) of an expression is the actual entity it stands for in the world (Frege, 1892/1948). Reference is the mechanism by which linguistic expressions such as proper names, definite descriptions, or pronouns point to specific individuals or entities. This is something that happens all the time, often without noticing: for example, when someone says "she replied already" or "that part was unclear," the listener must figure out who "she" refers to or what "that part" means. In order to resolve such expressions, people rely on context, memory, and the semantics of the expression.

Reference resolution in IM differs in certain aspects from spoken communication, such as phone calls or face-to-face conversation, and from written communication, such as emails or letters, because it is not purely linear. Messages can be edited after they are sent, withdrawn after delivery, and some platforms even support disappearing messages that are automatically removed after being viewed. Comments can refer back to earlier parts of the conversation, and feedback can take different forms, like emoji reactions that appear visually attached to the corner of the message they respond to. In addition, messages are displayed only after they are processed by the server, meaning that when multiple chat participants are typing at the same time, their messages will be shown in a fixed order based on when they are sent and received by the server, even if only milliseconds apart. These features can lead to misalignment between the order in which messages are written, the order in which they are shown on screen, and the way participants perceive the flow of the conversation. This creates uncertainty about which parts of the dialogue are still relevant for interpretation and which entities are still prominent. In such cases, resolving references becomes interesting to study.

From a theoretical point of view, reference resolution is known to depend on factors such as grammatical role, discourse structure, and recency. For instance, the notion of referential distance (RD) introduced by (Givón, 1983) captures how recently a referent was mentioned in terms of intervening clauses; the Right Frontier Constraint (RFC) (Asher and Lascarides, 2003) restricts anaphoric links to discourse segments that are structurally available. Studies also show that readers and listeners are faster to interpret references to entities introduced in the most recent clause (e.g. Clark and Sengul, 1979). But in IM dialogue, recency is no longer a simple matter of time. A message that was sent earlier may become recent again if it is edited or reacted to. However, its position in the chat interface may remain unchanged. This creates a misalignment between the temporal order of message production and the spatial positioning of messages on screen. This misalignment raises questions about how recency should be defined in IM discourse. Is recency determined by the order in which messages are sent, or by their position on the screen at the time of interpretation? And how do these different dimensions of recency affect the prominence of discourse entities? These questions form one of the central points of this dissertation: to examine how recency is defined in the context of IM, and whether it influences referent prominence in the same way as in other modes of communication.

Another dimension to consider is the layered structure of the IM interface itself. In face-to-face conversation, listeners often provide short feedback such as “uh-huh,” “okay,” or “hm?,” which helps maintain the flow of interaction (Yngve, 1970; Schegloff, 1982). In IM, users also have other ways of giving feedback. They can type out brief responses similar to those in spoken conversation, or choose from a wide range of emojis to express reactions. Even when using emojis, users have options: they can send an emoji as a separate message, or attach it directly to a previous message as a reaction. This flexibility raises a further question: does the way an utterance is presented, whether integrated into the main message thread or displayed in a separate layer of the interface, affect how users interpret the discourse, particularly when it comes to resolving reference?

These practical and conceptual issues provide the starting point for this dissertation. The goal of this dissertation is to investigate how IM specific features shape the way readers track reference and determine which entities remain accessible or become prominent in the ongoing dialogue.

Understanding how reference works in IM also has implications for platform design, human-computer interaction, and natural language processing (NLP). As more and more conversations move to digital platforms, it becomes essential to examine not just what people say, but how the platform influences how people say it and how their messages are interpreted.

## 1.2 Tracking Prominence in Instant Messaging Discourse

Prominence refers to the relative accessibility of elements in a discourse, the idea that among a set of potential referents, some are more accessible than others. These prominent entities serve as anchors for interpreting new information and maintaining coherence (Von Heusinger and Schumacher, 2019). When readers or listeners encounter a referring expression such as a pronoun, they rely on multiple cues to determine which entity is being referred to. Not all previously mentioned entities are equally likely candidates for reference; some are more "in focus" than others. The study of prominence, therefore, asks what makes one entity more accessible than another at a given point in discourse.

A number of theoretical frameworks have been developed to account for how prominence is established and maintained. The Givenness Hierarchy (Gundel et al., 1993) proposes that the form of a referring expression reflects the assumed cognitive status of its referent in the listener's mind. For example, using a pronoun implies that the referent is already highly accessible. Centering Theory (Grosz et al., 1995) further models the continuity of reference across utterances by predicting which discourse entities are carried forward and remain in focus. Both frameworks assume that prominence depends on the grammatical role, syntactic position, and

prior mention. Subjects are generally more prominent than objects, and recently mentioned or repeatedly referenced entities tend to stay more accessible than those mentioned only once. Prominence is also dynamic, it shifts with topic changes, speaker turns, and the unfolding of discourse.

Prominence is not determined by linguistic structure alone. It is also shaped by the modality of communication. In face-to-face interaction, speakers and listeners draw on a range of non-verbal cues, such as intonation, gaze, gesture, and body orientation, to signal focus and manage common ground (Clark 1996; Goodwin 1981; Kendon 2004, ch. 9; McNeill 1992). In written texts, features such as typographic emphasis (e.g., italics or bold) can highlight certain elements and guide the reader's attention (Maia, 2018). These modality-specific cues help establish coherence and enable readers or listeners to track which entities remain available for reference as the discourse unfolds.

IM lacks the prosody and physical co-presence of face-to-face interaction, replacing them with an interface shaped by platform. Here, cues such as spatial arrangement, edit history, message quoting, and emoji reactions play an additional role in shaping prominence. IM communication presents several new challenges for modeling prominence. Participants may type simultaneously or respond after extended delays, and this variability in response time may lead to fragmented discourse, where the connections between utterances must be inferred rather than directly perceived.

A second challenge arises from IM-specific platform features such as message editing, withdraw, and emoji reactions. These features introduce a level of flexibility not typically found in spoken discourse, like face-to-face conversation, or written discourse, like email or a letter. A message may be sent and then revised afterward, perhaps without the other participant noticing. A visible message may be deleted, leaving gaps in the conversation history. Such features complicate the linear flow of interaction and create ambiguities in tracking reference. In particular, they pose a challenge for recency-based models of prominence. For instance, a message that was originally sent long ago but recently edited may display updated content in place. Although the edit makes it temporarily the most recent contribution,

the message still remains higher up in the thread rather than at the bottom near the input box. In this case, the referent it introduces is temporally updated but spatially distant, raising the question of whether users treat it as currently prominent.

These spatial-temporal misalignments are linked to a third feature of IM: its layered interface structure. Most platforms distinguish between the main message thread and an additional interactive layer where reactions or edits can be made. Elements of the interactive layer often appear alongside an earlier message rather than forming new messages in the main flow of conversation. As a result, users can engage with prior content without explicitly reintroducing it as a stand-alone message. While these mechanisms provide new ways to give feedback or clarify meaning, they also raise important questions. Do edits or emoji reactions draw attention back to earlier referents? Do they affect prominence in the same way a new message would? And if they occur not in the main thread but in the interactive layer, do they shape pronoun interpretation at all?

## 1.3 Research Questions and Aims

The structural complexity of IM dialogue raises questions for discourse coherence and reference. IM communication can involve a misalignment between the order in which messages are written, edited, or responded to, and the order in which they appear visually on screen. It also introduces a distinction between utterances introduced in the main message thread and the interactive layer. These properties suggest that classic models of prominence may not fully capture the dynamics of digital interaction.

The central aim of this dissertation is to investigate how such features of IM environment influence the prominence of referents and their subsequent reference through expressions such as pronouns. More specifically, it tries to answer the overarching question: How is prominence established and maintained in IM dialogue, and what features make a referent more likely to be interpreted as prominent?

## Research Questions

This dissertation is structured around two research questions:

1. **Which distance has a stronger influence on pronoun resolution in IM dialogue: spatial distance or temporal distance between a referent's introduction and its subsequent mention?**

   Recency has been shown to play a role in reference resolution: all other factors being equal, entities mentioned more recently in time are more accessible in memory and therefore more likely to be interpreted as the antecedent of a pronoun. However, in IM, temporal order may diverge from visual layout. A message that was edited recently may still appear further up the screen, while the temporally distant utterance occupies a more recent spatial position. In this sense, recency in IM needs to be understood across two dimensions: Temporal Distance (TD) and Spatial Distance (SD). This raises the question of which cue has a stronger influence on reference resolution when the two diverge.

2. **Does the way an utterance is displayed, either in the main message thread or in an interactive layer, affect the prominence of its referents in IM dialogue?**

   Many IM platforms allow users to revisit previous utterances through message reactions or edits. These are often visually separated from the main thread of conversation. Do utterances in the interactive layer affect referent prominence in the same way as those in the main thread? Does the presentation of feedback as an interactive reaction versus as a stand-alone message influence whether the associated referent remains accessible in the ongoing discourse?

Addressing these questions requires both conceptual clarification and empirical testing, which this dissertation provides through a series of controlled experiments.

## Empirical Strategy

To address these questions, this dissertation reports four experiments, each targeting a different aspect of reference resolution in IM. All experiments were conducted in a chat-style interface that replicates the look of real-world IM while allowing precise control over message order and presentation format. In each case, participants interpreted ambiguous pronouns in carefully constructed dialogues, providing evidence of which referent was treated as most prominent under the given conditions.

- **Experiments I and II** focus on the effects of spatial and temporal distance. By deliberately misaligning the order of utterance production and visual display, these studies test whether participants resolve reference based on the most recently received message or the spatially closest one.

The results of Experiments I and II point to the importance of layering in IM, where message editing introduces a secondary layer, creating misalignments between spatial and temporal distance. IM platforms organize interaction across different interface layers. On the one hand, there is the layer where utterances appear in chat bubbles as stand-alone messages in the chat window. On the other hand, there is another interactive layer that allows users to edit or react to existing messages without producing a chat bubble. This layered organisation of dialogue motivated me to draw on the distinction between the main channel and back channel in IM. The main channel corresponds to the layer where stand-alone messages are displayed. The back channel contains actions that do not create new messages but are visually or structurally attached to earlier messages, such as edits or emoji reactions. Introducing this distinction makes it possible to investigate whether utterances presented in the main channel influence the prominence of the referent they introduced differently from those that appear in the back channel. This forms the guiding question for Experiments III and IV.

- **Experiments III and IV** turn to layering. They compare the effects of repair and feedback when presented either in the main channel or in the back channel. Experiment III uses self-repair constructions; Experiment IV tests feedback via emoji reactions.

Each experiment is designed to isolate the influence of a single factor (recency or channel) while keeping other variables constant, such as grammatical form and discourse relation. This control ensures that any observed effects can be linked to the experimental manipulation.

At the theoretical level, this dissertation contributes to the research that aims to adapt models of discourse processing to digital environments. It offers empirical evidence on how recency and channel function in contexts where messages are editable and presented in different layers. The results provide insights into how prominence is constructed when cues like prosody or gesture are absent.

At the applied level, the findings have implications for the design of IM interfaces and for computational systems that aim to model dialogue. Understanding how users track referents in IM can inform the design of more intuitive interaction features, such as better highlighting of edited content or more integrated feedback mechanisms. It can also improve NLP systems that perform tasks such as pronoun resolution, message summarization, or chatbot response generation in digital conversations.

## 1.4 Overview of the Dissertation

The chapters are organized to move step by step from background and concepts (Chapters 2–4), through the development of research questions and experimental studies (Chapters 5–8), and finally to a broader discussion and conclusion (Chapter 9).

Chapter 2 lays the theoretical foundation by introducing the concept of prominence as it has been understood in discourse processing. It adopts the definition of prominence from Von Heusinger and Schumacher (2019)

and outlines how prominence is shaped by grammatical role, discourse structure and recency.

Building on this foundation, Chapter 3 turns to the context of IM and explores how its unique technological and interactional affordances shape discourse. The chapter begins by situating IM within broader typologies of communication modes, focusing in particular on the dimension of synchronicity. As a hybrid form of communication, IM contains features of both synchronous (e.g., face-to-face conversation) and asynchronous (e.g., email) interaction. This gives IM discourse its own characteristics: temporally delayed responses, message editing after sending, and emoji reactions. These features raise new challenges for theories of discourse coherence and reference resolution.

Chapter 4 motivates the need to refine the concept of recency for IM discourse, where messages are not always processed in strict temporal order. The chapter develops the notion of recency as a form of distance, and proposes a distinction between two types of distance in IM: temporal distance, which concerns when a message is produced or received, and spatial distance, which concerns where a message appears within the visual layout of the chat interface. Because IM platforms allow for message editing, reactions, and layered feedback, often without altering a message's visual position in the thread, these two dimensions of distance can become misaligned. This leads to a central research question for the first set of experiments: when temporal and spatial distance point to different referents, which distance has a stronger effect on reference resolution?

Chapter 5 explores the role of repair in IM and how it may contribute to misalignments between spatial and temporal recency. While repairs in spoken dialogue often occur immediately, responding to hesitations, errors, or signs of misunderstanding, repairs in IM may take the form of message edits, which can occur with a delay and without visibly introducing a new turn, raising questions about how such edits affect the prominence of discourse referents. The chapter argues that repair in IM can contribute to misalignments in both spatial and temporal dimensions and sets the stage for empirically testing the consequences of these misalignments.

Chapter 6 presents two experiments designed to test how recency-based prominence is affected by the misalignment of spatial and temporal order. Results from both experiments demonstrate a consistent preference for spatially close referents, suggesting that visual layout exerts a stronger influence on pronoun interpretation than temporal order. These findings raise further questions about how IM users process repair in IM communication.

Before turning to the final experiments, Chapter 7 develops a theoretical account of the back channel in IM. It draws on foundational work by Yngve (1970), who introduced the concept of the back channel as a layer of listener feedback in face-to-face conversation. Building on this idea, the chapter distinguishes between two layers in IM: the layer where utterances are presented in stand-alone chat bubbles, which is conceptualized as the main channel, and the back channel, where participants provide emoji reactions or modify prior content without initiating a new message. This perspective provides the foundation for the design and interpretation of Experiments III and IV.

Experiments III and IV, presented in Chapter 8, examine how the main and back channels influence the prominence of referents in IM dialogue. Experiment III investigates message repairs, while Experiment IV focuses on evaluative emoji feedback. Across both studies, the results show a consistent pattern: when utterances (repairs or emoji comments) are presented in the main channel, they significantly increase the prominence of the related referent, whereas the same utterances in the back channel have a much weaker influence. These findings show that the channel in which an utterance appears plays a role in shaping how prominently it is perceived and how likely its referents are picked up in subsequent discourse.

# Chapter 2

# Prominence and Pronoun Resolution

The interpretation of pronouns is a fundamental process in discourse comprehension, allowing speakers and listeners to track referents across utterances and maintain coherence. Prominence plays a role in guiding this process, since it helps maintain accessibility over time. Its role becomes especially important, however, in contexts where multiple potential antecedents are available. In such cases, entities that are more prominent are more likely to be selected as antecedents for pronouns (Arnold, 2010; Von Heusinger and Schumacher, 2019).

The relationship between prominence and pronoun resolution has attracted considerable theoretical and empirical attention. Prominence not only informs our understanding of how reference is maintained in discourse, but also reveals much about the organization of information and the cues listeners and readers use to resolve ambiguity. Determining which factors make an entity prominent is a complex task, as a combination of syntactic, semantic, and discourse-level features shapes prominence. As a result, resolving a pronoun usually involves integrating multiple cues to determine the most plausible antecedent.

This chapter explores the relationship between prominence and pronoun resolution in detail. Section 2.1 provides definitions and theoretical perspectives on prominence. Section 2.2 discusses why prominence is central to pronoun interpretation. Section 2.3 reviews the features that contribute to the prominence of discourse entities, with particular attention to those that are most relevant for pronoun resolution, including grammatical roles, discourse structure, recency, and modality-specific considerations. Finally, Section 2.4 addresses open questions on prominence in IM dialogue.

## 2.1 Defining Prominence

Prominence in discourse refers to the property that makes certain entities stand out and become more likely targets for subsequent reference. It is not a fixed or inherent property of a discourse entity, instead, it is shaped dynamically through the interaction of multiple linguistic cues as discourse unfolds. Primus (2013) and Von Heusinger and Schumacher (2019) offer a widely used framework for understanding prominence, which can be summarized in three main points:

> **Def.1:** Prominence is a relational property that singles out one element from a set of elements of equal type and structure.
> **Def.2:** Prominence status shifts in time (as discourse unfolds).
> **Def.3:** Prominent elements are structural attractors, i.e. they serve as anchors for the larger structures they are constituents of, and they may license more operations than their competitors.

These criteria highlight three aspects:

**Singling out:** Prominence is understood as a relational phenomenon, typically established through contrasts among potential referents. For example, in the sentence below from Jasinskaja et al. (2015), there are two candidates for the pronoun “she,” but the context and the verb’s semantics make “Mary” more prominent, and therefore a more likely choice:

(1) Jane hit Mary because she$_{Mary}$ had stolen a tennis racket.

**Dynamicity:** Prominence can shift as the discourse context changes or as new information is introduced. For instance:

(2) a. Sarah was talking to Lisa at the party.
b. Then, Lisa left to get some drinks.
c. She returned a few minutes later with a tray.

Here, "Sarah" may initially be more prominent, but after the sentence (2-b), "Lisa" becomes the most accessible referent, so the pronoun "she" is most naturally interpreted as referring to "Lisa."

**Structural attraction:** Prominent entities function as structural anchors in discourse, supporting a wide range of referring expressions and contributing to coherence. For example, Von Heusinger and Schumacher (2019) demonstrate that a central referent such as "Ron" in the sequence below remains accessible across multiple sentences and can be referenced flexibly:

(3) a. Many athletes met at the annual award ceremony in Baden-Baden.
b. It was the first time for Ron, a member of the swimming team, to attend this get-together.
c. At the buffet, the swimmer talked for some time to a cyclist.
d. He/This swimmer/The athlete reminisced about the Olympics in Rio de Janeiro.
e. Later in the evening, he was introduced to a rower/*the rower/*him.

Here, "Ron" continues to serve as an anchor for reference throughout the text, while less prominent entities, such as "a rower," do not permit the same referential flexibility.

## 2.2 Why is Prominence Central to Pronoun Resolution?

Pronouns are inherently ambiguous: expressions such as “she,” “he,” or “it” do not themselves specify which entity they refer to in a given context. For communication to succeed, speakers and listeners must reliably identify the intended referent of a pronoun. The process of pronoun resolution, therefore, relies on the concept of prominence.

As Von Heusinger and Schumacher (2019) and others have shown, when a pronoun appears, comprehenders do not consider all potential referents equally. Instead, they are naturally biased toward those entities that are most prominent in the current discourse context. Understanding how prominence is established and updated is thus essential for explaining how pronoun resolution takes place. Several theoretical frameworks have emphasized the role of prominence in pronoun resolution:

Accessibility Theory (Ariel, 1990) proposes that pronouns and other referring expressions are chosen according to the cognitive accessibility of their antecedents. Pronouns are typically used for highly accessible referents, while less accessible entities require more explicit forms, such as definite descriptions or full noun phrases (NP). Factors such as recency, grammatical role, and discourse structure all contribute to accessibility, allowing speakers to use pronouns with minimal risk of misunderstanding.

The Givenness Hierarchy (Gundel et al., 1993) similarly argues that the choice of referring expression reflects the cognitive status of the referent in the listener’s mind. Pronouns are reserved for entities that are “in focus” or “activated,” while less accessible referents require more explicit expressions.

Centering Theory (Grosz et al., 1983, 1995) formalizes how local discourse coherence depends on the prominence of discourse entities and makes predictions about the acceptability and interpretation of pronouns in context. According to Centering Theory, each utterance contains a set of “centers” (potential discourse entities), and the most prominent one, the “backward-looking center” is typically the default referent for subsequent

pronouns. Transitions between centers are closely linked to how easily a pronoun is interpreted: the smoother the transition, the easier and more natural pronoun resolution becomes.

These frameworks support the idea that prominence acts as an organizing principle for reference in discourse. By establishing which entities are most accessible at each point, prominence enables listeners and readers to resolve pronouns efficiently and accurately, supporting coherent and successful communication.

## 2.3 Prominence-Lending Features Shaping Pronoun Resolution

Prominence arises from a complex interplay of features. This section reviews features that shape prominence, with a focus on those most directly relevant to pronominal reference and the research question of this thesis.

### 2.3.1 Grammatical Features

The grammatical role of a NP is an important predictor of prominence in discourse. Subjects have been shown to enjoy privileged status in terms of accessibility and pronominal reference (Keenan and Comrie, 1977; Crawley et al., 1990). When a pronoun could refer to multiple antecedents, listeners and readers tend to resolve it to the most recent subject, a phenomenon known as the subject bias (Gernsbacher and Hargreaves, 1988).

(4) John hit Bill and Mary kicked him.

In this example (from Crawley et al. 1990), the pronoun *him* is ambiguous: it could either refer to *John* (the grammatical subject) or to *Bill*, which hold the same grammatical role (object) as the pronoun. In such an ambiguous sentence, experiments have proven that the subject *John* holds a greater preference and thus is more likely to be referred to.

This preference is formalized in the Noun Phrase Accessibility Hierarchy proposed by Keenan and Comrie (1977), which ranks grammatical roles according to their accessibility for relativization and pronominalization: *subject* > *direct object* > *indirect object* > *oblique* > *genitive* > *object of comparison*. Entities higher in this hierarchy are more easily retrieved from memory and are more likely to serve as referents for subsequent discourse.

Beyond grammatical role, several additional features contribute to prominence in discourse. *Word order* can affect the accessibility of referents, with sentence-initial positions often being more prominent (Gundel et al., 1993). *Voice* alternations, such as the contrast between active and passive constructions, allow speakers to shift patient to subject position, thereby influencing prominence and guiding subsequent reference (Keenan and Dryer, 2007). Finally, *cleft constructions* highlight specific elements within a sentence, marking them as focal and thus prominent (Prince, 1978; Lambrecht, 1994, 2001).

### 2.3.2 Recency

The intuition that recency affects prominence and reference resolution is straightforward: entities mentioned more recently remain more accessible in memory, and are therefore more available as antecedents for subsequent pronouns or anaphoric expressions (Grosz et al., 1995; Givón, 1983). The importance of recency was demonstrated early on by Clark and Sengul (1979), who showed that readers comprehend and process referents more quickly when the antecedents appear in the immediately preceding sentence compared to those mentioned further back. This finding established recency as a factor in reference resolution.

Linear recency refers to the sequential order in which entities are introduced in discourse, typically measured by the number of intervening clauses or sentences since the last mention. This corresponds to what Givón (1983) terms referential distance (RD), defined as the number of clauses (or elapsed time) from the last occurrence in the preceding discourse.

Consider the following sequence:

(5) a. Anna cleaned the kitchen.
b. Lisa washed the dishes.
c. She made coffee.

Here, the pronoun *she* in (5-c) is most naturally interpreted as referring to "Lisa," the most recently mentioned potential antecedent, and the RD between the pronoun and its antecedent is one clause.

It is worth mentioning that the effect of recency matters at the clause-by-clause level rather than within a single clause (Arnold, 2025), as in the following example:

(6) a. Anna cleaned with Liz.
b. She used the broom.

Here, although "Liz" is linearly more recent, "Anna", as the subject, is preferred as the antecedent of "she". This illustrates that within-clause syntactic prominence can override recency.

More recent work has refined our understanding of how recency interacts with other prominence-lending cues. For example, Arnold (2025) further explored how recency affects pronoun interpretation, specifically examining the interaction of recency with topicality and attentional cues such as pointing and gaze.

In Arnold's study, participants were presented with short audio stories where one character *Matt* was established as the subject and the discourse topic in the first two sentences, and a second character *Will* was then introduced in a later sentence as the new subject, a setup named "topic switch." The manipulation was whether the pronoun *he* appeared immediately after the discourse topic (as in 7) or after the switch topic 8, and the critical question was to resolve the pronoun *he*:

(7) a. Matt was making a cake with Will.
b. Matt read the recipe.
c. He measured out two cups of flour.
d. Will got out the hand mixer.
e. They mixed together the rest of the ingredients and put the batter in the oven.

(8) a. Matt was making a cake with Will.
b. Matt read the recipe.
c. Will got out the hand mixer.
d. He measured out two cups of flour.
e. They mixed together the rest of the ingredients and put the batter in the oven.

Arnold also tested the effect of attentional cues, such as speaker gaze and pointing, using video stimuli. In these cases, the speaker's gestures could direct attention to one of the two characters. The aim was to see whether these non-linguistic cues could override recency or topic status.

Arnold's findings from the audio experiment replicated earlier evidence that pronouns consistently tend to resolve toward the most recently mentioned entity, regardless of topic status. While topicality has often been considered another important factor in pronoun resolution, Arnold demonstrated that recency bias is not reducible to topicality. Moreover, her further experiments found that recency effects persist even when competing cues such as visual or gestural indicators are present.

Although recency is an important cue for prominence, discourse processing is not purely linear. Structural organization also plays a role, as some referents remain accessible for anaphora even when not recently mentioned. This indicates that prominence is shaped not only by temporal order but also by the rhetorical relations linking utterances.

### 2.3.3 Discourse Structure

In Segmented Discourse Representation Theory (SDRT) (Asher and Lascarides, 2003), rhetorical relations connect segments in a discourse. And overall coherence is achieved through the establishment of these relations. As in the following example from Jasinskaja and Karagjosova (2020):

(9) a. I did two things on my seventy-fifth birthday.
b. I visited my wife's grave. Then I joined the army.

Utterance (9-b) is an elaboration of (9-a), as it specifies the two activities mentioned in the first utterance.

Rhetorical relations can broadly be divided into two types. Coordinating relations such as narration, result, parallel, contrast, or correction move the discourse forward. Subordinating relations such as elaboration, explanation, or consequence create hierarchical structures by providing supporting information (Asher, 1993; Asher and Vieu, 2005). [1]

A further principle central to SDRT is the Right Frontier Constraint (RFC), first introduced by Polanyi (1988) and later formalized by Asher and Lascarides (2003). The RFC states that, as a discourse unfolds, only certain "open" segments, i.e., those on the right edge, or "frontier" of the developing discourse structure are available for new attachments and for referential processes like anaphora. This means that referents introduced in these open segments remain accessible, even if they are not the most recent in linear order.

The interaction between rhetorical relation and the RFC influences prominence. In coordinating relations such as narration or parallel, only the rightmost discourse segment remains accessible on the frontier, so referents introduced in earlier segments lose their availability for

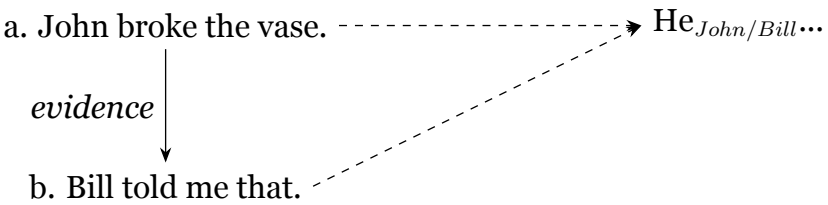
[1] It is important to distinguish rhetorical coordination and subordination from their syntactic counterparts. Syntactic coordination and subordination concern the grammatical relations between clauses within a sentence, whereas rhetorical coordination and subordination operate at the discourse level, connecting larger discourse segments. The two levels do not map one-to-one: a syntactically subordinate clause may still be in a coordinating discourse relation.

subsequent pronouns. Consider the following example from Jasinskaja and Karagjosova (2020):

(10) a. John broke the vase. $\xrightarrow{parallel}$ b. Bill broke the mirror. $He_{Bill}$...

Here, utterances (a) and (b) are connected by a coordinating parallel relation. Thus, only Bill remains available as an antecedent for the pronoun *he* in the continuation, since John is no longer accessible on the right frontier.

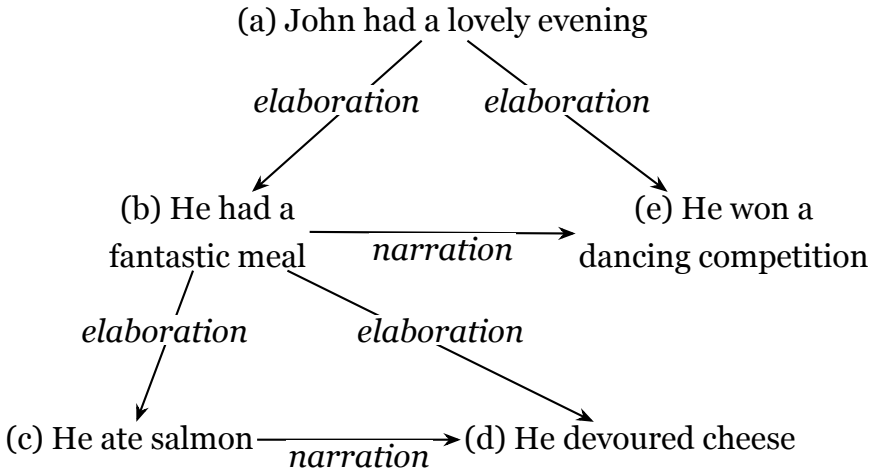
(11)




By contrast, in subordinating relations, multiple nodes may remain accessible. In example (11) from the same authors, utterance (b) provides evidence for why the speaker knows that *John broke the vase*. Since (a) and (b) are in a subordinate relation, both discourse segments remain available, and the subsequent pronoun *he* can refer either to *John* or to *Bill*.

This rule also applies in longer discourses, as illustrated by the following sequence adapted from Asher and Vieu (2005):

(12)



In example (12), utterances (b)–(d) elaborate on the main event introduced in (a). As the discourse develops, only the segments on the current right frontier remain available for new attachments. When the utterance continues with (e), reference is constrained by the structure of the discourse tree rather than by linear recency alone. Thus, when (e) is added, it cannot attach back to the earlier food descriptions because they are no longer on the frontier; instead, it links to (a). As a result, continuation such as *Then he had a great dessert* is awkward, because it attempts to attach back to the meal segment, which is no longer on the right frontier.

The RFC therefore helps explain why certain referents remain accessible even after new utterances are introduced, and why others drop out of prominence as the discourse structure becomes more complex. It shows that discourse accessibility, and by extension, prominence is achieved not just by linear order, but also by the structure of the discourse.

A further factor that shapes prominence is information structure, which refers to the cognitive and communicative organization of discourse, particularly how speakers and writers structure utterances in terms of given vs. new information (Lambrecht, 1994; Prince, 1981; Halliday, 1967), topic vs. comment and focus vs. background (Krifka, 2008). According to theoretical models, *given information* includes elements already active in the listener's consciousness or discourse context, thus readily accessible. Conversely, *new information* represents previously unmentioned or unknown elements, typically requiring additional cognitive resources for processing and integration (Prince, 1981). Similarly, the *topic* identifies the entity about which information is conveyed. And *comment* provides the propositional content about the topic (Lambrecht, 1994; Krifka, 2008). These structures facilitate efficient information processing, as given information and topics require less cognitive effort to access and integrate, promoting their prominence within ongoing discourse.

### 2.3.4 Modality-Specific Considerations

Prominence is also influenced by modality-specific considerations, which vary between spoken and written forms. Each modality has unique resources for signaling prominence, such as prosody and visual cues in face-to-face communication, and typographical features in written text.

In **spoken discourse**, *prosodic features* such as pitch, stress, intonation, and pausing play an important role in signaling prominence. These features help listeners identify which discourse entities are new, contrastive, or topical at any given moment. Experimental research supports the role of prosody in discourse comprehension. For instance, Cohen et al. (2001) investigate the effects of prosody by having participants listen to recorded passages with varying intonation patterns (normal, monotonous, or altered). Participants' performance on subsequent comprehension and word recognition tasks show that prosody significantly enhances oral discourse processing by improving comprehension and word recognition. Their results also suggest that prosody facilitates meaning extraction, supports the structuring of short-term memory, and activates associations in long-term memory.

Beyond prosodic cues, *visual cues* such as gaze direction also affect the prominence of referents in spoken discourse. Burmester et al. (2018), for example, explored how linguistic and visual prominence interact during sentence comprehension in German, considering both canonical (subject–object) and non-canonical (object–subject) word orders. Their experiments demonstrated that linguistic prominence (e.g., grammatical role, word order) generally cause a stronger influence on referent accessibility at the discourse level compared to visual salience. Nevertheless, visual cues, especially gaze direction towards specific referents, can facilitate sentence processing, particularly when the visually cued referent appears in sentence-initial position.

In **written discourse**, prominence can also be indicated through *typographical features*, such as capitalization, boldface, italics, and underlining. These typographical cues can be used to highlight specific words or phrases, guiding readers' attention and signaling important information. For example, Fraundorf et al. (2013) conducted recognition memory experiments in which participants read short texts containing typographically emphasized words. Consider their example:

(13) a. Both the British and the French biologists had been searching Malaysia and Indonesia for the endangered monkeys.
b. Finally, the BRITISH spotted one of the monkeys in Malaysia and planted a radio tag on it.

Their results demonstrated that emphasized elements, such as "BRITISH", are better remembered, and participants are subsequently better able to reject false claims involving related contrastive alternatives (e.g., *The French scientists found the endangered monkey*). This finding indicates that typographical emphasis not only strengthens memory encoding of the emphasized content but also leads readers to actively construct mental representations of alternative entities implied by the discourse.

## 2.4 Open Questions on Referential Choices in Instant Messaging Dialogue

Computer-mediated communication (CMC), particularly IM dialogue, blends features of textual communication with visual presentation and time-sensitive turn-taking, creating a complex environment for interpreting discourse prominence. Before turning to the open questions, consider first a dialogue exchange:

(14) a. *Lea: Hey, how is everything going?*
b. *Freya: Selina just washed the dog.*
c. *Freya: She also fed the cat.*
d. *Freya: Now she is petting it.*

In this exchange, Freya produces a sequence of utterances. The subject Selina remains constant across sentences, while two potential referents (*dog* and *cat*) are introduced. Because (14-b) and (14-c) stand in a coordinating relation, prominence, and therefore the interpretation of the pronoun *it* in (14-d), is resolved by recency: the most natural reading is that *it* refers to *cat*, the most recently mentioned non-human entity.

If the same exchange occurred in an IM setting, the dialogue could be presented differently. IM platforms add an interactional layer that allows for emoji comments or message edits (see Chapter 3). Figures 2.1 and 2.2 [2] illustrate such a case:


Figure 2.1: An example of IM dialogue


Figure 2.2: An example of IM dialogue with emoji comment

[2] All chat interface screenshots presented in this thesis are constructed examples created by the author for illustrative purposes, and do not represent real user conversations, unless otherwise stated.

Lea (messages in green chat bubble, right-hand side) initiates the exchange, and Freya replies by describing Selina's actions. Lea then attaches a confused-face emoji to Freya's first utterance (*Selina just washed the dog*). This emoji may signal confusion about why Selina washed the dog or express disapproval. Before the emoji is added, the pronoun *it* in Freya's third utterance would most likely be interpreted as referring to *cat*, the immediately preceding entity, due to its recent mention. However, the emoji-commenting feature could shift the discourse dynamics by bringing the first utterance about *washing the dog* back into conversational focus, making it temporally close again. This could be a potential shift in prominence and could increase the likelihood that the pronoun *it* would now be interpreted as referring to *dog* rather than *cat*.

Currently, we do not know whether such shifts in prominence actually occur in natural IM discourse, nor to what extent these features affect reference resolution. One of the aims of this dissertation is to investigate how technical features of IM, such as emoji comments and message edits used by speakers,[3] influence reference resolution in digital communication.

[3]For clarity, I use the term speaker to refer to the sender of a message, in order to avoid variations such as message sender, writer, or author.

# Chapter 3

# Instant Messaging Dialogue

At the end of Chapter 2, I showed how commenting on an utterance with an emoji might be able to shift the dynamics of interaction by bringing an earlier contribution back into conversational focus. Such affordances challenge traditional notions of recency, as will be discussed in Chapter 4, and they make IM discourse a particularly interesting material for investigating how pronoun resolution is shaped.

This chapter provides a broader overview of IM as a mode of communication. The aim is to highlight features that distinguish IM from both face-to-face and written communication. By doing so, the chapter establishes the foundation for understanding how the technological capabilities and linguistic practices of IM interact with discourse processes.

This chapter is organized as follows. Section 3.1 introduces different modes of communication with respect to synchrony and situates IM within this continuum. Section 3.2 outlines the characteristics of IM dialogue, focusing on three dimensions: its technical affordances, such as typing indicators, read receipts, and message-editing functions; its linguistic informality, including the frequent use of abbreviations, acronyms, and creative spellings; and its multimodal resources, such as images, GIFs, and emojis.

## 3.1 Modes of Communication and Synchrony

With the rapid advancement of communication technologies, mobile devices have become central to everyday interaction. People can now communicate through phone calls, video conferencing, or IM, making mobile-centric communication a global phenomenon. In Germany, for example, the JIM study reported that by 2015 every second teenager owned a smartphone (Dürscheid and Frick, 2014). This pattern reflects a worldwide trend: according to DataReportal (2025), by January 2025, 70.5% of the world's population was using a mobile phone. These figures illustrate not only the rapid adoption of smartphones but also their central role in facilitating communication.

As a digital medium, IM offers a dynamic environment characterized by informality, spontaneity, and multimodality, which influence conversational practices. Popular platforms such as WhatsApp, Facebook Messenger, WeChat, Telegram, and Signal further expand the range of interaction by integrating text, visual content, audio messages, and interactive features.

The development of digital communication platforms has also influenced everyday language use. Since the first text message was sent in 1992 (BBC News, 2022), texting quickly gained global acceptance. By 2005, questionnaire studies by Faulkner and Culwin (2005) already showed that text messaging had become a preferred mode of communication over email, fax, and even telephone calls. While Schlobinski et al. (2001) described SMS in 2001 as a "new" form of communication, it is now regarded as outdated in many contexts. Building on these earlier forms, IM allows for the real-time exchange of not only text but also multimedia content such as emojis, images, hyperlinks, and videos, making it a widely used mode of communication.

Before turning specifically to IM dialogue, it is important to clarify several key terms associated with CMC. CMC refers to communication between humans that is mediated by computers (Herring, 1996). It contains

a broad range of communicative practices facilitated by digital technologies, which can be classified according to their timing, mode of interaction, or communicative purpose. These dimensions often overlap; however, for present purposes, the classification based on timing is especially relevant. On this basis, CMC can be divided into three categories: synchronous, asynchronous, and hybrid forms.

While definitions of **synchronous communication** are more common in computer science, they provide a useful starting point for understanding the concept in interactional terms. For example, according to Cacciagrano and Corradini (2001), a communication is synchronous when sending and receiving information between a sender and a receiver are simultaneous events.

Strictly speaking, transmission and perception are never perfectly simultaneous, because sound waves or digital signals always take time to travel. However, if this delay is perceptually negligible, the interaction can still be considered synchronous. A feature of synchronous communication is that speech is produced and perceived incrementally. Listeners begin to hear and interpret an utterance while it is still being spoken, rather than only after it is complete. For example, in the sentence "Celina washed the dog," the word "Celina" is already perceived before the speaker produces "washed." This overlap between production and perception creates the real-time effect that characterizes synchronous communication. Applying this to the study of discourse, synchronous communication can be understood as interaction in which participants exchange utterances in real time, enabling immediate responsiveness and turn coordination.

Face-to-face conversation is the prototypical example of synchronous communication. In this mode, participants share the same physical space, perceive one another's verbal and non-verbal cues, and align their actions moment by moment. Phone calls are also synchronous, as they allow for the instant exchange of speech and joint participation despite physical distance.

CMC can likewise take synchronous forms when technology supports real-time interaction. Synchronous CMC requires the simultaneous presence of participants and allows for immediate exchange (Romiszowski and

Mason, 2013). Examples include *video conferencing* (e.g., Zoom, Microsoft Teams, or Google Meet), defined as a live visual connection between remote participants that simulates face-to-face interaction (TechTarget, 2021), and *voice calls* via platforms such as Skype or WhatsApp.

Various works have shown that synchronous CMC shares important similarities with oral communication, both in terms of participant roles (Warschauer, 1995; Darhower, 2002; Abrams, 2001; Böhlke, 2003) and discourse functions (Sotillo, 2000).

A communication is **asynchronous** when sending and receiving information between a sender and a receiver do not necessarily happen at the same time instant (Cacciagrano and Corradini, 2001). Applying this for discourse analysis, asynchronous communication can be understood as interaction occurs with a time delay and does not require the simultaneous participation of communication partners (Johnson, 2006).

Postal correspondence, for instance, is an example of asynchronous communication. Letters, postcards, or telegrams are sent and received at different times, with time delays ranging from hours to weeks depending on distance and delivery systems. Similarly, printed newsletters that invite responses or follow-up contributions are also of this type. In each case, the defining characteristic is the temporal separation between message production and response.

Similarly, asynchronous CMC allow users to respond at their convenience. *Email* is a classic example of asynchronous communication, as messages are sent and read at different times. *Forums and Discussion Boards*, such as those on Reddit or Quora, enable users to post messages and receive replies over time. *Social Media* platforms like Facebook, X (Twitter), and Instagram also support asynchronous interactions, as users can share posts and engage with them through comments and likes. *Blogs*, where written content is shared for others to read and comment on later, further illustrate this category.

**Hybrid Communication** is increasingly common in digital contexts, and IM provides an example of this mode. The classification of IM has been debated: some researchers (Baron, 2004; Schiano et al., 2002) describe

it as a form of synchronous CMC, where text messages are exchanged in real time via a web-based client (Peslak et al., 2010). Others, however, view IM as quasi-synchronous or even asynchronous, comparable to email or text messaging, since users frequently engage in multiple conversations simultaneously or respond while multitasking, which often results in delayed replies (Baron, 2013).

For the purpose of this dissertation, hybrid communication is defined as an interaction mode that combines elements of both synchronous and asynchronous communication, enabling participants to contribute either in real time or with temporal delays within the same conversational exchange. I consider IM dialogues as a type of hybrid communication: platforms such as WhatsApp support synchronous exchanges when users are online and responding quickly, yet they also allow delayed replies. Collaborative tools such as Slack or Microsoft Teams operate in a similar way, allowing immediate back-and-forth exchanges during active collaboration but also supporting delayed responses.

## 3.2 Features of IM dialogue

### 3.2.1 Technical features

IM dialogue is similar to face-to-face communication because it allows people to exchange messages quickly, respond in real time, and keep track of each other's participation in the conversation. Like spoken interaction, it supports immediate feedback, smooth changes in topic, and a shared sense of being "present" together. At the same time, IM is also similar to written communication because messages are typed, can be read again after they are sent, and often remain stored as a record of the exchange.

IM dialogue happens on platforms that offer a range of technical features that shape how users communicate online. Among these platforms, WhatsApp serves as an illustrative example due to its global popularity. By January 2025, WhatsApp had approximately 33 million downloads on the Google Play Store alone (Statista, 2025b), showing its central role in digital communication.

One of WhatsApp's defining features is that chat participants can communicate with an open conversation window almost real-time (Dürscheid and Frick, 2014). WhatsApp provides users with cues regarding interlocutors' presence and message reception, creating a sense of immediacy and presence. **Message status**, represented visually through ticks, offer feedback about delivery: a single grey tick (✓ ) indicates successful sending, double grey ticks (✓✓) signify successful delivery to the recipient's device, and two blue ticks (✓✓) confirm that the message has been displayed on the recipient's screen. For example, in Figure 3.1, the message *Love you all* 🤗 is marked as displayed to the recipients in the chat group.[1]


Figure 3.1: Message marked as read

[1]Screenshot from the official WhatsApp website. The "read receipt" indicator (two blue ticks) can be disabled in the app settings. When recipients choose to disable this option, senders can no longer infer from the ticks whether the message has been displayed on the recipient's screen. For details, see WhatsApp Help Center.

Many IM Apps, including WhatsApp, include a **message recall** function, also known as *Delete for Everyone*. This allows users to withdraw messages they've sent by mistake or no longer wish to share. After a message is recalled, it is removed from both there own and the recipients' chat interface (Figures 3.2, 3.3 and 3.4), leaving a notification that states, *This message was deleted*. In contrast, the *Delete for Me* option removes messages only from the sender's device, leaving the original content on recipients' devices.


Figure 3.2: Sent message before recall (Delete for Everyone)


Figure 3.3: Interface after recall (Delete for Everyone), sender's view


Figure 3.4: Interface after recall (Delete for Everyone), recipient's view

The availability and time limits of recall and edit functions, however, vary across platforms, as summarized in Table 3.1.

| **Platform** | **Edit** | **Time Limit for Edit** | **Recall** | **Time Limit for Recall** |
|---|---|---|---|---|
| WhatsApp | Yes | 15 minutes | Yes | 48 hours |
| Telegram | Yes | No time limit | Yes | No time limit |
| Signal | Yes | 10 times within 24 hours | Yes | 24 hours |
| iMessage | Yes | 5 times within 15 minutes | Yes | 2 minutes |
| Facebook Messenger | Yes | 15 minutes | Yes | No time limit |
| WeChat | No | N/A | Yes | 2 minutes |
| Microsoft Teams | Yes | No time limit | Yes | No time limit |
| Slack | Yes | No time limit | Yes | No time limit |
| Discord | Yes | No time limit | Yes | No time limit |

Table 3.1: Summary of recall and edit features in popular IM platforms [2]

[2]Data compiled based on testing and official documentation from WhatsApp, Telegram, Signal, iMessage, Facebook Messenger, WeChat, Microsoft Teams, Slack, Discord as of March 2025

Another noteworthy feature of IM platforms is the ability for users to edit sent messages. This **modification function** addresses a common frustration among users who notice typos, grammatical errors, or realize they have sent incomplete or inaccurate information after pressing the send button. With the edit feature, users can modify their messages within a time window that varies by platform. Edited messages are explicitly marked (for instance, with an *edited* label in WhatsApp or a pencil emoji ✏️ in Skype), so that all participants can see that an update has occurred. Figures 3.5 and 3.6, created to illustrate this function, show an example from WhatsApp. In Figure 3.6, according to the time that the message was sent, the speaker realized that the greeting *good morning* (in Figure 3.5) was no longer appropriate, and therefore modified it accordingly. This modification function differs significantly from face-to-face conversations or any other forms of CMC, where past utterances cannot be directly altered.


Figure 3.5: Message before modification


Figure 3.6: Message after modification

While recall and modification functions in IM offer practical benefits like error correction, their impact on conversational coherence remains complex and potentially problematic. Recalling messages disrupts dialogue by creating noticeable gaps, indicated only by a removal notification. Such gaps can confuse participants, who might wonder about the content and reasons behind the deletion, thus, this function interrupts the natural progression and flow of conversation. Similarly, when previous messages are edited, the original content is replaced, potentially confusing recipients who have already read and responded to the initial version. Although the *edited* label provides notification, it does not specify what and when content changed, leaving room for misinterpretation or uncertainty. The impact of message editing on discourse coherence remains unclear, particularly whether recipients notice the change, and how the timing of an edit influences the coherence of the ongoing interaction. In Chapter 6, I will empirically investigate how message modifications affect discourse coherence and pronoun resolution.

An additional essential feature is the **quoting function** (also known as "message replies" or "quoted replies"), it is designed to enhance communication by allowing users to respond directly to a specific message within a chat. Especially in busy or fast-moving group conversations, quoting ensures clarity by visually connecting the response directly to the original message. This is particularly helpful in maintaining conversational coherence and clarity within crowded or overlapping interactions, further discussions, see Chapter 4, section 4.1.2.

Lastly, WhatsApp and other IM apps allow users to react to messages quickly and visually with emojis. This **comment with emoji function** lets participants respond with a single tap. By choosing an emoji as a comment, users provide efficient feedback to a specific message. The selected emoji appears beneath the original message, visible to all participants. Figure 3.7,[3] for example, shows users reacting with a ❤️ and 👍 emoji to a shared photograph in the chat.

[3]Screenshot from the official WhatsApp website.

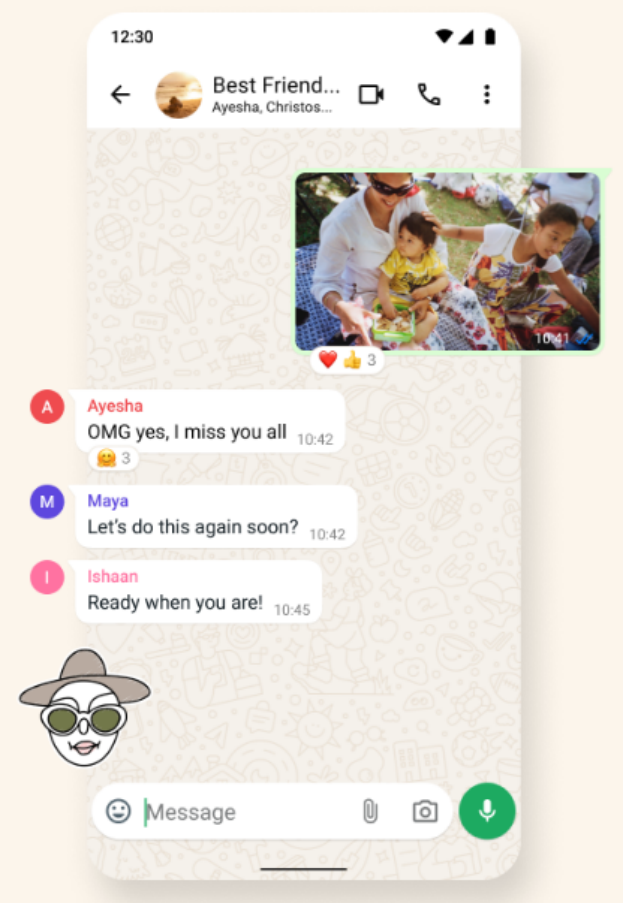



Figure 3.7: Example of the comment with emoji function in WhatsApp

However, the emoji reaction function differs from sending emojis as *part of a message*, whether alone or alongside text. Emoji reactions are visually tied to a specific earlier message rather than sent in the flow of new contributions. While in-message emojis enrich or modify an utterance's meaning, emoji reactions serve primarily as quick feedback, making them ideal for acknowledging messages or expressing emotions.

### 3.2.2 Informality and Spontaneity

IM offers a platform that combines the immediacy of spoken conversation with the permanence of written text. For instance, the use of abbreviations, acronyms, colloquial expressions, and minimal punctuation contributes to creating an informal way of communication (Baron, 2004; Crystal, 2006). To illustrate these features, consider the following exchange between two teenage users, adapted from Craig (2003):

(15) a. Girl One: sorry im talkinto like 10 ppl at a time
b. Girl Two: u izzyful person
c. Girl Two: kwel
d. Girl One: hey i g2g
e. Girl One: sry but my dinner is ready

In this exchange, abbreviations like “ppl” (people) and expressions such as “g2g” (got to go) show how IM speakers use reduced written forms that establish a casual tone.

Crystal (2006) describes IM as a new form of written communication, with features such as minimal capitalization (Thurlow, 2003) and the widespread use of acronyms, including sentence-length examples like AYSOS (“Are you stupid or something?”) and GTG (“Got to go”) (Crystal, 2006). Thurlow (2003) and Varnhagen et al. (2010) both propose a systematic and detailed classification for words and spellings that are especially used in IM. Varnhagen et al. (2010) consider shortcuts, pragmatic devices, and errors as the three main levels, with each containing several subcategories, as illustrated in Table 3.2.

| Category | Example | Example in context |
|---|---|---|
| Short cuts | | |
| Insider word | hottie | MAKE OUT WITH THAT HOTTIE |
| | fugly | is that like a fugly slut? |
| Abbreviation | feeling | how r u feelin? |
| | prolly | you could prolly look them up on the net |
| Word combination | wanna | i wanna sign up for the yhl |
| | gonna | i am gonna be gone sat and sunday |
| Acronym | bf | SHE HAS A BF |
| | omg | omg that is terrible |
| Alphabet/letter | u | what do u wanna talk about? |
| | 2day | did u go to skool 2day? |
| Phonetic | yer | you get yer stuff done tonight? |
| | wat | wat u doin? |
| Lower case | I | i almost cried |
| | elyssa | r u talking to elyssa????? |
| Contraction | im | im so excited |
| | thats | thats not cool |
| Pragmatic devices | | |
| Emotion word | hahahaha | hahahaha okay |
| | soooooo | wow im soooooo glad |
| Emotion acronym | lol | lol im not talking to you |
| | omg | omg for social we have to do this... |
| Upper case | THAT | Not THAT nervous though. |
| | WASH | u might have to actualy WASH that sweater |
| Emotion punctuation | ..... | so i was thinking..... |
| | :) | :) doo it hahah |
| Errors | | |
| Typographical error | carzy | im too carzy |
| | frwnch | just finished studying 4 frwnch |
| Misspelling | embarrasing | how embarrasing.... |
| | protect | it's for like a psychology progect |

Table 3.2: New language categories and examples of new language (Varnhagen et al., 2010)

König (2015) observes that WhatsApp communication differs from SMS in the way adjacency pairs are realized. Writers tend to split adjacency pair parts across multiple short messages, thereby creating a turn-by-turn rhythm that resembles spoken dialogue. As in this example:

(16) a. A: Hey Laura :-) Frohes neues Jahr erstmal noch! Die Arbeit hat doch ganz normal offen, oder? Weil ich sitze hier alleine und warte auf Annika;-)
*Hey Laura :-) Happy New Year to you, too! The office is open as usual, right? Because I'm sitting here alone and waiting for Annika ;-)*

b. B: Hey! Dir auch ein frohes neues Jahr
*Hey! A Happy New Year to you as well*

c. B: Ja, die Arbeit hat ganz normal geöffnet
*Yes, the office is open as usual*

d. B: Ist sie inzwischen da?
*Has she arrived by now?*

e. B: Aber du kommst auch alleine klar, oder?
*But you can manage on your own, right?*

f. A: Ja klar :-)
*Of course :-)*

Here, a single WhatsApp message containing multiple pair parts (New Year's greetings, a question, and an explanation) is answered by the recipient in a series of short, consecutive messages, each addressing one part of the first turn.

IM dialogue is also shaped by sociolinguistic factors such as age, gender, and cultural background. Younger users often show greater creativity and flexibility in adopting nonstandard forms, whereas older users tend to follow conventional norms more closely (Ling and Baron, 2007). Research on gender differences indicates that women generally use more expressive language and emoticons than men, who more often favor concise, task-oriented communication (Baron, 2004). Cross-cultural studies further show

variation: compared to North America, users in Asian contexts integrate more multimodal elements, such as multi-party chat, audio-video chat, and emoticons (Kayan et al., 2006).

### 3.2.3 Multimodality

Multimodality refers to an approach that emphasizes how multiple semiotic resources, such as language, images, sounds, and gestures, integrate and interact to create coherent meaning in communication (Jewitt et al., 2016; Jewitt, 2017). In IM dialogues, multimodality enriches communication by providing additional resources for users to convey meaning.

**Visual cues**, such as emojis, stickers, images, and GIFs, help enhance expressiveness. Emojis, in particular, are widely used across languages and cultures. According to Kelly and Watts (2015), emojis serve as emotional cues, helping users convey feelings that might be difficult to express through text alone. Emojis are not only decorative but play a pragmatic role by adding emotional cues to plain textual messages.

The use of visual cues in IM are not limited to emojis. Stickers, images, and especially animated GIFs are often more elaborate and context-specific. They are widely used to convey emotions, reactions, and even complex ideas. Schneebeli (2019) found that GIFs can effectively translate nonlinguistic cues, such as facial expressions and bodily actions, into the digital domain. They enable users to express feelings and affect in ways similar to face-to-face conversations, bridging the gap between digital and physical interaction. Herring and Dainas (2017) found that different different visual resources tend to fulfill specialized roles in conversations. For example, emoticons serve to express emotional reactions, while stickers are commonly used for interactional exchanges (reactions and mentions). Emojis are used especially to express reactions. This differentiation shows the specialized communicative functions that different visual resources fulfill within IM dialogues.

**Audio cues** such as voice messages offer a faster and more convenient alternative to typing, allowing users to multitask, such as recording a voice message while walking or participating in other activities. According to Sampietro and König (2024), voice messages are particularly effective for certain types of communication, such as long explanations, detailed narrations, or "vocal performances." Additionally, they are often favored for important and emotional topics (Haas et al., 2020), where intonation plays a critical role in conveying meaning.

While visual and audio cues highlight the diversity of semiotic resources available in IM dialogue, recent research suggests that these resources should not only be examined in isolation, but also in terms of how they are sequentially organized within interaction. König (2024) proposes the notion of **transmodal interaction** to describe how participants sequentially coordinate text and voice messages in WhatsApp chats. In this view, both textual and audio contributions are treated as interactional units that are ordered temporally and related to one another in the unfolding discourse. Messages build on prior contributions and shape subsequent responses, thereby forming a coherent interactional structure. This perspective also highlights that modality choice itself carries interactional meaning: voice messages, for instance, are perceived as more personal and committed contributions than text messages (König, 2024).

### 3.2.4 The use of emoji

Emojis first appeared in Japan (Holvikivi, 2019), with the term "emoji" originating from the combination of the Japanese words for picture ("e") and character ("moji"). They are small graphic symbols that represent various objects, ideas, or emotions from everyday life and rapidly become a universal method of digital expression. According to Emojipedia (2025), emojis can be classified into various categories, including smileys, people, animals and nature, food and drink, activity, travel and places, objects, symbols, and flags. Their number and diversity have significantly increased over the years, from only 76 in 1995 (Statista, 2025a) to 3,790 by September

2024 (Emojipedia, 2025), reflecting the expanding complexity and richness of digital communication. Due to Unicode standardization, emojis appear consistently across digital platforms, although their visual representations may differ slightly depending on the software or device. For instance, a "smiling face emoji" may appear differently on Google's platform (☺️) compared to Microsoft's (😊)[4].

**Clarification of the terms** Given the popularity of emojis, it is necessary to first clarify related terminology to avoid confusion, as terms like pictograms, emoticons, emojis, and stickers have been used inconsistently by researchers. For instance, Li and Yang (2018) used the term "emoticon" as a category that encompasses text-based emoticons, icon-style emoticons, and sticker emoticons. In contrast, Wang et al. (2014) used "emoticon" to refer specifically to graphic symbols that imitate facial expressions, what others might call emojis. Given this overlap, it is important to establish clear distinctions between these concepts.

In the context of IM, Arens (2014) clarified that **pictograms** can represent almost anything: animals, flowers, or even imaginary objects, but they do not include facial expressions. Therefore, symbols that mimic facial expressions should not be classified as pictograms.

The term **emoticon** is a blend of *emotion* and *icon*. According to the Oxford English Dictionary, an emoticon is "A representation of a facial expression formed by a short sequence of keyboard characters (usually to be viewed sideways), used in electronic mail, etc., to convey the sender's feelings or intended tone" (Oxford Online Dictionary, 2025). Examples of emoticons include textual sequences like ":)", ";-)", or ":D," to express happiness or ":-(" to express sadness.

The term **smiley** originally referred to :-), derived from the phrase "smiley face," which depicted a round, yellow face with a smile. Today, a smiley is a face-like symbol with various expressions and can be considered

[4]Google version retrieved September 2025; Microsoft version retrieved August 2025. For more information about smiling face emoji designs on different platforms, see https://emojipedia.org/smiling-face#designs.

Figure 3.8: A set of WhatsApp sticker named “cuppy”

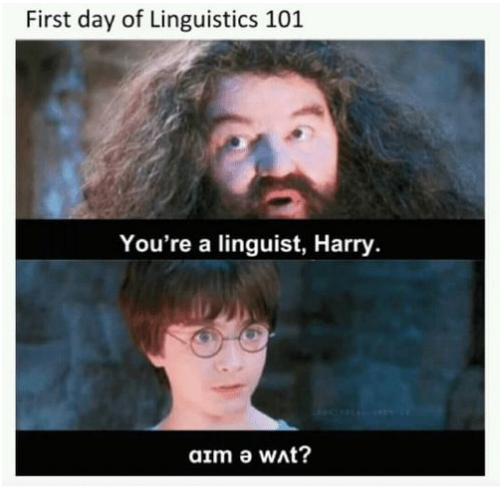



Figure 3.9: A meme

a subset of emojis.

Another category is **stickers**. According to De Seta (2018), stickers are “images, typically larger than graphical emoticons and emojis, offered as themed sets in the interfaces of instant messaging apps and social networking platforms (Figure 3.8[5]). They are often organized into tabs and personalized collections.” Internet **Memes** (Figure 3.9[6]) are usually images or GIFs that are often overlaid with text. They are copied and spread rapidly by users, often with slight variations. It reflects a shared cultural experience and evolves as it is remixed and reinterpreted by the online community (Shifman, 2013). Stickers and memes are characterized by their detailed, expressive designs and are typically used for reactions or humorous commentary. Unlike emojis and emoticons, stickers and memes usually cannot be inserted inline with text and are instead sent as standalone messages.

[5]Downloaded from: https://in.pinterest.com/pin/835769643328372178/ (accessed April 25, 2025)

[6]Downloaded from: https://share.google/t5y6fhsroIimugRmn (accessed April 25, 2025)

**Function of Emoji** Emojis are widely used across numerous platforms, appearing in monologues, stories on websites, social media posts (e.g., X or Instagram), and prominently in online chatting.

From a **syntactic** perspective, emoji can function as replacements or substitutes for words or letters (Dürscheid and Siever, 2017; Cramer et al., 2016). For instance, the sentence "I ❤️ u!" shows how the emoji ❤️ substitutes for the word *love*. In this case: "❌ 🎅 👉 🤞 ", emojis constitutes a full replacement of textual utterance, expressing *no gifts for you*. Emojis can also replace punctuation marks (Cramer et al., 2016; Hasyim, 2019; Albert, 2015), as in ending a sentence with a smiling face ☺️ instead of a period to convey friendliness or positive emotion, or pairing an exclamation mark with a celebratory emoji (*Happy birthday!* 🎉) to emphasize excitement. Emojis can also serve more elaborate syntactic functions, mirroring sentence structures and narratives. In guessing-games, "🎤 ☔" is interpreted *Singing in the Rain*, emojis directly correspond to a sentence-like structure, mirroring traditional syntactic patterns (Danesi, 2017).

**Semantically**, emojis have often been found to replace words in a message (Pohl et al., 2017). Siebenhaar (2018a) suggests that emojis can function semantically as substitutes at various linguistic levels: as replacements for subjects (17-a), verbs(17-c), entire propositions (17-b), or even complete communicative acts (e.g., narrations or descriptions (17-d)).

(17) a. 👸 fliegt mit 🚀 zu 👨 und 😽
👸 *flies with* 🚀 *to* 👨 *and* 😽

b. Ja eben und trotzdem kein 👫 sondern 👨 🏡 ———— 👸 🏡
*Yeah right, still no* 👫 *but* 👨 🏡 ———— 👸 🏡

c. Falls wir uns nicht vorher nochmal 📝, 👂 oder 👀, wünsche ich Dir einen guten 🏂 ins Jahr 12 10 11 14 [...] Also lass es 💥, 👊 rein und mach einen drauf...
*If we don't see each other again before then* 📝, 👂 *oder* 👀, *I wish you a good* 🏂 *into the year 12 10 11 14 [...] So let it* 💥, 👊 *in, and make the most of it...*

d. 🏃
🚌🏃
🚌🏃
🚌🏃
🚌🏃
🚌💭🚶
💭…🚶
🚶 nuttesohn *son of a bitch*

Hasyim (2019) also mentions that the semantic function of emojis is to "express connotation meanings in conversation," arguing that verbal text alone is often insufficient to convey speakers' feelings and emotions. However, the interpretation of emojis can vary, leading to potential miscommunication. As Miller et al. (2017) note, "in order to avoid miscommunication incidents, people must interpret emoji characters in their exchanges in the same way (and they must know that they are interpreting them the same way)". Studies have shown that the semantic interpretation of emojis can vary depending on factors such as gender (Chen et al., 2024), age (Weiß et al., 2020), context (Miller et al., 2017), platform (Franco and Fugate, 2020; Miller et al., 2016), and cultural background (Guntuku et al., 2019).

While much of the literature describes emojis in terms of substitution, recent work has taken a more formal perspective. Grosz et al. (2023) propose that face emojis, in particular, do not make independent assertions but often function to comment on the proposition expressed by the accompanying text. For example, in "Alex hates spaghetti 😟," the sad face does not influence the truth of the statement but signals the author's negative opinion toward the fact, compared with a discourse value (which is the expectation that Alex should enjoy spaghetti). In this sense, face emojis contribute affective or evaluative meaning that shapes the interpretation of discourse. This perspective supports the view that emojis are not merely decorative but are elements of multimodal discourse with specific semantic functions.

From a **Pragmatic** perspective, emoji function as tools for managing social interaction and conveying communication intentions. According to a corpus-based analysis by Li and Yang (2018), emoji fulfill multiple pragmatic roles: signaling attitudes or emotions, intensifying emotional expressions, modifying illocutionary force, indicating humor or irony, and facilitating conversational management (turn taking and backchanneling). For example, the 😠 emoji in the utterance "Sure, go ahead 😠" may expresses sarcasm, while the same sentence ending with a 😀 would convey agreement or glad (Pohl et al., 2017).

Cramer et al. (2016) further explains that emojis are particularly useful for managing conversations efficiently. They can serve as quick acknowledgments, provide a polite way to end a conversation, or act as a substitute for verbal responses when silence might seem rude. For instance, a simple smiley face 😄 allows users to respond without typing words. Emojis can also serve as comments or direct reactions to a previous utterance (Pohl et al., 2017; Dürscheid and Frick, 2014), as in example 18.

(18) a. Speaker A: 😢
b. Speaker B: 😪
c. Speaker A: Bisch no verrockt? (Bist du noch böse?)
*Are you still mad?*
d. Speaker B: Nei, und du? (Nein, und du?)
*No, and you?*

Additionally, Pohl et al. (2017) mention the decorative use of emojis, where they serve as flourishes to accompany text. For example, emojis can be used to add visual emphasis to congratulatory messages, such as "Happy birthday! 🎂 🎁."

Taken together, the features of IM dialogue help combine the immediacy of spoken interaction with the persistence and flexibility of written text. Functions such as message editing, quoting, or emoji reactions not only expand the ways in which participants can respond, but also reshape the temporal and structural organization of conversations. These affordances

challenge traditional expectations of how dialogue unfolds, particularly in terms of coherence and recency. In the following chapter, I will build on these insights to examine how coherence is established and maintained in IM, with a special focus on how feedback and recency are redefined within this hybrid communicative environment.

# Chapter 4

# Discourse Coherence and Recency in IM

Chapter 2 discussed the notion of prominence in discourse, drawing on theoretical accounts that characterize prominent entities as those that are singled out, structurally central, and dynamically updated. It also reviewed linguistic and cognitive factors that influence prominence, such as syntactic roles, information structure, and discourse relations. Building on this foundation, Chapter 3 turned to the medium-specific features of IM. It showed how technical features such as message editing, quoting, and emoji comments influence IM discourse by allowing interlocutors to revisit earlier utterances. These features complicate the linear flow of dialogue and make discourse coherence in IM especially sensitive to the way recency is construed.

This chapter develops the notion of *recency* as it will be used in this dissertation. Rather than treating recency simply as a matter of temporal order, I conceptualize it as a form of distance, specifically the separation between a referent's introduction and its subsequent mention along two dimensions: temporal distance and spatial distance. While temporal and spatial distance are usually aligned in spoken and written discourse, IM

introduces the possibility that they may diverge. This divergence raises questions about how coherence is maintained when the order of appearance on screen does not match the order of production and perception in time.

The chapter proceeds in three main steps. Section 4.1 discusses discourse coherence in IM, focusing on turn-taking and feedback in IM discourse. Section 4.2 introduces the distinction between temporal and spatial distance as two dimensions of recency. Finally, Section 4.3 considers cases where these two dimensions align or misalign, and discusses the implications of misalignment for reference processing in IM dialogue.

## 4.1 Discourse coherence in IM

Coherence refers to the logical connection and consistency of discourse, ensuring that a text or conversation is understandable and meaningful. In both spoken and written discourse, coherence is achieved through a combination of local and global mechanisms. *Coherence at the level of individual entities* refers to the relatedness of text at the level of sentence-to-sentence transitions (Barzilay and Lapata, 2008). Entity-based approaches to local coherence argues that discourses are about an entity, and continue mentioning the entity from sentence to sentence (Jurafsky and Martin, 2025). There are several theoretical perspectives on discourse coherence and prominence. In the influential framework of Halliday and Hasan (1976), coherence is closely related to cohesion: linguistic devices like pronouns, conjunctions, and lexical repetitions linking sentences together. Other theories explore coherence through notions of topicality (Chafe, 1974), predictability (Kuno, 1972), and accessibility (Gundel et al., 1993). These perspectives collectively underscore the importance of maintaining coherence to ensure logical flow within a discourse.

Beyond sentence-to-sentence transitions, *coherence at the level of textual units* concerns broader thematic and structural connections of a discourse. RST is one influential model addressing global coherence (Mann and Thompson, 1988). According to RST, coherence is maintained through relations between textual units, typically consisting of a nucleus (the central, independent unit) and a satellite (the supporting unit, dependent on the nucleus for interpretability). Global coherence is established through rhetorical relations (e.g., elaboration, justification, contrast) that helps to organize discourse into coherent, purpose-driven units.

Studies in conversation analysis have further enriched our understanding of coherence in real-life interactions. Schegloff (2007) highlights how sequential organization of turns, use of repairs, and strategic pauses contribute to the dynamic process of meaning-making. However, IM dialogue presents unique challenges to discourse coherence. As Siebenhaar (2018b) argues, chat communication is conceptually grounded in spoken dialogue yet realized in written text. This duality, combined with IM-specific features such as message editing, deletion, and multimodal elements, introduces additional complexity while at the same time, offering new opportunities for creating meaning and maintaining interactional flow.

Thus, the study of discourse coherence in IM dialogue requires attention both to linguistic strategies (as outlined in Chapter 2) and to the ways in which technological features shape conversational interaction (as discussed in Chapter 3).

### 4.1.1 Turn-taking and relevance

Turn-taking is the process by which conversational participants alternate speaking roles to create a coherent and coordinated dialogue. In face-to-face interactions, turn-taking has been studied to reveal how individuals manage speaking rights, overlaps, and silences.

Two foundational approaches have shaped our understanding of turn-taking. Duncan (1972) investigates the conversational signals that speakers and listeners use to coordinate turns. In face-to-face communication, *turn-yielding signals* indicate when a speaker is ready to yield their turn. Examples include intonation cues such as a rising or falling pitch at the end of an utterance; body motion cues such as relaxing previously tense gestures; and sociocentric sequences such as phrases like *but uh* or *you know* that signal a possible transition point; and syntactic cues like the completion of a grammatical clause serve to mark the end of a speaking turn. On the other hand, speakers use *attempt-suppressing signals* such as continuous hand gestures to indicate that they may wish to maintain their turn. Listeners complement these cues with *Back-channel signals*, including verbal affirmations ("mm-hmm"), non-verbal gestures (such as nodding), or brief requests for clarification, demonstrating their attentiveness.

Sacks et al. (1974) propose a *rule-based model* of turn-taking. They argue that conversation is managed through rules that operate at transition-relevance places, which are points in a conversation where a change of speaker is possible. These rules can be summarized as follows:

- If the current speaker selects the next speaker, the selected person has the right and obligation to speak next.
- If the current speaker does not select the next speaker, any participant may self-select; the first to begin gains the turn.
- If no other participant self-selects, the current speaker may continue with another turn-constructional unit.

Building on this understanding, Clark and Wilkes-Gibbs (1986); Clark (1996) explored the collaborative nature of dialogue, emphasizing the importance of grounding, where interlocutors build shared understanding through mutual signals and feedback.

In IM, messages are sent and displayed in a chat bubble as a unit, and several participants may be typing at once. As a result, the sequential order

of displayed messages may not reflect the actual timing of their production, raising questions about how relevance and coherence are managed in this environment.

**Loosened coherence** Herring (1999, 2013) characterized IM conversations as having *loosened coherence*, while Dürscheid and Frick (2014) similarly described this phenomenon as a *loosened structure*. Loosened coherence happens because conversational turns in IM do not always follow immediately or sequentially, leading to a more fragmented structure. Specifically, Herring (1999) identified two major obstacles to coherence in CMC:

1. Lack of simultaneous feedback: caused by reduced audiovisual cues and the fact that messages cannot overlap. The absence of visual and auditory cues requires participants to rely on textual cues to manage turn-taking.

2. Multi-Participant interaction: when multiple participants are involved in a chat, a message may be separated in linear order from a previous message it is responding to, if another message or messages happen to have been sent in the meantime.

A distinguishing feature of turn-taking in IM is the prevalence of *fragmented and overlapping turns*. Unlike face-to-face conversation, where overlap and interruptions are typically minimized and quickly resolved, IM diaogues often involve multiple participants sending messages simultaneously, making turn sequences less predictable. This phenomenon is illustrated by Figure 4.1, elicited from Dürscheid and Frick (2014). Here, speaker B's utterance (19-f) logically continues an earlier comment (19-d) but is disrupted by speaker A introducing a new topic (19-e). The entire interaction occurs within a short three-minute window, suggesting

simultaneous typing and minimal time differences between sent messages. Even immediate responses in IM may experience slight delays, influenced by factors such as bandwidth issues, network traffic, or technical glitches (Dürscheid and Frick, 2014), further complicating coherent turn-taking.


Figure 4.1: Loosend coherence due to asynchronicity

(19)
a. A: Yes, nevertheless... I have some new ammunition so we can gripe about NAME.
b. B: ?!?! Tell me.
c. A: I'll tell you tomorrow it's a bit tricky via WhatsApp.
d. B: Okay... then I'll tell you the rest too. ☺
e. A: And my date from today met someone else over the weekend.
f. B: And you from yesterday
g. B: No?!?! Really?!
h. A: Yeah, well, that could just as easily happen to me.

Example (20) from Herring (2013) illustrates multi-participant interaction challenges in IM. Two parallel discussions occur simultaneously, causing disrupted turn adjacency. Arch's narration of personal needs (line 23, 25, 26) is interrupted by Auxilary responding to Mike (line 24), Arch later joined the discussion of "tuxedo mask" (line 30 and 31), again interrupting Auxilary's response to WendyCA about a credit card application (line 29 and 32).

(20)
**9***** Arch- (ademon@netcom10.netcom.com) has joined channel #bayarea

**10*** Arch- yawns.. hey folx

(...)

**23**<Arch-> I need a nap

**24**<Auxilary> mike: will u dress up like toxedo mask for ya?

**25**<Arch-> I need to fill my fridge with food

**26**<Arch-> I need to get laid

**27**<Auxilary> err for me

**28**<WendyCA> the way to a woman's heart is through your credit card

**29**<Auxilary> wendy: i got a VISA platinum

**30**<Arch-> sorry no tuxedo.. nor black cape.. although I might be tempted to dress up as the Moonlight Night

**31**<Arch-> Knight

**32**<Auxilary> wendy: it has a 25k limit on the card :)

**33**<WendyCA> sorry, I require the guys I date to have a personality :)

**34**<Arch-> If you get a promotion over a woman it's favoritism, if a woman gets a promotion over you it's equal opportunity.

**35**<Auxilary> wendy: u can date my right hand

Specific devices have been developed in order to maintain coherence in CMC. Werry (1996); Herring (1999) mentioned *addressivity* in multi-party synchronous conversation, where speakers indicate the intended addressee by including their name at the start of an utterance, as in example (20), line 24 and 29, Auxilary directs messages to different individuals by starting with their names. Herring (1999) also introduced *linking* as another strategy, where participants re-mention a previous message in their utterance, such as when a response begins with, "I would like to respond to Diana's comment about landmines". Another strategy, known as *quoting* (Herring, 1999), involves copying portions of a previous message in one's response, serve as a subtype of linking. Quoting is not unique to instant messaging: it has

long been a feature of email and earlier text-based platforms, where users manually inserted or prefixed quoted lines (often with the symbol ">") to mark earlier contributions. For instance, in example 21, the speaker first quotes the message before presenting their own utterance. Modern IM applications adopt similar principles but offer technical support, enabling users to select and attach earlier messages directly to their replies.

(21) janed@ABC.bigtel.com (Jane Doe) writes:
>I can't believe how horrible Natalie looks. Has
>she put on a lot of weight?
I agree, but she has always had a somewhat round face, so if she did put on weight, I think that would be accentuated.

Additionally, Berglund (2009) studied disrupted turn adjacency in IM and found that such disruptions do not necessarily lead to misunderstandings or confusion. This is because IM tools provide timing information and other features that help maintain coherence, even when turns are not sequential.

For instance, in example (22) from Berglund (2009), Dina (line 8) returned to Bea's earlier statement (line 3) using anaphoric reference ("it") and lexical repetition. Almost simultaneously, Dina added a new elaboration about the software (line 9). Bea's reply ("Yes ... I saw it today," lines 10–11) appeared right after Dina's elaboration, which could make the antecedent ambiguous. However, because the timing showed that Dina's elaboration (line 9) and Bea's "Yes" (line 10) were posted at nearly the same time, Bea's contribution was interpreted as the second pair part to Dina's earlier clarification request (line 8). Thus, timing, sequential structure, and distinctions between given and new information worked together to maintain coherence, even though turns were disrupted.

(22)

| Line | Mode | Time | From | To | Message |
|---|---|---|---|---|---|
| 1 | MSN | 21:03:19 | Dina@school... | beay@umea | u know adobe cs3 will release |
| 2 | MSN | 21:11:49 | beay@umea | Dina@school... | Err |
| 3 | MSN | 21:11:05 | beay@umea | Dina@school... | i think mr. charles is already using it |
| 4 | MSN | 21:12:41 | beay@umea | Dina@school... | beay@umea sends Farewell_TCAng 157.jpg |
| 5 | MSN | 21:13:43 | beay@umea | Dina@school... | You have successfully received C:\Documents and Settings\zhghau04\Desktop\Farewell_TCAng 157.jpg from beay@umea. |
| 6 | MSN | 21:13:57 | beay@umea | Dina@school... | thi sphoto was shoot when we in Malaysia |
| 7 | MSN | 21:14:35 | Dina@school... | beay@umea | haha, u have potensial at that time |
| 8 | MSN | 21:14:46 | Dina@school... | beay@umea | really?he already use it<???? |
| 9 | MSN | 21:15:00 | Dina@school... | beay@umea | sam said in embbed flash inside |
| 10 | MSN | 21:15:00 | beay@umea | Dina@school... | Yes |
| 11 | MSN | 21:15:03 | beay@umea | Dina@school... | i saw it today |
| 12 | MSN | 21:15:04 | beay@umea | Dina@school... | really? |
| 13 | MSN | 21:15:06 | beay@umea | Dina@school... | Coooooooll |
| 14 | MSN | 21:15:30 | Dina@school... | beay@umea | i heard that, so that's why i want to tell u |
| 15 | MSN | 21:16:14 | beay@umea | Dina@school... | but it hink the cs2 have also simillar |
| 16 | MSN | 21:16:16 | beay@umea | Dina@school... | but easy flash |
| 17 | MSN | 21:16:18 | beay@umea | Dina@school... | i dunno man |
| 18 | MSN | 21:16:28 | beay@umea | Dina@school... | but is going to be cool |

**Utterance chunking** Another feature of turn-taking IM dialogue is Utterance chunking, where speakers break their contributions into multiple, shorter messages. This behavior is often driven by the medium's constraints, such as limited message lengths, typing speed, or the speaker's attempt to maintain the recipient's attention. König (2015) found out through a corpus study that a tendency in WhatsApp dialogues is that users tend to break IM turns into separate messages, as shown in example (23).

(23) a. A: Habs voll vercheckt (29.04.2014 - 08:50:00)
*Totally forgot*

b. A: Deine ma hat gestern angerufen weil sie heute nach Paris kommt und deine Schwester nicht da ist (29.04.2014 - 08:50:30)
*Your mom called yesterday because she is coming to Paris today and your sister isn't there*

c. A: Und sie wollte wissen ob du sie vom Bahnhof abholen kannst (29.04.2014 - 08:50:45)

*And she wanted to know if you could pick her up from the train station*

d. A: Sorry (29.04.2014 - 08:51:00)
*Sorry*

e. B: Kein Problem, sie hatte mich auch aufm Handy angerufen (29.04.2014 - 08:56:00) *No problem, she had also called me on my mobile*

In this example, Speaker A first signals forgetfulness, then elaborates across two subsequent messages on what was forgotten, and finally adds an apology. The contributions are semantically connected but distributed over several short messages.

Utterance chunking can be used to manage multiple topics within a single turn. Example (24), adapted from Crystal (2006), illustrates such *multi-theme chunking*.

(24) a. Emma: anything else you need for The day that we could bring?
b. Joe: me
c. ...
d. Jane: Will mail the address of the ocation too just in case they are delayed
e. Jane: hello!
f. Jane: not absorbed then!
g. Jane: baby lables
h. Jane: (labels)
i. Jane: and taxis

Here, Jane sends several short messages, each addressing a different point. Messages (24-e) and (24-f) respond to Joe, who has just joined the conversation; messages (24-g) and (24-i) address separate questions Emma had asked earlier; and message (24-h) is a self-correction.

Another type of chunking is *single-theme chunking*, where a single semantic point is divided into multiple messages, as in (25-a) and (25-d). Jeff breaks his opinion into two messages, interrupted by Matt [1].

(25) a. Jeff: There are so much bad design
b. Matt: no kidding
c. Matt: But I still get surprised sometimes
d. Jeff: And so much bad grammar
e. Matt: Bad grammar are everywhere

**IM features and their role in turn-taking** Unlike face-to-face interaction, where turn-taking is achieved by both verbal and non-verbal cues, IM communication takes place primarily through text and multimodal elements. The lack of immediate visual and auditory feedback introduces challenges to synchrony, since participants cannot rely on the same temporal coordination that regulates turn-taking in spoken dialogue. Consequently, IM platforms have introduced technical features designed to compensate for the absence of traditional non-verbal cues. These features help users manage conversational coherence and turn-taking effectively.

*Typing indicators* are often visually represented as "User is typing..." or by animated symbols (e.g., three moving dots within a speech bubble, see Figure 4.2). They act as digital counterparts to face-to-face attempt-suppressing signals, indicating to conversation partners that someone is currently composing a message. Such indicators aim to help prevent overlapping or simultaneous message exchanges, providing a sense of continuity and predictability that is otherwise missing in text-based communication.

[1] Example taken from Voida et al. (2002)

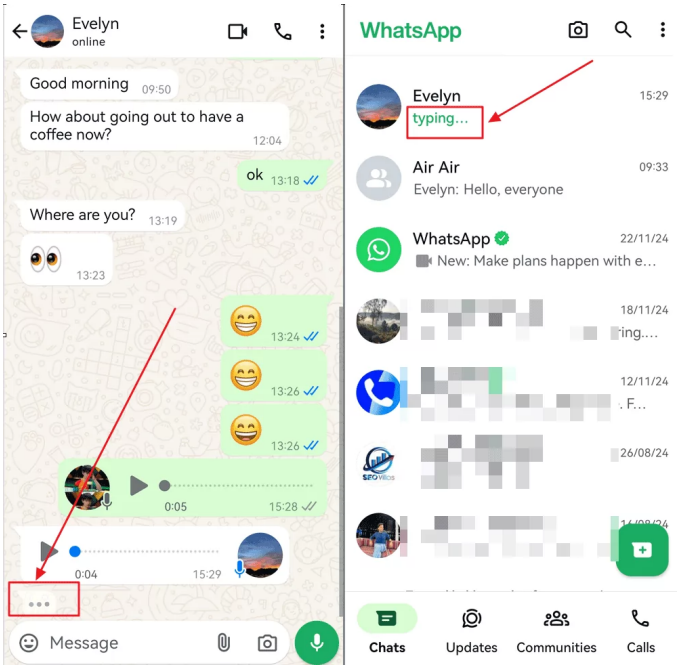



Figure 4.2: WhatsApp typing indicator. Source: AirDroid (2024)

*Read receipts* indicated through visual cues like double checkmarks (as in WhatsApp) or labels such as *seen* serve as digital equivalents of back-channel signals, providing information to the sender that their message has been delivered and displayed, but not necessarily that it has been read or understood. Still, when it is clear that both interlocutors are online and actively engaged in an exchange (e.g., online or typing indicators are visible), the speaker may treat read receipts as a signal of attention, similar to a nod or verbal acknowledgment like "mhm" in face-to-face conversation. It gives the speaker reassurance that their messages have reached the recipients, thus reducing uncertainty regarding message reception. Read receipts influence turn-taking by shaping expectations: once a sender sees that their message has been read, they may wait for a response and pause before sending further contributions. However, such indicators are imperfect: they can usually be disabled by recipients, may be delayed by connectivity or device settings, and do not guarantee attention, comprehension, or uptake.

### 4.1.2 Feedback

Feedback is crucial for maintaining a coherent discourse. It acts as a dynamic mechanism that ensures mutual understanding, alignment of meaning, and the continuous adaptation of messages among interlocutors.

**The notion of feedback** Effective conversations rely on coordinated interaction. The speaker must ensure they are being listened to, heard, and understood, while the listener provides clear signals to confirm this (Clark and Schaefer, 1989). Unlike monologues, conversations are inherently interactive. Clark and Schaefer (1989) argues that each contribution involves two phases: a *presentation phase*, in which the speaker provides information, and an *acceptance phase*, in which the listener either confirms understanding or signals trouble in understanding. During the acceptance phase, listeners typically provide either positive evidence, indicating successful understanding, or negative evidence, signaling difficulties or misunderstanding (Brennan and Hulteen, 1995).

Clark and Schaefer (1989) identified four possible states representing different levels of understanding that a listener (B) may achieve regarding a speaker's (A) utterance

**State 0** B did not notice that A uttered any U.

**State 1** B noticed that A uttered some U (but was not in state 2).

**State 2** B correctly heard U (but was not in state 3).

**State 3** B understood what A meant by U.

Corresponding to these listener states, Clark (1994) outlines four related levels of action for the speaker (A):

**Level 1** Vocalization and attention

**Level 2** Presentation and identification.

**Level 3** Meaning and understanding.

**Level 4** Proposal and uptake.

At each level, the listener provides evidence that signals progress toward understanding. If no such evidence is provided, the speaker must adapt their approach to resolve any misunderstandings or to clarify their intended meaning.

**Feedback in face-to-face conversation** The basic mode of language is face-to-face dialogue (Clark, 1996). As previously discussed, successful communication involves an acceptance phase, in which listeners provide clear evidence of their understanding or misunderstanding of the speaker's utterance (Clark and Schaefer, 1989). Specifically, feedback during this phase can be categorized into two main types: *evidence of understanding* and *evidence of trouble in understanding*.

Clark and Schaefer (1989) identify five primary types of evidence of understanding: *Continued attention*, where the addressee shows that they are still attending to the speaker's contribution and are satisfied with the presentation. *Initiation of the relevant next contribution*, by which B starts in on the next contribution that would be relevant at a level as high as the current one. *Acknowledgment*, B nods or says "uh huh," "yeah," or the like. *Demonstration*, B demonstrates all or part of what he has understood A to mean. And *Display*, B displays verbatim all or part of A's presentation.

Other scholars have offered similar classifications. Allwood et al. (1992) describe reaction types such as contact, perception, understanding, and attitudinal responses, which correspond closely to Clark and Schaefer's notions. Schegloff (1982) further emphasizes that conversation is an

interactive achievement, where brief vocalizations like "uh huh" and non-verbal signals (e.g., nods) play a key role. Gardner (1998) refines this by distinguishing among minimal response tokens: "Mm," "Yeah," and "Mm hm", each conveying varying degrees of acknowledgment.

When understanding fails, listeners provide *evidence of trouble in understanding*. Clark and Schaefer (1987) outline four states of understanding, ranging from State 0 (no awareness of the utterance) to State 3 (full comprehension of the utterance). Listeners may signal trouble through various strategies, such as: asserting no hearing, presuppose no hearing, presuppose incomplete hearing and presuppose fallible hearing.

These strategies are often referred to as clarification requests (CRs). Derriks and Willems (1998) categorize communication difficulties into levels such as linguistic (phonetic or syntactic issues), pragmatic (reference or intention recognition), and cognitive (comprehension or memory failures). CRs can take various forms, including partial CRs (focusing on unclear parts), alternative CRs (offering interpretations), and reformulations (rephrasing for clarity).

Rodríguez and Schlangen (2004) classify CRs by surface form (e.g., questions, repetitions) and function (e.g., identifying the source of the problem). Schlöder and Fernández (2015) extend this by categorizing questions based on the listener's state of intention, such as low CRs (clarifying content) or intention adoption CRs (accepting the speaker's goal). Listeners may also provide implicit feedback, for example through gaze, hesitation, or gestures, which indicate trouble in understanding.

**Feedback in IM dialogue** In face-to-face conversations, non-verbal cues such as nodding or shaking one's head provide synchronous feedback, allowing speakers to track the listener's understanding in real time. In IM interactions, these visual and auditory cues are absent, and the timing of feedback is less flexible: it can only be given once a message has been sent

and displayed, not during its production.

As discussed in Section 3.1, IM belongs to hybrid communication, because it has features of both synchronous and asynchronous communication. Feedback features in IM, such as *read receipts*, are designed to mediate between these two modes. On platforms like WhatsApp, a message goes through three phases after the user clicks the "send" button (Figure 4.3):


Figure 4.3: Message status phases in WhatsApp. Source: (AirDroid, 2025)

**Phase 1** A single grey tick indicates that the message has been sent successfully. At this stage, the addressee is in State 0 of grounding: they have not yet noticed any utterance.

**Phase 2** Double grey ticks indicate that the message has been delivered to the addressee. This only ensures that the message is available to be noticed, but it does not guarantee that the addressee has actually seen it. Technically, the addressee may still be in State 0.

**Phase 3** Double blue ticks indicate that the message has been displayed on the recipient's interface. At this point, it is reasonable to assume that the addressee has at least noticed the utterance, corresponding to State 1 of grounding. However, whether the message is fully understood (State 2) is not ensured by the platform.

At Phase 3, the system indicates that the message has been opened in the chat interface. This creates the opportunity for grounding to advance, but it does not itself constitute evidence of understanding. For successful communication, the addressee still needs to provide uptake, either by responding in a way that demonstrates understanding or by signaling difficulty, thereby contributing to the grounding process.

Besides read receipts, IM enables other forms of feedback. Users may type and send *verbal feedback* such as emotion words like "haha," "arrrgh," and "ahhh," similar to those in face-to-face conversation. Alternatively, emojis provide *non-verbal* feedback, comparable to facial expressions or gestures in face-to-face interaction.

Emojis signal agreement, disagreement, surprise, amusement, or other responses, helping to maintain discourse coherence. Li (2021) investigated the use of emoji as feedback strategy in IM dialogue and found out that emojis such as *face with raised eyebrow* (🤨), *face with monocle* (🧐) and *thinking face* (🤔) provide evidence of trouble in understanding and

emojis like *OK hand* (👌), *clapping hands* (👏) and *thumbs up* (👍) provide evidence of understanding.

IM Apps provide multiple methods for emoji-based feedback. Users can respond directly by sending standalone emojis (Figure 4.4) or by using the emoji reaction function, as shown in Figure 4.5.

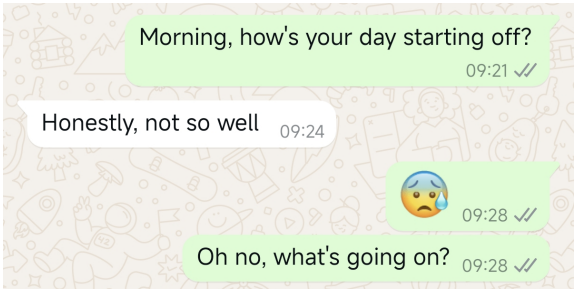



Figure 4.4: Emoji feedback in chat

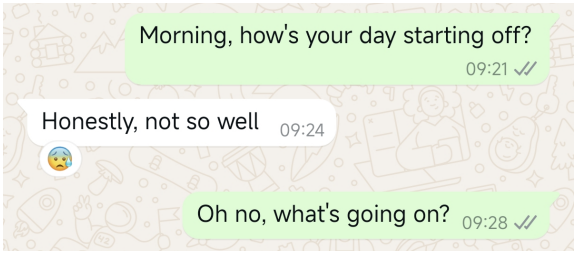



Figure 4.5: Emoji feedback using comment function

When two speakers type simultaneously in IM, messages may overlap (Figure 4.6). But if the conversation in Figure 4.6 happens face-to-face, when the speakers realize that they are talking at the same time, one of the speakers would probably stop talking and wait for the other speaker to finish. In such interactions, comments on earlier utterances are relatively uncommon, since listeners can provide feedback at the moment they wish to do so. But when a speaker does want to comment on a previous contribution, they cannot simply "travel back in time" to insert the feedback; instead, they must explicitly reintroduce the earlier utterance. In the case of Figure 4.6,

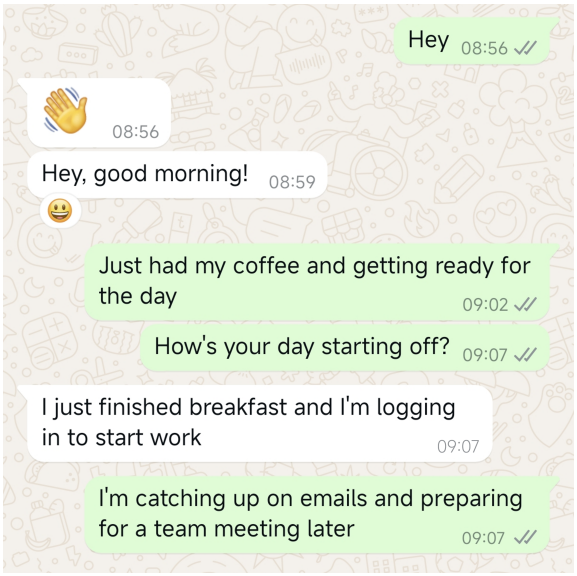



Figure 4.6: Message overlaps

if the speaker wants to finish the utterance of catching up on emails first and then provide feedback on the other speaker's previous message, they might say something like "You just said that you finished your breakfast and started to work, good for you!"

In IM exchange, in the case of message overlaps, if a speaker wishes to respond specifically to an earlier message (e.g., "I just finished breakfast and I'm logging in to start work"), IM platforms offer solutions such as comment with emoji function (Figure 4.7) or quoting the message and adding emoji feedback in the reply (Figure 4.8).

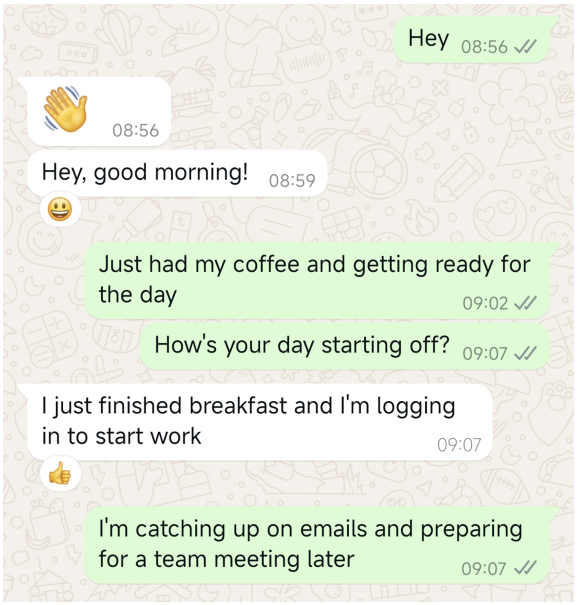



Figure 4.7: React to previous message using comment function

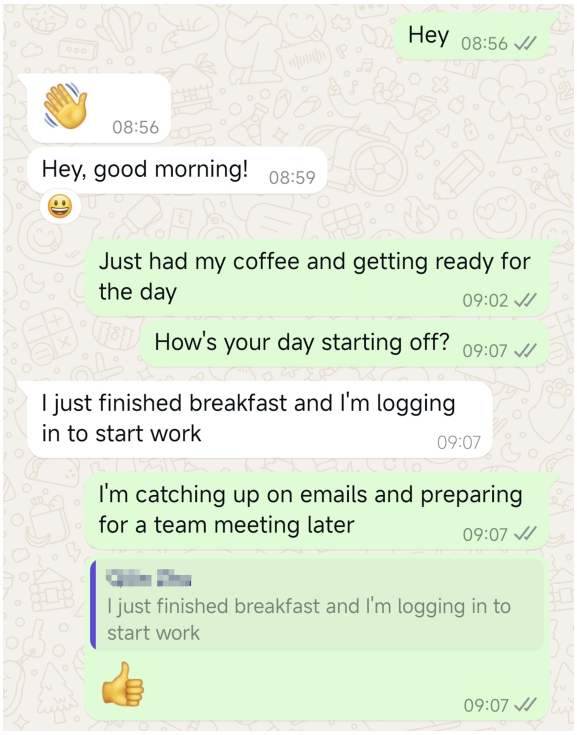



Figure 4.8: React to previous message using quote function

As this example shows, in IM dialogues, although emoji reactions and quoting help resolve overlaps and maintain clear reference to previous utterances, they may also introduce new complexities for discourse coherence. Consider again Figures 4.7 and 4.8. Both show an emoji reaction (👍) responding explicitly to the previous utterance, "I just finished breakfast and I'm logging in to start work," yet the reactions occur after a subsequent message ("I'm catching up on emails") has already been sent. In Figure 4.7, the emoji is directly attached to the earlier message as a reaction, while in Figure 4.8, the emoji feedback appears as a separate, standalone message referencing the original utterance through quoting. In the former scenario, this feedback causes a misalignment: the emoji feedback is temporally the latest communicative act but visually appears attached to an earlier message, disrupting the linearity between spatial and temporal sequence. This misalignment can potentially affect how speakers in IM interpret and continue the conversation.

These features of IM dialogue raise several questions. First, how does the spatial–temporal misalignment caused by "comment with emoji" function affect conversational flow? Second, does how emoji is presented: as a feedback attached to the corner of a previous message (Figure 4.7) versus being sent in a new message (Figure 4.8), influence how the dialogue continues? Third, does the timing of emoji feedback, immediate (Figures 4.4 and 4.5) versus delayed (Figures 4.7 and 4.8) shape the development of discourse?

To address these questions, it is necessary to clarify the notion of *recency*. In this context, recency can be understood in different ways. From a temporal perspective, the 👍 is the most recent contribution. From a spatial perspective, however, in Figure 4.7, the message "I'm catching up on emails and preparing for a team meeting later" appears at the bottom of the screen and is therefore spatially closest to the point where new messages appear.

## 4.2 Recency as Temporal and Spatial Distance

Givón (1983) introduces the notion of *referential distance (RD)*, defined as the number of clauses (or elapsed time) from the last occurrence in the preceding discourse. In this sense, recency can be measured by how much time or how many intervening units separate the current point of reference from the most recent mention of an entity. This temporal view of recency presupposes a model of discourse in which utterances are produced in time sequence and interpreted in order. It fits well with spoken interaction, where discourse is generally processed sequentially and synchronously.

Digital communication environments, however, complicate this picture. Messages may be delayed due to typing pauses or network lag; they may be edited or deleted after sending; and participants may scroll back to read, quote, or respond to earlier contributions. In this environment, the assumption that production order and interpretation order align no longer holds. Instead, interlocutors encounter utterances in a visual layout that preserves past messages and allows reactivation of prior turns at any time. The temporal succession of contributions is therefore only one dimension of recency.

### 4.2.1 Temporal Distance

To measure temporal distance (TD), it is necessary to first specify what counts as the *temporal order* of utterances. Two notions of temporal order can be distinguished.

> The **temporal order of production** is the order in which utterances are *produced* by speakers. In IM, it is determined by the moment an utterance is sent to the system. For multi-

utterance messages, [2] the internal order follows the textual sequence of utterances within that message.

The **temporal order of perception** is the order in which utterances are *perceived* by an addressee. In IM, this is the sequence determined by the moment an utterance is first read by the receiver on screen. [3]

In synchronous spoken interaction, production order and perception order normally align, since speech is produced and perceived incrementally. In IM, the two orders often align when recipients read messages as they arrive and no edits or withdrawals occur. They may also diverge, for example when network delays alter the order of delivery, or when the receiver read the message later.

**Temporal distance** is defined as the number of intervening utterances between the introduction of a referent and its subsequent mention, calculated according to the temporal order of perception. A referent is considered *temporally close* when it is mentioned in the most recent utterance, and *temporally distant* when several intervening utterances separate its introduction from the current point of reference.

According to this definition, in Example 26, utterance (26-c) follows (26-b) in time, which means the referent cat is *temporally closer* than the referent dog.

[2] In IM, the platform introduces the *message* (or chat bubble) as a technical feature: each message is a block of text submitted in one action of pressing "send." A message may contain one utterance, several utterances, or only part of one. I therefore use *message* to describe the structural unit of IM, while using *utterance* as a more general term for any produced segment.

[3] In practice, some users may open a chat with many unread messages and begin by reading the message at the bottom before scrolling upward. Such strategies are excluded here, because they represent individual reading habits.

(26) a. A: Hey, how is everything going?
b. B: Selina just washed the dog
c. B: She also fed the cat
d. B: Now she is petting it

### 4.2.2 Spatial Distance

Spatial distance (SD) refers to the visual separation between the mention of a referent and its subsequent reference in the interface of the discourse. In hand-written letters and printed books, as well as screen-based environments such as chat windows, comment threads, or digital documents, discourse entities are not only processed sequentially but are also navigated spatially. To measure SD, it is necessary to first specify what counts as the spatial order of utterances.

> The **spatial order of utterances** is the order in which utterances appear on the page or screen, determined by the visual arrangement of text.

It is important to note that the spatial order of utterances in IM is not purely a matter of vertical distance. For languages such as English and German, this corresponds to the linear progression from left to right and from top to bottom on the page or screen.

> **Spatial distance** is defined as the number of intervening utterances between the introduction of a referent and its subsequent mention, measured according to the spatial order of utterances. A referent is considered *spatially close* when it appears in the most recent position of the interface,[4] and *spatially distant*

[4] In applications such as WhatsApp or Signal, the most recent position corresponds to the bottom of the chat window, i.e., the point at which new messages are entered.

when several intervening utterances separate it from the current point of reference.

SD in IM is shaped by the presence of multiple layers in the interface. The main conversational thread occupies one layer, while interactive features such as emoji reactions or message edits are displayed in additional layers, visually attached to earlier messages. Both message placement and interface layering contribute to how close a referent appears at the time of reference, making spatial distance an interface-dependent phenomenon.

Consider again Figure 4.9 from Chapter 2, reproduced here for convenience. In WhatsApp, new utterances appear at the bottom of the screen, meaning that the referent *the cat* is not only temporally close but also spatially close compared to *the dog*. This example illustrates a case in which TD and SD align, ensuring the prominence of the same referent along both dimensions.


Figure 4.9: An example of IM dialogue

## 4.3 Aligned and Misaligned Distance

In different modes of communication, temporal and spatial order relate to one another in different ways. In a face-to-face conversation, discourse unfolds purely in time, and production order aligns with perception order. In written texts, such as novels or emails, the order of production may diverge from perception, since authors may draft chapters out of sequence,

but once the text is finalized, the spatial order on the page determines the perception order for readers.

However, in digital communication environments, such as IM, this alignment may break down. The layout of messages on screen introduces a spatial dimension that may diverge from temporal progression. This divergence complicates recency-based reference resolution and motivates the need to examine how interlocutors respond to misaligned cues.

### 4.3.1 Aligned Distance in Communication

In *face-to-face communication*, discourse unfolds entirely in time. As such, SD is not a meaningful concept in this modality. There is no visual layout of utterances that can be navigated or referred back to. Spoken utterances cannot be recalled or deleted, they are bound to the moment of their production. As the saying goes, "A word once spoken cannot be recalled."[5] Utterances cannot be visually revisited, scrolled through, or reordered. TD is thus the only relevant factor in determining recency.

In *written discourse*, such as novels, essays, or textbooks, SD and TD typically align. Readers process texts in a linear and directional manner, for instance from left to right and top to bottom in English. Thus, the most temporally recent utterance is also spatially closest to the current point of reading, typically found near the bottom of the page or the end of a paragraph. For example:

(27) a. Why do birds fly south in winter?
b. Because it's faster than walking.

In this case, utterance (27-b) is both temporally recent (just introduced) and spatially recent (on the immediately preceding line). Even in complex

[5] Horaz, *Ars Poetica. Die Dichtkunst*, ed. and trans. Eckart Schäfer (Stuttgart: Philipp Reclam jun., 2008), v. 390: *Nescit vox missa reverti.* Schäfer's translation: „das Wort, das du von dir gabst, kennt keine Rückkehr."

texts, paragraph structure and formatting typically preserve this spatial-temporal alignment. It is worth mentioning that in these written discourses, all editing, reordering, and deletion happen before the text is presented to the reader. As a result, even if the production order differs from the intended perception order during the writing process, the reader only reads the finalized version in which spatial and temporal orders align. This means that recency-based prominence remains a consistent and reliable cue.

### 4.3.2 Misaligned Distance in Digital Communication

As introduced in Chapter 3, IM platforms offer features such as message quoting, editing and emoji reactions, which allow users to revisit, modify or highlight temporally earlier messages. These tools can blur the linear flow of discourse by reactivating referents that are both spatially distant and temporally distant, making them accessible again.

Consider the IM exchange illustrated in Figures 4.10, 4.11 and 4.12:


Figure 4.10: Before Message Edit

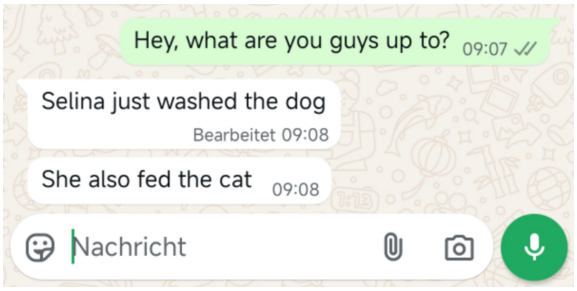



Figure 4.11: After Message Edit


Figure 4.12: Following Message

In Figure 4.10, SD and TD are aligned: the utterance *She also fed the cat* follows an earlier mention of *washed the birdie*. In Figure 4.11, the speaker edits the earlier message and replaces *birdie* with *dog*. This modification introduces the *dog* as a new referent in temporal terms, as it has just been added to the discourse. However, the edited message remains in its original position on the screen, meaning that *the dog* is now temporally close but spatially distant. By contrast, *the cat* remains spatially close, as this utterance is still positioned near the bottom of the interface, where new utterances appear.

In Figure 4.12, the speaker continues with the utterance *Now she is petting it*. The pronoun *it* here is ambiguous: both *the dog* and *the cat* are possible antecedents. They share the same grammatical role and are both direct objects. Thus, the disambiguation of *it* in this case relies on which form of distance is more prominent: spatial or temporal. If SD determines the prominence of the referent, the referent *cat*, being visually closer to the point of reference, should be preferred. If TD determines the prominence of the referent, then *dog*, temporally newly introduced via the edit function, should be more accessible. This example illustrates the ambiguity that arises in IM discourse when temporal and spatial cues diverge, and it highlights the need to investigate how spatial distance and temporal distance compete during reference resolution.

Features such as message editing introduce new referents out of temporal sequence, resulting in cases where referential cues conflict across SD and TD. Such cases are not only a matter of interface design but are also deeply connected to how speakers manage corrections and revisions in dialogue. In IM, corrections can not only follow traditional conversational repair patterns, but also rely on editing tools. This raises questions about how corrections affect coherence in dialogue. The next chapter takes up this question by examining how correction functions as a speech act in dialogue, setting the stage for the experimental hypotheses that follow.

# Chapter 5

# Repairing Words, Rewriting Distance

This chapter explores self-repair as a factor contributing to the misalignment of spatial and temporal distance in IM dialogue, with a focus on its implications for pronoun resolution and discourse coherence. Building on previous discussions of prominence (Chapter 2), the structural features of IM (Chapter 3), and recency effects (Chapter 4), this chapter outlines the types and strategies of self-repair in IM dialogue and discusses their role in referential ambiguity. Self-repair is a universal feature of human communication, enabling speakers to maintain coherence, correct errors, and update shared understandings. However, in the digital environment of IM, different types of self-repair can influence reference tracking and discourse interpretation. This chapter first outlines the general characteristics and types of repair in dialogue, then focus on the difference of self-repair in spoken versus IM dialogue, and finally draws attention to self-repairs using IM message editing function which cause spatial-temporal misalignment and thereby lead to referential ambiguity.

## 5.1 Repair in Dialogue

Repairs are the means by which speakers edit and reformulate conversational turns, and are a characteristic feature of natural dialogue (Colman and Healey, 2011). Repair allows speakers to resolve misunderstandings, correct errors, and maintain coherence across turns. It plays a crucial role in dynamically managing the accessibility and prominence of discourse elements and is also tied to processes of reference, prominence, and alignment, which are core concerns of this dissertation. While repair has been widely studied in spoken communication, its implications for referential process also remain relevant in digital communication, especially IM dialogue, where spatial and temporal distance could misalign.

### 5.1.1 Self and other-initiated Repair

Repair in dialogue can be made either by the speaker themselves or by their conversational partner, resulting in different interactional dynamics. Consider the following self-repair example from Spilker et al. (2000):

(28) on Thursday I cannot · no I can meet ah after one

In this case, the speaker recognizes a potential error in their own utterance and interrupts themselves to make the necessary correction. This is an example of self-initiated self-repair. Here, both the initiation and completion of repair is completed within the same conversational turn by the same speaker.

Repairs can also be initiated by another interlocutor, known as other-initiated repair. For instance, in the following dialogue from Schegloff et al. (1977):

(29) a. Ben: Lissena pigeons.
b. Ellen: Coo-coo::: coo:::
c. Bill: Quail, I think.
d. Ben: Oh yeh?
e. Ben: No that's not quail, that's a pigeon.

Bill initiates the repair by suggesting an alternative interpretation ("Quail"), thus prompting Ben to reconsider his original utterance. In this interaction, the recipient addresses a communicative difficulty and corrects the original utterance.

However, the initiator of a repair does not necessarily have to be the one who completes it. The distinction becomes clear in another example from Schegloff et al. (1977):

(30) a. Ken: 'E likes that waider over there
b. Al: Wait-er?
c. Ken: Waitress, sorry
d. Al: 'At's bedder

Here, Al initiates the repair ("Wait-er?"), but it is Ken, the speaker of the previous utterance, who completes the repair by providing the correction ("Waitress"). This example illustrates other-initiated self-repair, where the listener identifies the communicative problem but the original speaker resolves it.

To systematically capture such variations, Schegloff et al. (1977) developed a typology within the framework of Conversation Analysis (CA), categorizing repairs based on who initiates and who completes them: self-initiated self-repair, other-initiated self-repair, self-initiated other-repair, and other-initiated other-repair.

Signals for other-initiated repair may differ in their degree of specificity. According to Dingemanse et al. (2015), strategies that recipients use to

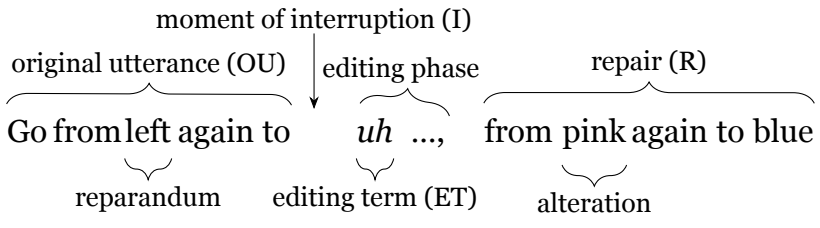
indicate trouble of understanding or hearing range from an open request (e.g., *Huh?*) to signal a general understanding problem to a restricted request (e.g., *Who?*), explicitly seeking clarification. Another approach is the restricted offer, where the recipient suggests a candidate interpretation and seeks confirmation (e.g., *She had a boy?*).

While all forms of repair contribute to the construction and maintenance of mutual understanding, self-repair is especially important. Schegloff et al. (1977) show that conversational participants prefer self-repair. That is, speakers generally attempt to correct their own utterances whenever possible, even if it is the listener who initially signals the need for correction.

### 5.1.2 Self-Repair in Spoken Dialogue

According to Levelt (1983), self-repair proceeds in three phases. First, speakers continuously monitor their own speech, interrupting themselves upon noticing an error. Second, the interruption is frequently characterized by hesitations, pauses, or editing markers. Finally, in the third phase, the speaker provides the correction, thereby completing the repair. The most common structural format of self-repair thus includes three main components: the original utterance, the editing phase, and the repair. Consider the following illustrative example from Levelt (1983):

(31)




In this example, the reparandum (the trouble spot) is *left*, the editing term is the hesitation marker (*uh*), and the alternation is the correction (*pink*). The editing term, such as filled pauses (*uh*, *um*) or expressions like *I mean*

or *well*, serves as a cue for listeners that a correction is forthcoming (Spilker et al., 2000) and signals a shift from the original problematic utterance toward the repaired formulation.

**Types of self-repair** Self-repairs can be categorized based on their structural and functional characteristics. One distinction identified by Levelt (1983) differentiates between *error repairs* and *appropriateness repairs*. Error repairs address inaccuracies in lexical choice (*Go then to the wrong..., to the other side*), syntax (*And black ... from black to right to red*), or phonetic (*A unut, unit from the yellow dot*), reflecting real-time monitoring of speech production (Levelt, 1983). Appropriateness repairs refine utterances to enhance clarity, avoid ambiguity, or ensure that expressions remain contextually appropriate, coherent, and aligned with previously used terms or expressions.

Another categorization differentiates between self-repairs occurring within the *same turn-constructional unit (TCU)* and those happening *after the TCU*. Same-TCU repair involves immediate corrections within the ongoing turn (Schegloff et al., 1977). Example 32, taken from Schegloff (2013), illustrates a common strategy of replacement: at line (32-e), the phrase *by the t[ime]* is replaced with *when*.

(32)
a. Bee: Becuz they're gonna do the operation on the teeuh duct.
f [ f i : r s ] t. Before they c'n do=
b. Ava: [Mm–hm,]
c. Bee: =t [he cata] ract]s.
d. Ava: [Right. ]Yeah,]
e. Bee: .hhh So I don' know I haven:'t yihknow, she wasn' home
→ **by the t–** yihknow ***when*** I lef' fer school tihday.=
f. Ava: =Mm hm,
g. Bee: Tch! .hh So uh I don't kno:w,

Other common repair methods within same-TCU include inserting, deleting, searching (or word searches), parenthesizing, aborting, sequence-jumping, recycling, reformatting and reordering (for a detailed description and examples, see Schegloff 2013).

Besides self-repair in same-TCU, repairs can also happen in later turns or after intervening utterances by other participants. Such delayed repairs commonly occur in the third turn, or third position (Kitzinger, 2012). Schegloff et al. (1977) observes that a speaker may produce an initial utterance (a trouble-source turn), followed by an intervening contribution from the addressee who do not explicitly signal misunderstanding or confusion. Nevertheless, the original speaker may subsequently realize the need for correction and initiate a delayed self-repair. Consider the following example from Kitzinger (2012), demonstrating delayed self-repair occurring in a third turn:

(33)
a. Sar: >uh uh< Did Sean: go ho:me fer- (0.5) break=
b. Lis: =Ye::ah. He went back t' Mi:chigan.=Friday.
c. T1 <'r actually (0.5) Thurs:da::y ni:ght.
d. Sar: T2 Oh- Oh.
e. (0.2)
f. Lis: T3 Afternoo:n.

Here, in Turn 1, Lisa initially indicates the timing as *Friday*, then quickly self-corrects it to *Thursday night* (same-TCU self-repair). In Turn 2, Sarah rsponds minimally (*Oh-Oh*), neither indicating nor acknowledging a problem explicitly. Nevertheless, Lisa provides a further delayed correction in Turn 3, replacing *night* with *afternoon*. This delayed self-repair shows that the speaker continue to monitor their prior discourse across multiple conversational turns. Further analyses and examples of delayed self-repair can be found in Schegloff (1997) and Schegloff (1992).

**Correction as a coherence relation** The phenomenon of (self-)repair has also been studied in theoretical approaches to discourse coherence based on coherence relations. Most coherence relations, such as elaboration or explanation, add to what has already been said and build up shared information. Correction, by contrast, establishes a relation in which a later contribution overrides or replaces part of an earlier one (Asher and Lascarides, 2003). In dialogue, this often occurs when one speaker contradicts another, as in the following example from Asher and Lascarides (2003):

(34) a. A: John distributed the copies.
b. B: No, Sue distributed the copies.

Here, speaker B's utterance corrects speaker A's, and after the correction the referent *Sue* becomes the most recent and therefore the most prominent entity.

The same process applies when speakers repair their own talk. Asher and Lascarides (2003) note that self-repair can be understood as a form of correction:

(35) It was raining on Saturday. Oh no, wait a minute, I think it was Sunday.

In this exapmle, self-repair functions as a coherence relation that the speaker's later contribution corrects their own previous utterance. Classic accounts link referential choice to factors such as recency and accessibility (Givón, 1983; Ariel, 1990; Gundel et al., 1993). A correction resets recency and makes the correcting expression accessible, and thus the referent it introduced becomes the potential candidate for pronominal reference. For instance, in example 35, if the speaker continues with *Later on that day...*, *that day* is resolved to the corrected NP *Sunday*, not *Saturday*, because the

coherence relation explicitly signals that the repaired utterance overrides the earlier referent.

**Self-repair and Temporal recency** Beyond its communicative and corrective functions, self-repair also influences the temporal organization of discourse. As discussed in Chapter 4, a referent is considered temporally close when it is mentioned in the most recent utterance of the perception order, and temporally distant when several intervening utterances separate its introduction from the current point of reference. Temporally closer referents are generally more accessible and are more likely to serve as antecedents for subsequent pronouns and definite expressions (Clark and Sengul, 1979; Arnold, 2025). When a speaker corrects or clarifies a referent through self-repair, that entity is not only updated in terms of informational content but is also reintroduced as temporally close within the conversational flow. In this sense, self-repair can reset the temporal distance of a discourse entity, making it recent in the discourse, regardless of where it first appeared.

Consider the following dialogue:

(36) a. A: How is everything going?
b. B: Selina just washed the dog.
c. B: She also fed the cat.
d. B: I mean she washed the neighbor's dog.
e. B: Now she is petting it.

Although the referent *cat* is mentioned after the original reference to *dog*, the self-repair in utterance (36-d) reactivates and emphasizes the referent *neighbor's dog*, making it temporally close again. This update makes the corrected NP more prominent, so that in the follow-up utterance *Now she is petting it*, the pronoun *it* is preferentially interpreted as referring to *the*

*neighbor's dog* rather than *the cat*. Thus, self-repair can influence discourse coherence and referential dynamics by reestablishing previously mentioned referents as accessible.

## 5.2 Self-Repair in IM Dialogue

The discussion now turns specifically to the context of IM. While the underlying motivations for self-repair remain consistent across modalities, IM introduces new constraints and affordances. The absence of immediate auditory feedback, the availability of textual editing, and the persistent visibility of previously sent messages influence how speakers repair and how recipients interpret them.

### 5.2.1 Types of self-repair

Repair in IM differs from spoken dialogue due to the text-based nature of digital interaction. According to Meredith and Stokoe (2014), repairs in digital communication broadly fall into two categories: *message construction repair* and *visible repair*.

*Message construction repairs* occur during the formulation of the message itself, and thus remain invisible to recipients. Consider the following example adapted from Meredith and Stokoe (2014):

(37)
a. 03.19 Isla: right sorry buti'm off to bed
b. I*: ✍ absolutely knackerd ~~d~~ ed! ✍
c. 00.05 Isla: absolutely knackered!
d. I*: ✍ speak t ~~speak t~~ let me know when (.) ~~when~~ (.) ~~let me know~~ you training tuesdA ~~dA~~ a a day? ✍
e. 00.14 Isla: you training tuesday?

In this example, the shaded turns are those visible to the recipient in the chat interface, while the intervening edits remain hidden. Participants frequently edit or delete segments of their messages before sending, with deletions represented here by strikethrough text. Using screen-capture methods, Meredith and Stokoe (2014) demonstrate that these repairs share functional similarities with repairs in spoken interactions, including the correction of lexical errors, adjustments to stance, and refinements in action formation.

By contrast, *visible repairs* are corrections observable by all participants. These occur after the message has been sent, as illustrated in the following example from Meredith and Stokoe (2014):

(38)
a. 0.56 Callum: 😐 lol im not going to argue with you tonight
secret diaries of a call girl is coming on lol ha-
ha tell her i say hiya and well done for a ew pb xxx
b. I*: ✍ wait wait wait ✍
c. 0.15 Isla: wait wait wait
d. I*: ✍ i ~~i~~ ✍
e. 0.06 Callum: new

In this chat, Callum notices the error (*ew* instead of *new*) after Isla has already posted another message. This scenario is similar to a third-turn repair in spoken conversation, as noted by Meredith and Stokoe (2014). Interestingly, while visible repairs in IM sometimes mirror spoken dialogue sequences, they are most commonly targeted at text-specific issues, such as spelling or typing errors.

In this dissertation, the focus is placed explicitly on visible repairs, given that they appear directly on the conversational interface, potentially influencing the spatial and temporal distance and reshaping conversational flow and reference tracking.

### 5.2.2 Repair Strategies in IM

In the absence of prosodic cues typical of spoken conversation, both users and online platforms have developed a range of textual and interactional strategies for visible self-repair. These strategies vary in explicitness and formality. While some strategies mark the correction in chat interface, others rely on platform-specific tools such as message editing, which allow speakers to modify earlier messages without introducing new turns or visibly altering the structure of the dialogue.

A common repair strategy in IM involves posting a follow-up message explicitly correcting a preceding utterance. These follow-up repair can be done without (as in example 38) or with markers. An example of the marked variety, known as *-repair (Collister, 2011), is given below:

(39) a. 11/27 21:14:52.750 [Party] Aniko: when i run ot
b. 11/27 21:14:54.765 [Party] Aniko: out*

Here, the asterisk explicitly signals a correction of the typographical error (”ot” to ”out”). This strategy has become widely understood across online environments, especially within informal digital communities such as online gaming or social media platforms.

Besides *-repair, sometimes speaker re-sends the entire corrected message. This approach relies on the recipient recognizing the corrected version and understanding that it replaces the original, as shown in Figure 5.1. The recipient has to recognize that the corrected version (*I'll mail the documents on Thursday*) is intended to replace the earlier mention (*Tuesday*).

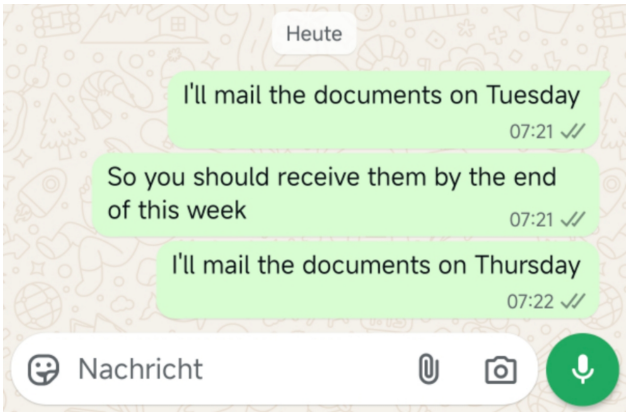



Figure 5.1: IM repair: stand-alone utterances

These follow-up repairs often take the form of *stand-alone messages*.[1] Unlike in spoken discourse, where such repairs would be integrated prosodically or syntactically into the same turn, IM self-repair externalizes the correction into a new message. This separation can affect how clearly the repair is linked to the original utterance, especially in fast-moving or multi-threaded exchanges.

An alternative repair strategy involves *quoting or referencing* the original message and appending a correction, especially in group chats or asynchronous conversations where turn adjacency is disrupted, as in figure 5.2. This format makes the repair explicit and allows the speaker to anchor the correction to a previous utterance, which can be particularly helpful when the conversation has moved on.

More recently, some IM platforms have introduced *message editing* function that allows users to modify sent messages after they appear (detailed in Chapter 3.2). Such editing features typically mark the message with an "edited" tag but retain the original spatial position within the dialogue structure (Figure 5.3).

[1] In this dissertation, I use *turn* to describe a speaker's contribution before another participant takes the floor. In IM, turns are technically realized as *messages*, that is, blocks of text submitted by pressing "send."

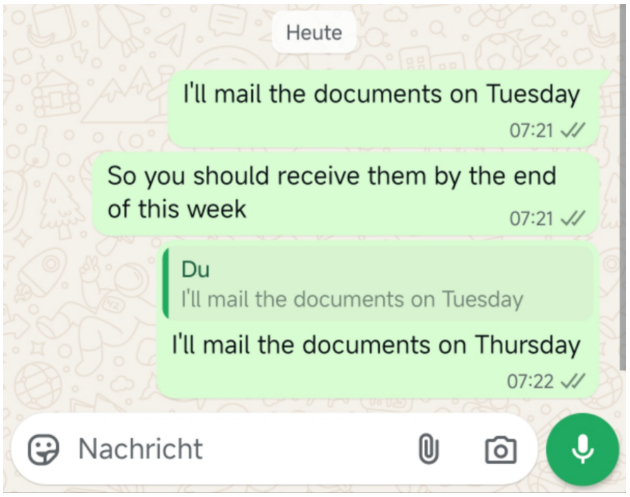



Figure 5.2: IM repair: Quoting

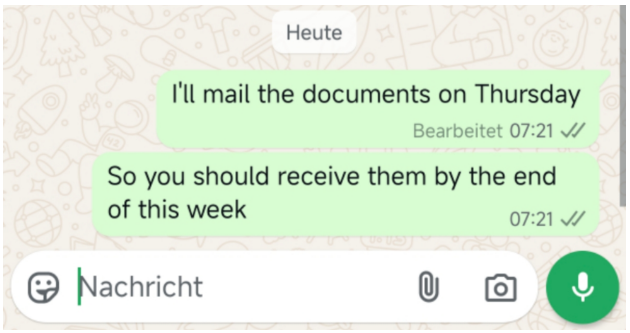



Figure 5.3: IM repair: Message editing

Figures 5.1 to 5.3 illustrate how the explicitness of repair strategies varies depending on their structural presentation. Quoted repairs explicitly anchor the correction to the original utterance, making the intended revision transparent to the recipient. In contrast, stand-alone messages rely primarily on temporal recency, assuming that recipients will infer that the temporally recent repair overrides the original utterance. This strategy requires greater inferential effort from the recipient, especially when multiple intervening turns are present. Repairs performed through message editing, which are specific to IM platforms, depend on the recipient paying attention to the edit. Because edited messages appear in their

original spatial position while their content is temporally updated, this strategy can introduce misalignment between spatial and temporal distance, which is of particular relevance to this dissertation as it directly concerns how such misalignment affects reference resolution. Thus, beyond serving as a mechanism for error correction, the question arises whether self-repair in IM also reshapes discourse coherence by influencing referential prominence. This question will be examined in the empirical studies that follow this chapter.

## 5.3 Self-Repair and Referential Ambiguity

As discussed in Chapter 4, IM dialogue may involve misalignment of spatial and temporal distance: the visual arrangement of messages on screen (SD) does not always reflect the chronological progression of interaction (TD). This misalignment becomes especially relevant in the context of self-repair, where the act of correcting or updating an earlier utterance introduces new layers of complexity for reference resolution.

For instance, as shown in Figure 5.3, a message originally stating *I'll mail the documents on Tuesday* can be edited to *I'll mail the documents on Thursday* without changing its visual position on the screen. While the correction is temporally close, its spatial position remains distant. Such edits may lead to a misalignment between what is displayed, what was intended, and what was interpreted, thus raising the risk of a mismatched interpretation of the ongoing conversation.

An example of how misaligned spatial and temporal distance could influence the pronoun resolution is provided in Figure 5.4, previously introduced in Chapter 4. In this chat, the pronoun *it* can ambiguously refer either to *the cat*, which is spatially close (appearing directly above the current message), or to *the dog*, which is introduced through a temporally recent self-repair. The self-repair thus creates a situation in which spatial


Figure 5.4: Misaligned spatial and temporal recency in IM dialogue

distance and temporal distance point to different referents, raising the question of whether such misalignment complicates pronoun resolution. More broadly, the question is whether self-repair in IM, beyond functioning as a corrective device, also reshapes the prominence structure of discourse. In particular, does modifying a referent after a delay disrupt the correlation between recency and prominence, especially when spatial and temporal distance point to different referents?

As discussed extensively in Chapter 2 and 4, recency is a factor influencing reference resolution in discourse: in spoken discourse, entities mentioned more recently tend to be more accessible and are more likely to be interpreted as antecedents for pronouns (Clark and Sengul, 1979; Arnold, 2025). However, self-repair through editing function in IM can disrupt temporal recency as a prominence-lending feature by introducing new referent in spatially distant utterance through correction. When a previously introduced entity is corrected or clarified, it may gain renewed prominence, even though it does not appear in a spatially closer message. This results in competing cues for prominence: a spatially close but temporally distant referent may compete with a temporally close but spatially distant referent introduced through repair.

These competing cues form the basis for the experimental studies presented in the next chapter. By systematically manipulating the type and structure of self-repairs in IM-style dialogues, the studies aim to investigate how readers prioritize spatial and temporal recency as prominence-lending factors when resolving pronouns in IM dialogues.

# Chapter 6

# Studies on Repair and Recency Misalignment in Instant Messaging Dialogue

As discussed in Chapter 2, prominence refers to the relative accessibility of discourse entities within a developing discourse (Von Heusinger and Schumacher, 2019). Various factors, including grammatical role, discourse structure, and recency, have been identified as factors influencing prominence (Keenan and Comrie, 1977; Gundel et al., 1993; Lambrecht, 1994; Prince, 1978, 1981; Thompson, 1987; Krifka, 2008).

Chapter 3 extended this discussion to the environment of IM dialogue. IM dialogue presents challenges to discourse organization, like features such as emoji reactions, message modification, and quoting, which can lead to ambiguity in referential expressions and change the visual layout of the conversation on the screen. As a result, spatial and temporal distances of the utterance can be misaligned.

This misalignment raises the question of how recency should be defined in IM, a topic discussed in Chapter 4, where it is defined along two dimensions: temporal distance and spatial distance. TD is measured by the number of intervening utterances between a referent's introduction and its

subsequent mention, whereas SD is measured by the number of intervening utterances in the visual order of the chat interface. According to this definition, a referent is considered temporally close when it is introduced in the most recent utterance in the sequence of perception, and spatially close when it is visually close to the position where new messages are produced. These two types of distance introduce uncertainty about which of them serves as the primary factor that influences pronoun resolution.

Various factors can influence TD and SD in IM. One such factor is repair, introduced in Chapter 5. Repair, especially self-repair, in IM interaction can be divided into two types: message construction repair and visible repair (Meredith and Stokoe, 2014). Visible repair may take the form of a new message, a quotation of a previous message, or the use of the editing function. From the perspective of discourse organization, self-repair through the editing function is particularly interesting, as it can introduce a new entity that is temporally close but still visually tied to earlier discourse units.

To examine referent prominence under conditions of spatial and temporal misalignment, this chapter presents a two-part empirical study on pronoun resolution in IM dialogue. Section 6.1 begins with an example illustrating the misalignment caused by self-repair, which serves to motivate the research question and continues with an introduction of the research question and hypotheses. Section 6.2 reports on a pretest conducted to develop controlled stimulus materials.

Based on the stimulus set, two experiments are carried out. Section 6.3 presents Experiment I, which investigates whether temporal or spatial distance has a stronger influence on pronoun resolution in IM dialogue. The results suggest that SD has a stronger effect when the two cues conflict. However, because this misalignment is introduced through message editing as a form of self-repair, it is unclear whether the effect reflects SD or the repair act. To address this, Experiment II (Section 6.4) isolates the effect of

repair by presenting it as a stand-alone message with the repaired utterance quoted. This design tests whether repairs function like other contributions with respect to pronoun resolution in IM discourse.

## 6.1 Research Question and Hypothesis

This is one of the main research questions for this dissertation:

> RQ I: *Which has a stronger influence on pronoun resolution in IM dialogue: spatial distance or temporal distance between a referent's introduction and its subsequent mention?*

Let me illustrate this with an example. Consider the following dialogue exchange in Figure 6.1, an English translation is provided in Example 40.


Figure 6.1: Example: Initial message sequence

(40) a. *Lea: Hello, how are you?*
b. *Freya: Greta sliced the ham.*
c. *Freya: After that, she thawed the bacon.*

In this short exchange, Lea initiates the conversation, and Freya responds by narrating what Greta has been doing: first slicing the ham and then thawing the bacon. After sending these two messages, Freya realizes that the first message contains an error: Greta had actually sliced the roast, not the ham. She corrects this by editing her earlier message using the message editing function available in many IM applications. This update is shown in Figure 6.2.

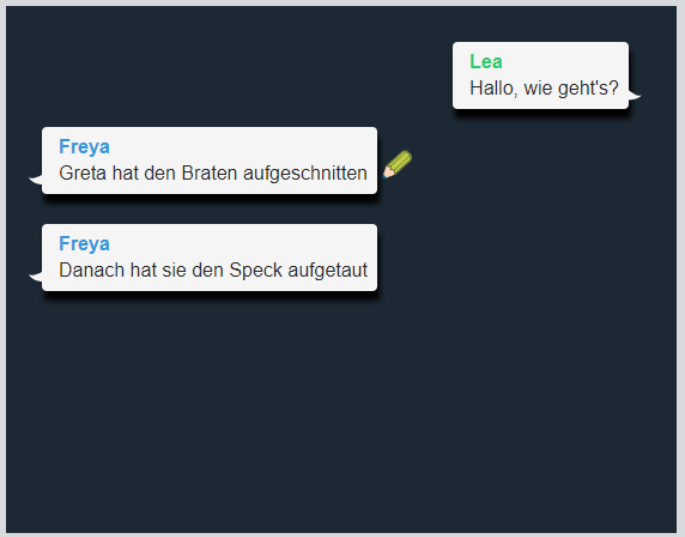



Figure 6.2: Example: Repair via message editing

The repair introduces a misalignment between SD and TD. The most recently produced utterance is now the repair (*Greta sliced the roast*), which is temporally close but spatially distant. The utterance that appears last in the interface (*After that, she thawed the bacon*) was in fact produced earlier, making it spatially close but temporally distant. This raises an empirical question: in such cases of misalignment, which referent, the roast or the bacon, is perceived as more prominent?

Now consider the continuation in Figure 6.3, where Freya continues the story with a nonce verb: How would the pronoun *ihn* ("it") be resolved? In this specific case, if the resolution of the ambiguous pronoun *ihn* relies

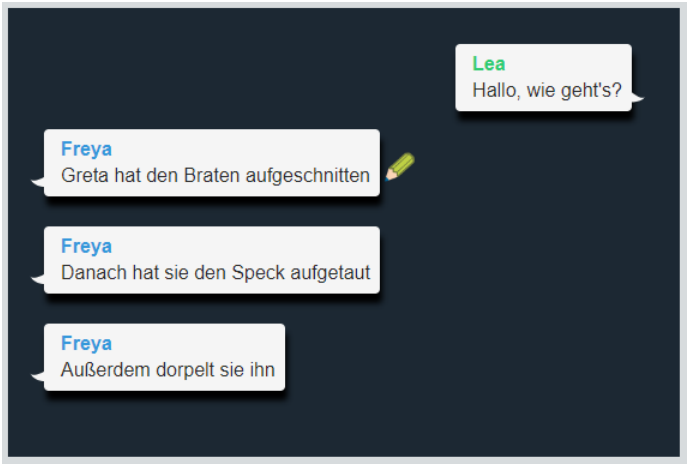



Figure 6.3: Example: Continuation of the dialogue

on SD, it would more likely refer to *den Speck* (“the bacon”). However, if it relies on TD, the temporally closer referent *den Braten* (“the roast”) is more likely to be chosen. Based on this, two hypotheses are proposed:

**Hypotheses for RQ I:**

- **H1**: When temporal distance and spatial distance are aligned, the referent of the closer nominal expression is more likely to be referred to by a subsequent pronoun than the referent of the more distant nominal expression.
- **H2**: When temporal and spatial distance are misaligned, pronoun resolution is guided by temporal distance, such that a subsequent pronoun is more likely to refer to the referent of the temporally closer nominal expression.

## 6.2 Pretest: Identifying Nonce Verbs

### 6.2.1 Purpose

Real verbs are often tied to world knowledge that shapes expectations about possible arguments. For instance, in a discussion of pets, in a sentence such as *Now I am washing it*, people are more likely to interpret the pronoun as referring to *the dog* rather than a less typical one like *the snake*, based on prior knowledge of what kinds of pets are usually washed. Such verb-driven expectations can introduce interpretive biases that confound the effect of spatial or temporal distance. Therefore, before testing these factors, I conducted a pretest to establish a reliable set of nonce verbs for use. This step was taken to minimize semantic biases in participants' interpretations of referents.

This approach follows previous work in psycholinguistics and discourse processing where nonce verbs have been used to control for semantic influence in referential choice and argument interpretation (e.g. Naigles, 1990; Fisher et al., 1994; Kako, 2006). In the current study, candidate nonce verbs were first created and then evaluated for plausibility and interpretability. The nonce verbs selected from this pretest were then used in the main experiments.

### 6.2.2 Method

#### Materials

Items consisted of 105 verbs: 75 nonce verbs and 30 real German verbs. The real verbs were evenly split into two categories: 15 high-frequency verbs, drawn from frequency rankings by Tschirner and Jones (2006), and 15 low-frequency verbs identified using criteria from the Duden corpus, where infrequent verbs occur, on average, just over once per million word forms (Dudenredaktion, 2023).

The nonce verbs came from three sources: 45 were provided by Timo Buchholz (personal communication), [1] created for experimental purposes; 10 were generated using ChatGPT; [2] 20 were constructed by the author. All verbs were presented in the third-person singular present tense with the subject pronoun *er* "he", each category is illustrated in 41.

(41) a. real verb, frequent *er glaubt* "he believes"
b. real verb, infrequent *er schilt* "he scolds"
c. nonce verb *er truntelt* "he 'truntelts'"

### Participants

55 speakers of German (25 female, 25 male; mean age = 36.34 years, range = 21–71) participated in the pretest. Participants were recruited via Prolific[3] and were compensated £4 for their participation. All participants gave informed consent, and Prolific's pre-screening system was used to ensure that they had no history of language or cognitive disabilities.

### Procedure

The pretest was conducted online using Ibex Farm (Drummond, 2023), a browser-based platform for psycholinguistic experiments. Each participant was presented with all 105 verbs in a randomized order.

For each item, participants were first asked whether the verb conveyed any meaning. If they answered *yes*, an open text field appeared, asking them to describe what the verb meant briefly. An illustration of the procedure is provided in Figure 6.4. The first two trials were always real verbs, presented in the same format as the rest of the experiment, to familiarize participants with the task.

[1] I thank Timo Buchholz for generously sharing a set of nonce verbs for use in this study.
[2] ChatGPT (OpenAI, GPT-4) was used to generate morphologically plausible nonce verbs, which were then manually reviewed and screened for suitable candidates.
[3] Prolific is an online participant recruitment platform (https://www.prolific.co)

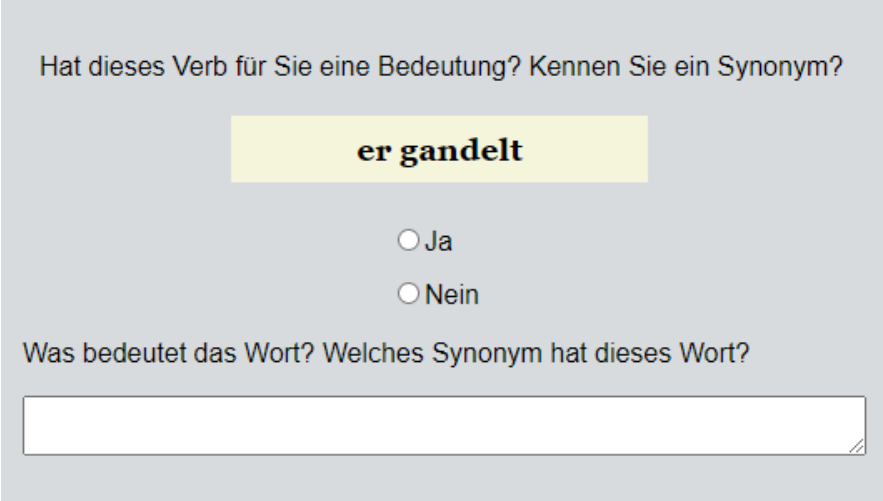



Figure 6.4: Pretest: experimental design

### 6.2.3 Outcome

Of the 75 nonce verbs tested, 54 were unanimously judged by all participants as conveying no identifiable meaning. These verbs were considered semantically neutral and therefore suitable for the main experiments. From this subset, 30 were randomly selected to create critical items.

Another 11 nonce verbs were judged meaningful by exactly one participant each. These were excluded from the critical set, but 6 were randomly chosen for use in filler items, alongside the 24 remaining verbs from the original set of 54 that were not used as critical stimuli[4].

[4]The complete list of the 54 nonce verbs unanimously judged to be non-meaningful is provided in Appendix A.

## 6.3 Experiment I: Pronoun Resolution after Edit-Based Repair

### 6.3.1 Aim and Hypotheses

Experiment I was designed to empirically test the hypotheses introduced in Section 6.1, which ask whether spatial or temporal distance has a stronger influence on pronoun resolution in IM dialogue. Specifically, it examines the case where repair in IM creates a misalignment between the two distances. For clarity, the hypotheses are restated here:

- **H1**: When temporal distance and spatial distance are aligned, the referent of the closer nominal expression is more likely to be referred to by a subsequent pronoun than the referent of the more distant nominal expression.
- **H2**: When temporal and spatial distance are misaligned, pronoun resolution is guided by temporal distance, such that a subsequent pronoun is more likely to refer to the referent of the temporally closer nominal expression.

### 6.3.2 Method

#### Materials

Thirty IM-style chat dialogues were constructed as critical items for this experiment. Each dialogue began with an opening question from speaker A (e.g., Lea), followed by a sequence of utterances[5] from speaker B (e.g., Freya). An example item is given in (42), with an English translation in (43).

[5] In the following descriptions, Utterance is abbreviated as U (e.g., Context-U1, Context-U2, Repair-U).

(42) a. Lea: Hallo, wie geht's? (Opening question)
b. Freya: Greta hat den Schinken aufgeschnitten. (Context U1)
c. Freya: Danach hat sie den Speck aufgetaut. (Context U2)
d. Freya: Greta hat den Braten aufgeschnitten. (Repair-U)
e. Freya: Außerdem dorpelt sie ihn. (Target-U)
f. *FRAGE:* Was dorpelt Greta?
A: Schinken B: Speck C: Braten

(43) a. *Lea: Hello, how are you?* *(Opening question)*
b. *Freya: Greta sliced the ham.* *(Context U1)*
c. *Freya: After that, she thawed the bacon.* *(Context U2)*
d. *Freya: Greta sliced the roast.* *(Repair-U)*
e. *Freya: Besides, she is 'dorpeling' it.* *(Target-U)*
f. *QUESTION: What is Greta 'dorpeling'?*
*A: Ham B: Bacon C: Roast*

The opening questions (e.g., *Hallo, wie geht's?* "Hi, how are you?"; *Alles gut?* "All good?") served as conversation starters, drawn from the same set of ten informal German greetings used in an earlier experiment (Jasinskaja et al., 2024). To ensure distributional balance, each greeting was used in three of the 30 items.

Each response sequence from speaker B included four utterances:

- **Context-U1** and **Context-U2** were perfect-tense utterances holding a coordinating discourse relation. The subject was held constant as a human referent, while each utterance introduced a different masculine accusative object.
- **Repair-U** was a correction of Context-U1, replacing it on the screen through the simulated message-editing function. The object in Context-U1 was repaired by a new masculine accusative object that

was contextually plausible within the dialogue.

- **Target-U** was a present-tense utterance, which contained a subject pronoun matching the gender of the subject in the context utterances, an object pronoun *ihn* ("it") serving as the critical referring expression, and a nonce verb selected from the pretest.

Following Target-U, participants completed a multiple-choice question identifying the referent of the object pronoun in Target-U.

The gender of speakers and referent subjects was counterbalanced across items. In half of the stimuli, the chat interface was mirrored, such that speaker A's messages appeared on the left and speaker B's on the right, and vice versa in the other half. This controlled for possible effects of speaker's position.[6]

Thirty filler items were also constructed with a similar structure but contained other types of modification, such as changes in subject or modifier. These served to reduce predictability and mask the experimental manipulation.

### Participants

Data were collected from 41 participants. One dataset was excluded because the participant reported not being a native speaker of German, and another was removed due to a failed concentration check. This resulted in a final sample of 39 native German speakers (21 male, 17 female, 1 diverse; mean age = 37.90 years; age range = 23–71 years). All participants were recruited via Prolific and received £6.83 for their participation. All participants gave informed consent, and Prolific's pre-screening system was used to ensure that they had no history of language or cognitive disabilities.

[6]A complete list of the stimuli used in this experiment is provided in Appendix B.

### Design

The experiment used a within-subjects design with three conditions, manipulating the alignment between spatial and temporal distance through simulated message editing. Each participant saw every item in only one of the following three conditions, with items counterbalanced across participants.

- **Late Repair**: *Repair-U* was produced after Context-U2, making it the most *temporally close* utterance in the sequence. However, because it replaced Context-U1 via simulated message editing, it appeared spatially before Context-U2, making it *spatially more distant than* Context-U2.
- **Early Repair**: *Repair-U* was produced immediately after Context-U1 and before Context-U2, making it *temporally more distant* than Context-U2. Since it replaced Context-U1 on the interface, it appeared spatially before Context-U2, thus *spatially more distant* as well.
- **No Repair**: Context-U1 remained unchanged, and Context-U2 directly followed it. This condition served as a baseline for natural distance effects, with the referent introduced in Context-U2 expected to be more prominent.

### Procedure

The experiment was conducted online using Ibex Farm (Drummond, 2023). After giving informed consent, participants received instructions and completed three practice trials to become familiar with the interface. They then proceeded to the main experiment.

Each trial displayed a short IM-style dialogue between two speakers in a chat-like interface to mimic natural IM conversations. The dialogue was presented dynamically to simulate the real-time flow of IM exchanges. Participants read the dialogue and then answered a forced-choice question.

Each trial began with an opening message from Speaker A, followed by a sequence of messages from Speaker B. After a brief delay corresponding to the estimated reading time of the opening message, an animated typing indicator (three looping dots) appeared, signaling that Speaker B was constructing a message. Each utterance was then displayed after a delay consistent with its length and the assumed typing speed. This procedure was used for all utterances. Reading and typing times were calculated from average rates: 52 words per minute (WPM) for typing (Dhakal et al., 2018) and 200 WPM for reading (Siegenthaler et al., 2011).

In the **Late Repair** condition, Speaker A (e.g., Lea) initiated the conversation with a greeting (Figure 6.5a). After a reading delay based on the message length, Speaker B (e.g., Freya) was shown to begin composing Context-U1, indicated by three looping dots (Figure 6.5b). Following a typing delay calculated from U1's length, the typing indicator disappeared, and Context-U1 appeared on screen (Figure 6.5c).

The same process applies to Context-U2: the dots reappeared (Figure 6.6a) and Context-U2 was displayed after the appropriate delay (Figure 6.6b). Then, after a short pause equivalent to Context-U1's reading time, the typing indicator appeared again (Figure 6.6c). This represented the composition of Repair-U. The duration of this typing phase was determined by the estimated time required to delete the original noun (e.g., *Schinken*) and type the new one (e.g., *Braten*).

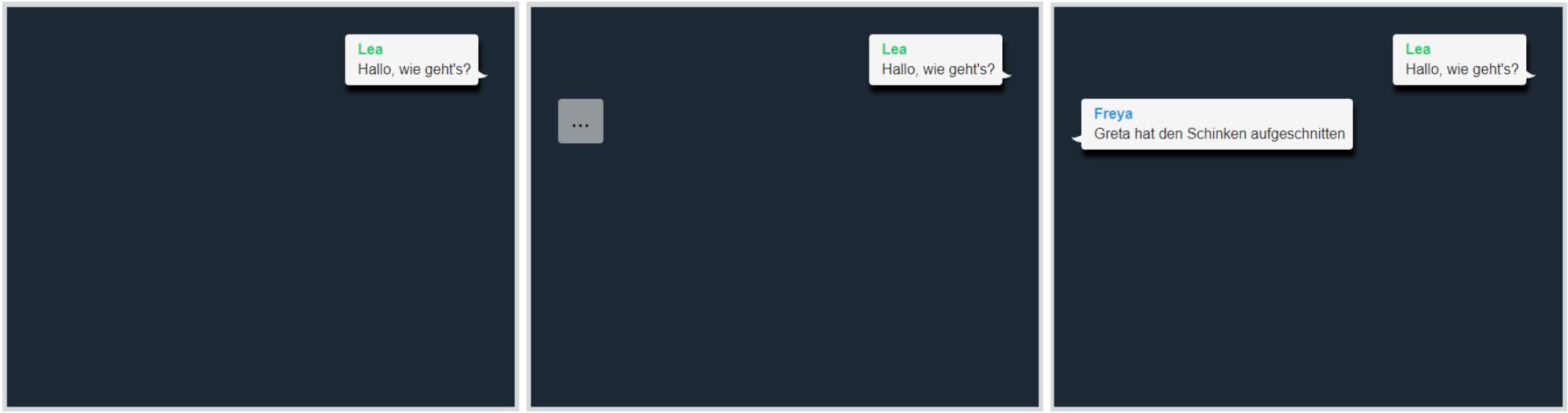
Repair-U then replaced Context-U1 in the interface and was marked with a pencil symbol (Figure 6.7a). In this condition, compared with Context-U2, *Repair-U was spatially distant but temporally recent*.

After another instance of the typing indicator (Figure 6.7b), Target-U appeared (Figure 6.7c). This critical utterance included a nonce verb and an ambiguous object pronoun, *ihn*, whose referent could be any of the three previously introduced referents.

In the **Early Repair** condition, the dialogue presentation began in the same way as in the LR condition: Speaker A initiated the conversation with a greeting (Figure 6.8a), followed by the appearance of a typing indicator (Figure 6.8b) and Context-U1 (Figure 6.8c). After Context-U1 was displayed, the typing indicator appeared again, indicating that Speaker B was preparing a new message (Figure 6.9a). Repair-U then replaced Context-U1 via simulated message editing and was marked with a pencil emoji (Figure 6.9b). After Repair-U was presented, the dialogue continued with the display of Context-U2 (Figure 6.9c) and 6.10a, followed by Target-U in the same format as the other conditions (Figure 6.10b and 6.10c). In this condition, Repair-U occurred before Context-U2, meaning that *Repair-U was both spatially and temporally more distant* than Context-U2.

In the **No Repair** condition, no message was repaired. The dialogue unfolded sequentially: Speaker A sent an opening message, followed by Speaker B's Context-U1, Context-U2, and then Target-U (Figure 6.11). In the forced-choice task, participants always saw three possible referents, as in the other conditions. However, in this condition, only two of them (the objects from Context-U1 and Context-U2) had actually been introduced in the dialogue. This condition served as a baseline, with *spatial and temporal order fully aligned* across utterances.

In all conditions, once the dialogue was complete, participants had four seconds to reread it before being redirected to a new screen with a multiple-choice question (Figure 6.12). This question asked them to identify the referent of the pronoun in Target-U. In filler items, the question instead targeted the subject or a modifier, to reduce predictability and distract participants from the experimental manipulation.




(a) Late Repair: Opening message

(b) Late Repair: Typing indicator before Context-U1

(c) Late Repair: Context-U1 appears

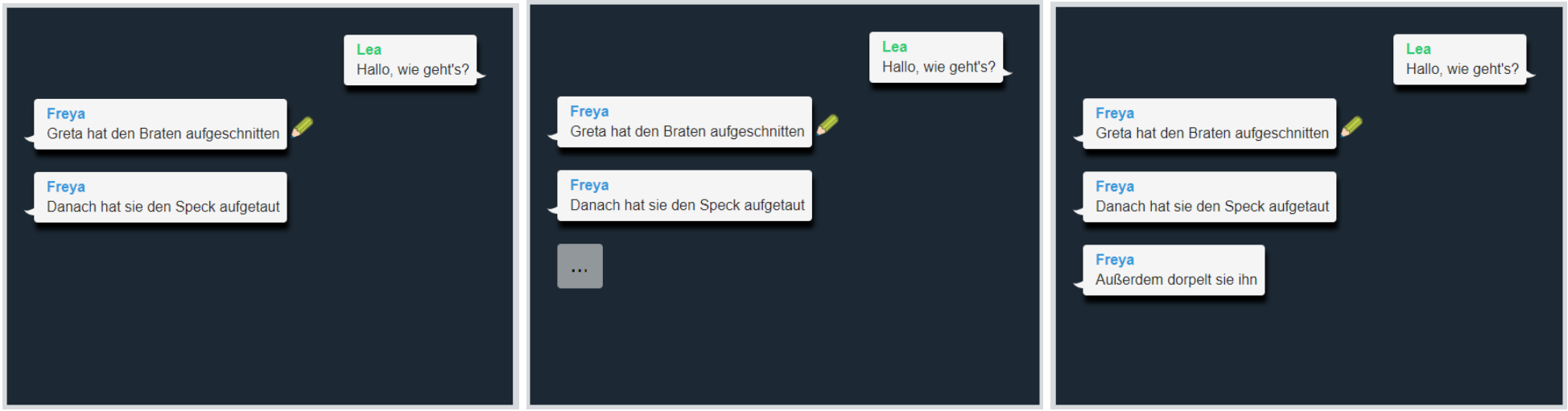
Figure 6.5: Exp I: Example screenshots from the Late Repair condition (1-3)


(a) Late Repair: Typing indicator before Context-U2

(b) Late Repair: Context-U2 appears

(c) Late Repair: Typing indicator before Repair-U

Figure 6.6: Exp I: Example screenshots from the Late Repair condition (4-6)

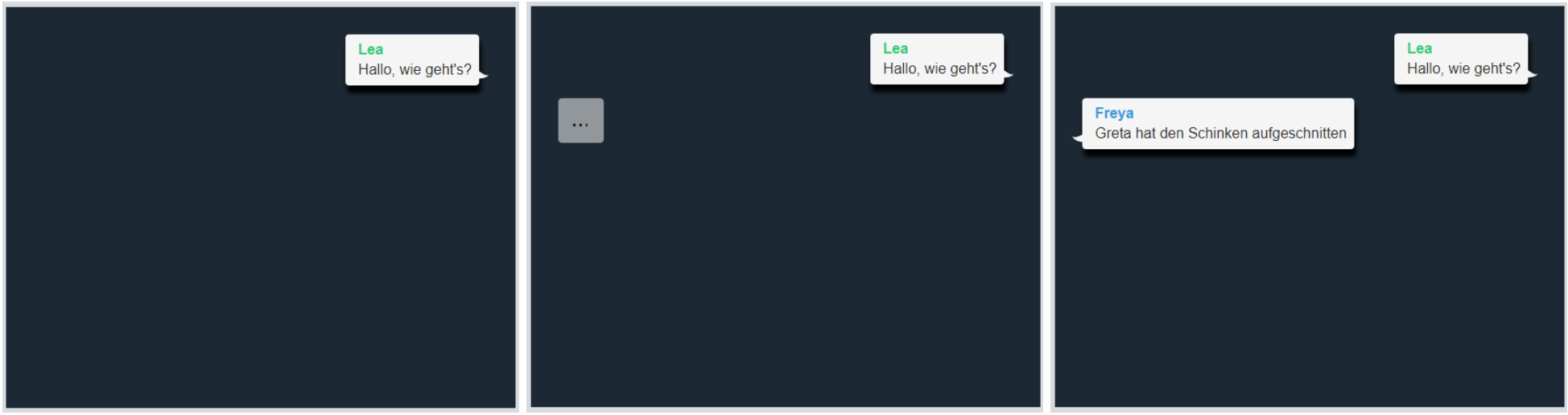



(a) Late Repair:
Repair-U replaces Context-U1,
marked with a pencil emoji

(b) Late Repair:
Typing indicator before Target-U

(c) Late Repair:
Target-U with nonce verb
and ambiguous pronoun

Figure 6.7: Exp I: Example screenshots from the Late Repair condition (7-9)

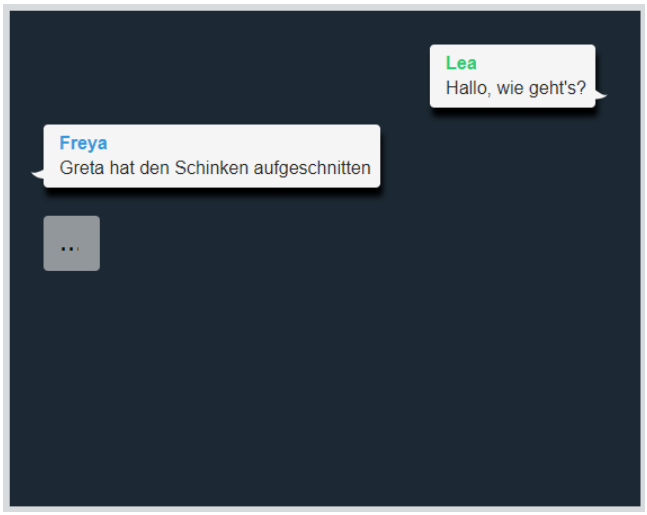



(a) Early Repair: Opening message

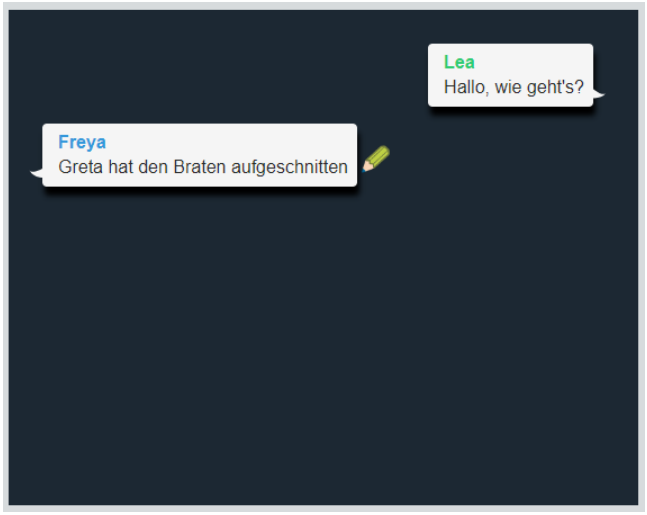
(b) Early Repair: Typing indicator before Context-U1

(c) Early Repair: Context-U1 appears

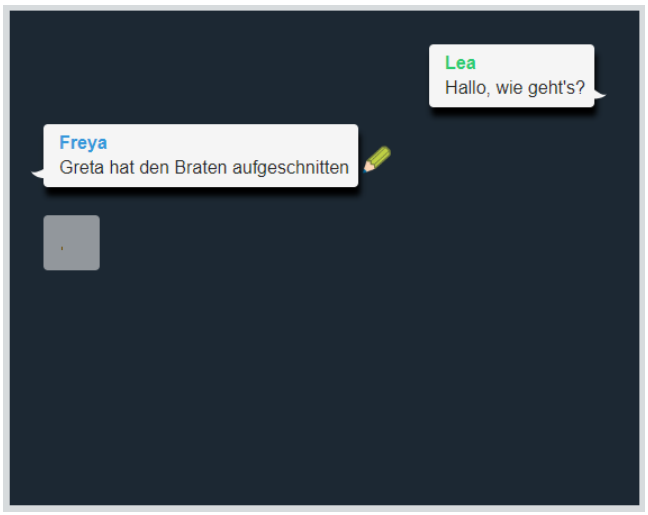
Figure 6.8: Exp I: Example screenshots from the Early Repair condition (1-3)




(a) Early Repair:
Typing indicator before Repair-U




(b) Early Repair:
Repair-U replaces Context-U1
Marked with a pencil emoji




(c) Early Repair:
Typing indicator before Context-U2

Figure 6.9: Exp I: Example screenshots from the Early Repair condition (4-6)


(a) Early Repair: Context-U2 appears

(b) Early Repair: Typing indicator before Target-U

(c) Early Repair: Target-U with nonce verb and ambiguous pronoun

Figure 6.10: Exp I: Example screenshots from the Early Repair condition (7-9)

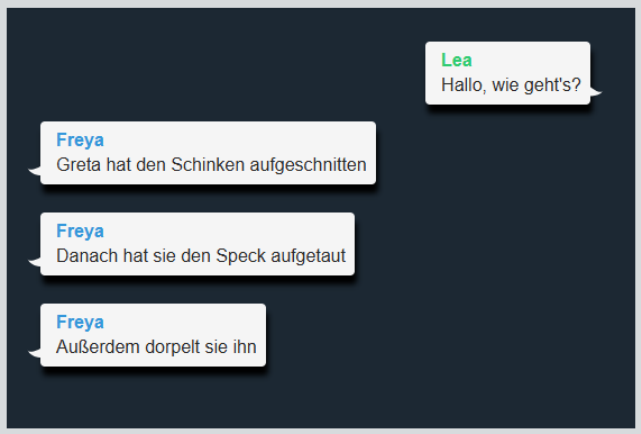



Figure 6.11: Exp I: No Repair: Full dialogue without repair

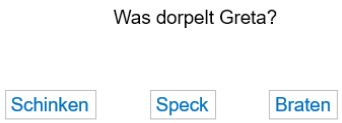



Figure 6.12: Exp I: For all conditions: Multiple-choice question following the dialogue

### 6.3.3 Results

Figure 6.13 illustrates the proportional distribution of referent choices across the three experimental conditions.[7] In the aligned conditions (Early Repair and No Repair), participants showed a strong preference for referents that are both spatially and temporally close (Ref2[8]), with 73.1% and 77.2% respectively. In the Late Repair condition, where temporal and spatial distance were misaligned, participants favored the spatially close referent (Ref2, 75.9%) over the temporally close one (RefRepair, 21.3%). These results suggest that participants generally prefer close referents over distant ones, and that when spatial and temporal cues conflict, spatial closeness has a stronger influence on referent selection.

To investigate how temporal and spatial distance affect pronoun resolution in IM dialogue, a Bayesian multinomial logistic regression model was fitted using the `brms` package (Bürkner, 2017) in R (version 4.4.2; R Core Team, 2024) within the RStudio environment (RStudio Team, 2025). The model predicted participants' referential choices: Ref1, Ref2, and RefRepair based on the experimental condition (No Repair, Early Repair, and Late Repair). The reference levels were set to `No Repair` for `Condition` and `Ref1` for `referential choice`. Random intercepts were included for `participant ID` and `item number` to account for individual- and item-level variation. Following recommendations for multinomial logistic models (Gelman et al., 2008), weakly informative priors (`normal(0, 2)`) were placed on the condition effects, with default priors on intercepts and group-level variance components. Model estimation was carried out using four chains with 4000 iterations each.

[7] In the No Repair condition, the proportion of RefRepair is shown at the top and Ref1 at the bottom, because RefRepair was never actually introduced. In the Early and Late Repair conditions, however, Ref1 was replaced by RefRepair by the time the dialogue was complete. Thus, the final interface presented Ref2 and RefRepair as the competing candidates. The arrangement was chosen to facilitate direct comparison across conditions.

[8] Ref is used as an abbreviation for referent. *Ref1*, *Ref2*, and *RefRepair* denote the referents of the NPs introduced in Context-U1, Context-U2, and Repair-U, respectively. This notation is used consistently across all experiments.

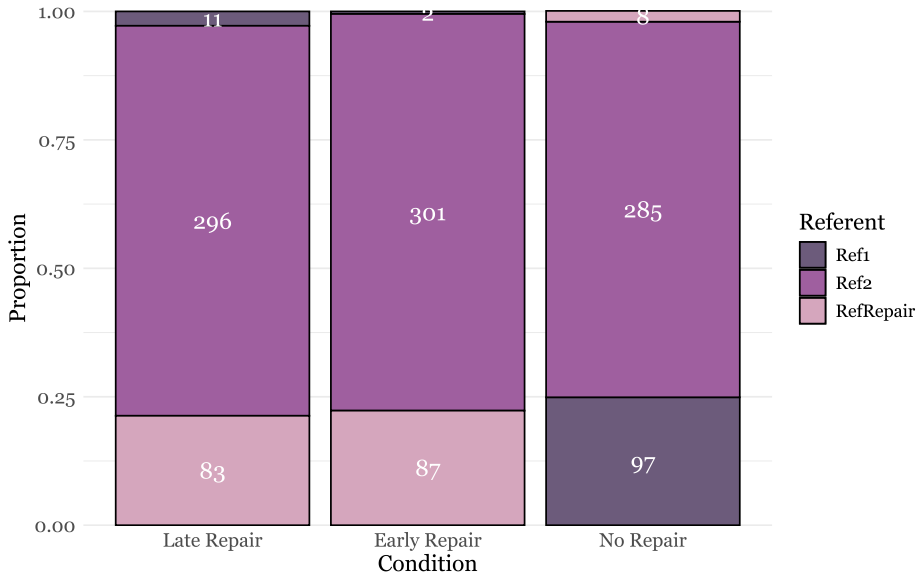



Figure 6.13: Proportional distribution of referent choices across experimental conditions in Experiment I. Each bar represents the proportion of referent choices within each condition. Exact counts are shown on each bar.

Model diagnostics indicated excellent convergence (all R-hat values = 1.00; effective sample sizes > 1000). Posterior predictive checks confirmed that the model accurately captured the observed data distribution. The model returned strong evidence for condition-specific differences in referential choice. Posterior summaries of the fixed effects are presented in Figure 6.14.

To test the experimental predictions, targeted hypotheses were tested using the `hypothesis()` function in `brms`. H1 predicted that when spatial and temporal distance are aligned (in the No Repair and Early Repair conditions), participants would prefer the close referent (Ref2) over the distant alternatives (Ref1 or RefRepair).

In the *No Repair* condition, Ref2 was both spatially and temporally close. Participants showed a clear preference for Ref2 over Ref1, with a posterior estimate of 1.39 (Est. Error = 0.26), 90% credible interval [0.96, 1.82], and posterior probability = 1.00 (Evid.Ratio $\rightarrow \infty$). In the *Early Repair* condition, where Ref2 was again both spatially and temporally close (and

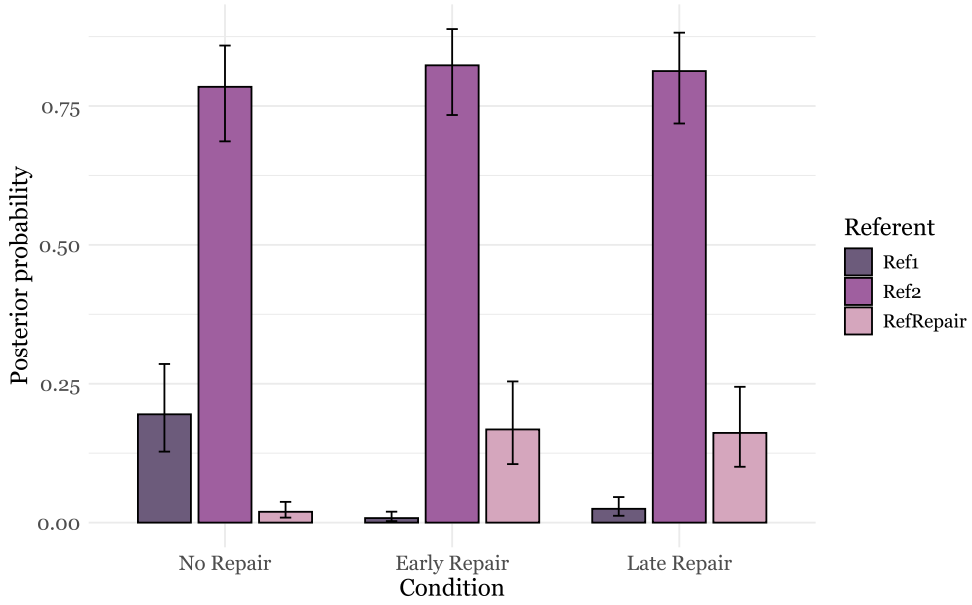



Figure 6.14: Posterior estimates of predicted response probabilities for referent choice across conditions in Experiment I. Estimates are derived from a Bayesian multinomial logistic regression model with `Ref1` as the reference level for `Referent` and `No Repair` as the reference level for `Condition`. Error bars represent 95% credible intervals.

RefRepair was spatially and temporally distant), Ref2 was also preferred. The contrast between Ref2 and RefRepair yielded an estimate of 1.59 (Est. Error = 0.27), with a 90% credible interval of [1.15, 2.05] and a posterior probability of 1.00. These results strongly support Hypothesis 1.

H2 predicted that when spatial and temporal distance are misaligned (in the Late Repair condition), temporal distance would outweigh spatial distance, leading to a preference for RefRepair over Ref2. In the *Late Repair* condition, RefRepair was temporally close but spatially distant, while Ref2 was spatially close but produced earlier. The contrast between RefRepair and Ref2 showed a clear preference for Ref2, with a negative posterior estimate of −1.62 (Est. Error = 0.28), 90% CI [−2.08, −1.17], and posterior probability = 0.00. This provides strong evidence against Hypothesis 2, suggesting that spatial distance outweighs temporal distance when the two cues conflict.

### 6.3.4 Discussion

The results of the present study show that, when spatial and temporal distance are aligned, close referents are more likely to be referred to by subsequent pronominal expressions than distant ones. In both the No Repair and Early Repair conditions, participants overwhelmingly selected referents introduced in the most temporally and spatially close utterance. This replicates previous findings that recency affects referent accessibility and pronoun resolution (Grosz et al., 1995; Arnold, 2010; Givón, 1983). The Late Repair condition, however, presented a situation in which spatial and temporal distance were misaligned. Findings in this condition suggest that spatial presentation can outweigh temporal sequence when the two are brought into conflict by repair.

An interesting observation is the similarity in the absolute number of choices for the Context-U2 referent (the spatially close referent) across all three conditions: 285 in No Repair, 301 in Early Repair, and 296 in Late Repair. Despite the difference in when the repair was introduced: immediately after Context-U1 in Early Repair, or after Context-U2 in Late Repair, participants' preference for the spatially close referent remained remarkably stable. This raises a question: why does the spatially close referent receive nearly equal preference regardless of whether a repair occurred, and irrespective of its temporal relation to other utterances?

One possible explanation is that the spatial position of a referent on the screen plays a decisive role in determining prominence in digital communication settings such as IM. In all three conditions, apart from Target-U, Context-U2 consistently occupied the final visible position in the chat interface. This made Ref2 spatially close at the point of reference resolution, regardless of the temporal order in which it was uttered or repaired. This suggests that spatial closeness at the point of interpretation has a stronger influence on reference resolution than temporal sequence.

Another possible explanation is that repairs, whether early or late, are not fully integrated into the reader's discourse model, at least not in a way that competes with spatial layout. In the Late Repair condition, the repairing

utterance (Repair-U) introduced a new referent that was temporally closer than Ref2, yet spatially distant. If readers treat Repair-U as an "interruption" rather than as a fully integrated discourse move, its influence on prominence may be reduced. This interpretation aligns with theories that conceptualize repairs as a form of metadiscourse, functioning as a discourse bracket (Schiffrin, 1980) rather than introducing new focal information. In IM, where edited messages overwrite previous content and lack auditory or prosodic cues, such edits may be perceived as revisions rather than as new contributions.

A more far-reaching possibility is that self-repair through message editing functions as a different kind of communicative act in IM, one that is not treated as a new contribution to the conversation. This idea is consistent with Clark and Brennan's (Clark and Brennan, 1991) theory of grounding in dialogue, where communication is built collaboratively and shaped by signals of shared understanding. In spoken conversation, repairs are typically marked by hesitations or prosodic cues, which make their purpose and status clear to listeners (Schegloff, 1997, 2000). In IM, however, repairs introduced via the editing function are shown in the interactional layer rather than in the main conversational thread. As a result, recipients may treat them differently from other utterances, perceiving them as revisions rather than as new contributions. Consequently, edited messages may not influence referent prominence in the same way that new utterances do.

Moreover, the stable preference for Context-U2 across three conditions indicates that discourse processing strategies may be shaped by medium-specific expectations. In everyday IM use, readers may rely on simple strategies, such as always focusing on the last visible message, as a way of keeping track of information. Such strategies could easily be carried over into experimental settings. As a result, participants may continue to prioritize spatially close information, even when temporal order would point to a different referent.

In sum, these results suggest that in digitally mediated dialogue, recency must be reconceptualized to include both spatial and temporal dimensions. When these two dimensions are misaligned, forcing a choice between a

spatially close and a temporally close referent for a pronominal expression, the spatially close referent is preferred.

This conclusion, however, must be treated with caution. While the data support the idea that pronoun resolution is guided by spatial distance in IM, the design does not allow me to rule out the independent role of repair. It remains possible that referents introduced in repair utterances were perceived as less prominent, not because they were spatially distant, but because repair in IM is, by nature or convention, treated as less prominent. If so, the reduced preference for referents from the repairing utterance may reflect a property of repair itself, rather than temporal distance being generally less influential.

To address this issue, the next experiment introduces the referents in a similar way but presents the repair as a stand-alone message rather than through the message editing function. This change ensures that the repair is clearly presented as a new discourse contribution. If referents introduced in such stand-alone repairs are still less preferred, this would suggest that repair in IM is generally treated as less prominent, regardless of how it is presented. By contrast, if referents in stand-alone repairs behave like other utterances in the main message thread, that is, if they gain prominence by being temporally close, then the influence of repair in the current experiment can be ruled out. In that case, the findings would support the broader conclusion that, in IM dialogue, spatial distance has a stronger influence than temporal distance on pronoun resolution.

## 6.4 Experiment II: Pronoun Resolution after Quoted Repairs

### 6.4.1 Aim and Hypothesis

This experiment was designed to address an open question from Experiment I: whether the lower preference for referents introduced in repair utterances was due to their spatial distance or to repair being inherently less prominent in IM dialogue. The same stimuli from the first experiment were used in this experiment, but the presentation of repair changed. Instead of replacing earlier content through editing, the repair was shown as a stand-alone message using a simulated quoting function, which visually anchored it to the utterance it corrected (see Figure 6.15).

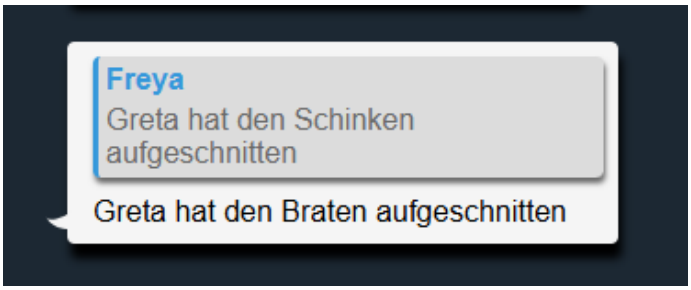



Figure 6.15: Example: A repair presented as a stand-alone message, with the repaired utterance shown in quotation

If referents introduced in stand-alone repairs are treated like those from other utterances in the discourse sequence, they should be just as likely to serve as pronoun antecedents. If, however, such referents remain dispreferred, this would suggest that repair in IM is inherently processed as a less prominent communicative act, regardless of its presentation format.

The hypothesis of this experiment is:

- **H3:** When a repair is presented as a stand-alone message and introduces a referent, that referent is treated in the same way as referents introduced in other stand-alone utterances in terms of being selected as the antecedent of a subsequent pronoun in IM discourse.

### 6.4.2 Method

**Materials**

The experimental materials were identical to those used in Experiment I. Each stimulus consisted of a short IM-style exchange between two interlocutors, designed to create a referential ambiguity in the final utterance (Target-U). Target-U contained the masculine pronoun *ihn* ("it"), whose antecedent could potentially be any of referents from the preceding utterances (Context-U1–Repair-U).

The following example (previously shown in Experiment I) illustrates the structure of a typical stimuli:

(44) a. Lea: Hallo, wie geht's? (Opening question)
b. Freya: Greta hat den Schinken aufgeschnitten. (Context U1)
c. Freya: Danach hat sie den Speck aufgetaut. (Context U2)
d. Freya: Greta hat den Braten aufgeschnitten. (Repair-U)
e. Freya: Außerdem dorpelt sie ihn. (Target-U)
f. *FRAGE:* Was dorpelt Greta?
A: Schinken B: Speck C: Braten

(45) a. *Lea: Hello, how are you? (Opening question)*
b. *Freya: Greta sliced the ham. (Context U1)*
c. *Freya: After that, she thawed the bacon. (Context U2)*
d. *Freya: Greta sliced the roast. (Repair-U)*
e. *Freya: Besides, she is 'dorpeling' it. (Target-U)*
f. *QUESTION: What is Greta 'dorpeling'?*
*A: Ham B: Bacon C: Roast*

The stimuli followed the same four-part structure. Context-U1 and Context-U2 were perfect-tense utterances connected by a coordinating relation; they shared the same human subject but each introduced a different masculine accusative noun phrase. Repair-U functioned as a correction of Context-U1, replacing its object with a new, contextually plausible masculine object. Target-U appeared in the present tense and

contained a subject pronoun coreferential with the subjects in the context utterances, as well as the ambiguous pronoun *ihn*, which served as the critical referring expression. Filler items from Experiment I were also reused.

**Participants**

The experiment was conducted with 43 participants. 7 datasets were excluded due to failed concentration checks, resulting in a final sample of 36 native German speakers (17 male, 19 female; mean age = 35.5 years; age range = 19–60 years). Participants were recruited via Prolific and received £6.83 as compensation for their participation. All participants gave informed consent, and Prolific's pre-screening system was used to ensure that they had no history of language or cognitive disabilities.

**Design**

Experiment II followed a within-subjects Latin Square design, mirroring Experiment I. The three experimental conditions were defined as follows:

- **Late Repair:** Repair-U appeared after Context-U2 immediately and before Target-U. In this condition, the referent from Repair-U was both temporally and spatially close to the subsequent pronoun.
- **Early Repair:** Repair-U appeared after Context-U1 but before Context-U2. Here, the referent from Repair-U was temporally and spatially more distant from the subsequent pronoun than the referent from Context-U2.
- **No Repair:** Context-U1, Context-U2, and Target-U were presented sequentially without repair. This condition served as a baseline.

While the condition labels are retained for continuity, the presentation of the repairing utterance differed from Experiment I in a crucial way. In Experiment I, Repair-U was introduced using a simulated message editing function that visually replaced Context-U1 in the chat layout,

creating a misalignment between temporal distance and spatial distance in the Late Repair condition. In this setup, Repair-U explicitly appeared in a new message bubble following either Context-U1 or Context-U2, depending on the condition, and was accomplished by a quote of the original utterance it repaired (again, see Figure 6.15). This ensured that repair was clearly marked as a new, temporally and spatially aligned utterance, while maintaining transparency about what was being corrected.

**Procedure**

Similar to Experiment I, each dialogue was presented dynamically to simulate real-time chat interaction. Utterances unfold one by one, accompanied by visual cues such as typing indicators. Typing and reading times were dynamically calculated to reflect natural conversational flow. Based on feedback from the first experiment, in which some participants reported that the original typing speed (52 WPM) led to boredom due to long waiting times, the typing speed was adjusted to 61 WPM in this experiment. The reading speed remained unchanged at 200 WPM, following average estimates for digital reading reported by Siegenthaler et al. (2011).

In the **Late Repair** condition, each item began with an opening message from speaker A (e.g., Lea), as shown in Figure 6.16a. After a brief delay based on the message length and reading speed, a looping typing indicator (three animated dots) appeared (Figure 6.16b). Once the adjusted typing time passed, the indicator disappeared and the first utterance (Context-U1) from speaker B (e.g., Freya) appeared (Figure 6.16c). This sequence was repeated for Context-U2, which was preceded by another typing phase (Figures 6.17a and 6.17b).

The third utterance (Repair-U) served as a correction of Context-U1. Different from Experiment I, where Repair-U replaced Context-U1 via editing function, here Repair-U appeared as a stand-alone message spatially below Context-U2, visually referencing the original utterance through quoting. The quoted Context-U1 was shown in a grey background styled to mimic a typical quoting feature in WhatsApp (Figure 6.18a). As in Experiment 1, the delay before Repair-U appeared was adjusted to reflect

the time required to delete and retype the corrected object (Figure 6.17c). After a brief reading pause and a final typing indicator (Figure 6.18b), the final utterance (Target-U), containing the nonce verb and critical pronoun, was displayed (Figure 6.18c).

A similar sequence was used in the **Early Repair** condition, where Repair-U appeared before Context-U2. In the **No Repair** condition, since there was no repair, only Context-U1, Context-U2, and Target-U were shown. Example screenshots of these conditions are provided in Figures 6.19 (Early Repair) and 6.20 (No Repair).

After the full dialogue was presented, participants were given four seconds to review the conversation before being automatically redirected to a new screen displaying a multiple-choice question (Figure 6.21). The question asked participants to identify the referent of the nonce verb in Target-U. The same fillers from Experiment 1 were used in this experiment.


(a) Late Repair: Opening message

(b) Late Repair: Typing indicator before Context-U1

(c) Late Repair: Context-U1 appears

Figure 6.16: Exp II: Example screenshots from the Late Repair condition (1-3)


(a) Late Repair: Typing indicator before Context-U2

(b) Late Repair: Context-U2 appears

(c) Late Repair: Typing indicator before Repair-U

Figure 6.17: Exp II: Example screenshots from the Late Repair condition (4-6)


(a) Late Repair: Repair-U appears as a stand-alone repair message quoting Context-U1

(b) Late Repair: Typing indicator before Target-U

(c) Late Repair: Target-U appears

Figure 6.18: Exp II: Example screenshots from the Late Repair condition (7-9)


Figure 6.19: Exp II: Early Repair: Full dialogue with Repair-U appearing as a stand-alone message that quotes Context-U1, occurring directly after Context-U1 and before Context-U2


Figure 6.20: Exp II: No Repair: Full dialogue without Repair

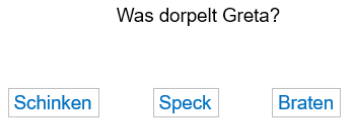



Figure 6.21: Exp II: For all conditions: Multiple-choice question following the dialogue

### 6.4.3 Results

Figure 6.22 visualizes the distribution of response types across the three experimental conditions.[9] In the *No Repair* condition, where temporal and spatial recency were aligned, *Ref2* was selected in 81.7% of trials, compared to 15.8% for *Ref1*, indicating a strong preference for the most recent referent. In the *Early Repair* condition, this preference remained robust: *Ref2* was chosen in 85.3% of cases, while the spatially distant referent *RefRepair* received 12.2% of responses. The *Late Repair* condition showed that repair presented as a stand-alone utterance behaves similarly to other stand-alone utterances concerning referent prominence in the sense that *RefRepair* was chosen most frequently (79.7%), while *Ref2* received 18.1% choices.

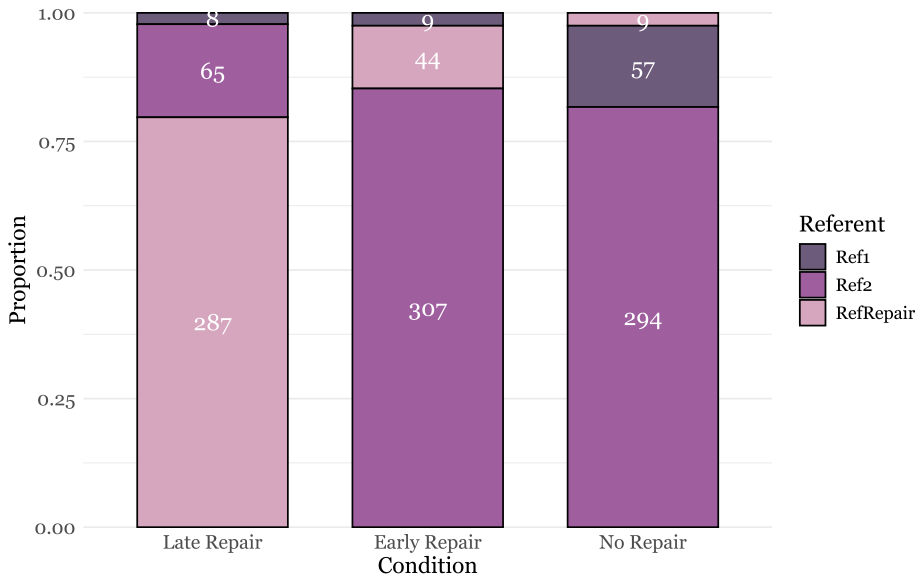



Figure 6.22: Proportional distribution of response types across experimental conditions. Each bar represents the proportion of response types within each condition. Exact response counts are shown on each bar.

[9]The stacking order of referents differs across conditions to reflect their relative spatial distance at the point of pronoun resolution. Experiment II tests whether referents introduced in stand-alone repairs behave like those in other stand-alone utterances. To allow for a direct comparison, referents in each condition were stacked according to spatial distance: in the Late Repair condition, RefRepair was spatially closest to the subsequent pronoun, whereas in the No Repair and Early Repair conditions, Ref2 was closest.

The analysis of Experiment II followed the same general process as in Experiment I. All statistical analyses were conducted in R (version 4.4.2; R Core Team, 2024) using the `brms` package (Bürkner, 2017) within the RStudio environment (RStudio Team, 2025). The goal was to test whether referents introduced in stand-alone repair utterances are selected as antecedents of subsequent pronouns in the same way as referents introduced in other stand-alone utterances.

The model predicted participants' referential choices (*Ref1*, *Ref2*, *RefRepair*) based on the experimental condition (No Repair, Early Repair, Late Repair). The reference levels were set to *No Repair* for Condition and *Ref1* for referential choice. Random intercepts were included for participant ID and item number to account for by-participant and by-item variability. Following recommendations for multinomial logistic models (Gelman et al., 2008), weakly informative priors were specified: normal priors for fixed effects and Student-*t* priors for intercepts and group-level standard deviations. Model estimation was carried out using four chains with 4000 iterations each.

Model diagnostics indicated excellent convergence (all $\hat{R}$ values = 1.00; effective sample sizes > 1000). Posterior predictive checks confirmed that the model accurately captured the observed data distribution. The model returned strong evidence for condition-specific differences in referent choice. Posterior estimates of predicted response probabilities are presented in Figure 6.23.

To test the experimental predictions, targeted hypotheses were evaluated using the `hypothesis()` function in the `brms` package.

When **No Repair** occurs, the referent introduced in Context-U2 is more likely to be selected as the antecedent of the subsequent pronoun than the referent from Context-U1. This preference is confirmed, with an effect estimate of 1.96 (Est. Error = 0.25), 90% CI [1.56, 2.37], and a posterior probability of 1.00. This indicates that, in the absence of repair, the referent introduced at the shortest distance is preferred.

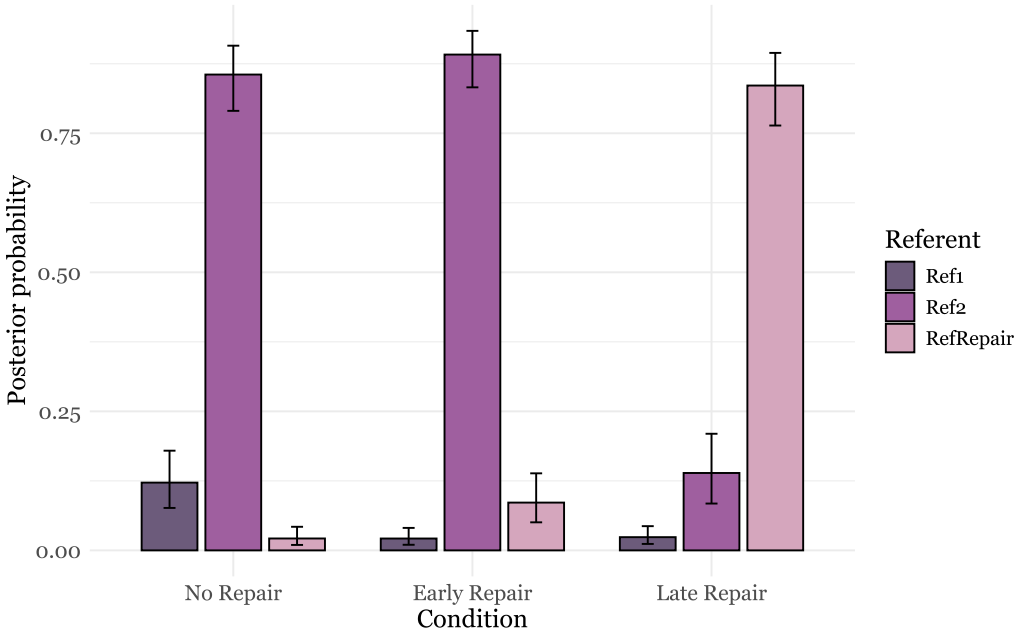



Figure 6.23: Posterior estimates of predicted response probabilities for referent choice across conditions in Experiment II. Estimates are derived from a Bayesian multinomial logistic regression model with `Ref1` as the reference level for `Referent` and `No Repair` as the reference level for `Condition`.

In the **Early Repair** condition, the referent in Context-U2 is expected to be more likely chosen than the referent introduced in the Repair-U. Data shows that Ref2 was preferred over RefRepair, with an estimate of 2.34 (Est. Error = 0.29), 90% CI [1.88, 2.82], and posterior probability = 1.00. This suggests that the closer referent is consistently selected.

In the **Late Repair** condition, repair is presented after Context-U2, and according to H5, the referent it introduces is more likely to be selected as the antecedent of the subsequent pronoun than the referent from Context-U2, indicating that stand-alone repairs function like other utterances in contributing to prominence. This is supported by the data: participants preferred RefRepair over Ref2, with an effect of 1.80 (Est. Error = 0.27), 90% CI [1.38, 2.27], and a posterior probability of 1.00. This indicates that when a repair occupies the position closest in distance to the subsequent pronoun, it functions like any other utterance in shaping referent prominence.

### 6.4.4 Discussion

The goal of Experiment II was to test whether repairs, when presented as stand-alone messages in IM dialogue, are treated in the same way as other stand-alone messages in terms of their potential to serve as antecedents of subsequent pronouns. This experiment builds directly on the findings of Experiment I, in which repairs were introduced through simulated message editing. That manipulation created a misalignment between temporal and spatial distance: although the repair was produced later in time, it visually replaced an earlier utterance, making it spatially distant. The result of Experiment I showed a strong preference for spatially close referents, but it remained unclear whether this was due to the reduced prominence of repair itself, the spatial position of the repair, or a combination of both.

Unlike the message-editing format used in Experiment I, in this experiment, repairs were displayed as stand-alone messages, using a quoting function to show which earlier utterance was being corrected. This preserved the natural alignment between spatial and temporal distance while maintaining a clear connection to the original utterance.

Results of this experiment support H5. In the Late Repair condition, participants overwhelmingly preferred the referent introduced in the repair (Repair-U), which was both spatially and temporally close. This mirrored the preference for Context-U2 in the No Repair and Early Repair conditions, which likewise occupied the closest position.

This result suggests that repair in IM is not inherently treated as a metalinguistic adjustment or backgrounded update. Unlike spoken interaction, where repairs are typically accompanied by prosodic cues, disfluencies, or pauses that mark them as disruptions to the ongoing flow of speech (Levelt, 1983; Tree, 1995; Clark and Fox Tree, 2002), IM lacks such auditory signals. Instead, readers rely more heavily on visual presentation and message order to interpret discourse structure. When repairs are displayed as stand-alone messages, they appear to function more like elaborations or clarifications that extend the discourse.

However, this interpretation must be treated with caution. The type of repair tested in this experiment represents a specific and narrow subset of repair, namely, self-initiated self-repair that is made visible to the recipient. This corresponds to what Meredith and Stokoe (2014) term visible repair and to what Levelt (1983) classify as error repair rather than appropriateness repair. It remains an open question whether other forms of repair, such as rephrasings, corrections of misspellings, or repairs initiated by another speaker, would have similar effects.

In sum, the findings from Experiment II suggest that repairs presented as stand-alone messages contribute to discourse prominence in the same way as non-repair utterances, consistent with previous findings that recency is a key determinant of referential choice when other prominence-lending cues are held constant (Clark and Sengul, 1979; Arnold, 2025). The status of an utterance as a repair, therefore, does not appear to block or override this recency effect.

## 6.5 General discussion

### 6.5.1 Theoretical Implications

Experiments I and II investigated how prominence is established in IM dialogue, with a particular focus on the role of recency. While recency is an established factor influencing referent accessibility, the IM environment complicates the notion of recency and suggests a better understanding of recency in terms of spatial and temporal distance.

Experiment I used a simulated message editing function to create a misalignment between temporal and spatial distance. The results showed a clear preference for the closer referent when spatial and temporal distance aligned, and in the case of misalignment, the spatially closer referent was preferred.

However, because the misalignment in Experiment I was created through simulated message editing, it remains unclear whether the reduced prominence of the referent in the repair was due to its spatial distance or

to the repair itself. In other words, participants may have considered a referent less prominent simply because it appeared in a repair, regardless of its temporal or spatial position.

To investigate this further, Experiment II presented repairs as stand-alone messages, with a quoting function indicating which message was being corrected. This design kept temporal and spatial distance aligned, allowing for a direct comparison between referents introduced in repairs and those in other recent utterances. The results showed that referents from stand-alone repairs were resolved in much the same way as referents from other stand-alone messages, so that the potential influence of repair in Experiment I can be ruled out. Taken together, the two experiments demonstrate that in IM dialogue, spatial presentation plays a decisive role in reference resolution. When spatial and temporal cues are aligned, recent referents are consistently preferred. But when the two are in conflict, spatial distance appears to override temporal distance.

A comparison between the Late Repair condition in Experiments I and II is particularly interesting. In both cases, the repair was introduced temporally most recently in the discourse. However, only in Experiment II did it also appear spatially most recent. In that condition, participants showed a clear preference for the referent introduced in the repair, suggesting that it is not the status of repair per se that reduces prominence, but rather how the repair is presented.

This comparison also offers theoretical insight into the status of repair in digital communication. In spoken dialogue, repair is often marked by disfluencies, pauses, or prosodic shifts, and is sometimes interpreted as a special kind of act that interrupts the discourse (Levelt, 1983; Tree, 1995; Clark and Fox Tree, 2002; Schegloff, 1997). In IM, such auditory cues are absent. Instead, repairs are conveyed through visual means: edited messages, quoted utterances, or metalinguistic markers. Results support the view that the contribution of repair to prominence depends on how it is presented.

The findings from Experiments I and II provide insight into how discourse prominence is established in IM. Recency as a prominence-

lending feature should be considered as two dimensions: temporal and spatial distance. This distinction is absent from most traditional models of discourse processing, which assume that utterances unfold linearly in time and are processed in order, an assumption valid for spoken dialogue and conventional written discourse, but challenged by IM.

While the experiments clearly showed that spatial distance overrides temporal distance in IM, they do not fully explain why this happens. One explanation is that participants rely on habitual reading strategies developed through reading of traditional written text, where, in German culture, the most recent text is displayed at the bottom. Over time, people may develop an attentional bias toward the last visible utterance, treating it as the most relevant. This would align with research in visual cognition showing that spatial location influences language processing and memory (Spivey et al., 2002; Gleitman et al., 2007; Esaulova et al., 2021).

Another possibility concerns the layering of the repair utterance itself. In Experiment I, repairs introduced via message editing created a misalignment. At the same time, it placed the repair in the interactive layer, replacing earlier utterances. By contrast, in Experiment II, when the repair was presented as a stand-alone message at the bottom of the chat, it appeared in the main conversation thread like other contributions. This difference leads to another theoretical insight: modifications that are not presented in the main conversational layer as a stand-alone message may behave differently.

### 6.5.2 From Recency to Interactional Structure

Findings of the two experiments point to the role of layering in shaping how dialogue is processed in IM. Building on this, the next part of this dissertation turns from questions of distance to questions of layering. Specifically, whether utterances in the main conversation thread and in the interactive layer are interpreted differently in IM dialogues. Just as message editing introduced a misalignment between the spatial and temporal distance, utterances in the interactive layer often disrupt the expected linearity of conversation. They are typically out-of-turn, spatially

anchored to earlier messages, yet they still contribute to the ongoing interaction. This observation points to a broader question about the nature of layered discourse in IM, which will be taken up in the next chapter.

# Chapter 7

# The Back Channel in Instant Messaging Dialogue

The two experiments presented in Chapter 6 investigated referent prominence in IM dialogue, focusing on how temporal and spatial distance influence the resolution of ambiguous pronoun references. Misalignment between spatial and temporal cues was introduced by manipulating the order and position of self-repair. The results showed that repairs introduced through the message editing function, i.e., when they did not appear as stand-alone utterances, were treated as less prominent. By contrast, stand-alone repair utterances behaved similarly to other stand-alone turns with respect to referent prominence. These findings suggest that not all contributions in IM dialogue are treated equally with respect to prominence, and that their prominence status may depend on how they are integrated into the unfolding discourse. In particular, a distinction emerges between utterances that appear as stand-alone messages and those that are attached to earlier contributions.

To account for this distinction, this chapter revisits the original notion of the back channel, as introduced by Yngve (1970), and distinguishes it from the more widespread use of the term "backchannel(s)" as short feedback

signals. Building on this foundation, the chapter proposes a structural account of the back channel in IM dialogue, suggesting that features such as message edits and emoji reactions constitute a parallel interactional layer. This reframing invites a reconsideration of how prominence is assigned in digital communication, where the hybrid form of communication and multi-channel structures reshape the organization of dialogue.

## 7.1 The Back Channel and Backchannels

The concept of the *back channel* was first introduced by Yngve (1970) in his foundational work on turn-taking in conversation. It describes a structural component of dialogue that runs parallel to the main speaker's contributions. In Yngve's model, conversation is not a strictly turn-based alternation but a layered system in which both speaker and listener may contribute simultaneously. As he writes, "both the person who has the turn and his partner are simultaneously engaged in both speaking and listening. This is because of the existence of what I call the back channel, over which the person who has the turn receives short messages such as 'yes' and 'uh-hun' without relinquishing the turn. The partner, of course, is not only listening, but speaking occasionally as he sends the short messages in the back channel" (Yngve, 1970, p. 568). Crucially, for Yngve, the back channel is not merely a collection of expressions but a structural layer in which one participant contributes while the other speaks.

Over time, however, the term *backchannel* (often pluralized as *backchannels*) has come to refer more narrowly to the specific forms of listener feedback rather than to the structural layer itself. Early conversation analysis research focused on identifying and classifying such expressions, particularly in terms of their timing, sequential placement, and interactional function (e.g., Schegloff, 1982; Jefferson, 1984; Heritage, 1985; Gardner, 1997). Under this view, backchannels were treated not as a structural space,

but as a class of responsive behaviors. This approach has produced extensive typologies of backchannel expressions, including verbal tokens such as *hm*, *uh-huh*, or *okay*, as well as nonverbal behaviors like nodding or smiling (e.g., Gardner, 2001; Bavelas et al., 2000; Stivers, 2008; Drummond and and, 1993).

It is now widely acknowledged that *backchannels* are brief addressee responses, including verbal forms (e.g., *yeah*, *oh*, *okay*, *mhm*) and visual displays (e.g., facial expressions, nods, gestures). Backchannels can be seen as a subset of communicative acts that take place within the structural layer defined as the *back channel*. In Yngve's words, back channel activity may range from a listener's attitude of attention and interest to short comments such as "Oh, I can believe it," or even brief questions like "You've started writing it then — your dissertation?" (Yngve, 1970). Thus, while backchannels are individual forms or expressions, the back channel refers to the structural domain in which they occur. In this view, the back channel is the space for listener contributions, which is non-floor-claiming and simultaneous, that support the main discourse.

The concept of *feedback* further complicates this picture. Feedback is a broader category referring to communicative behavior by which participants exchange information about mutual understanding, attention or perception. As defined by Allwood et al. (1992), feedback includes the exchange of cues related to four communicative functions: contact, perception, understanding, and attitudinal reactions. Backchannels, then, are one type of feedback, which is a minimal and spontaneous form that expresses understanding or alignment without disrupting the speaker. However, not all feedback is back channel behavior. For example, explicit corrections, evaluations or clarification requests may require speaker engagement or even shift the floor. Such responses typically belong to the main channel, not the back channel, because they change the topical progression or interrupt the speaker.

Consider the example in 46, adapted from Allwood et al. (1992):

(46) a. A: and the candy papers they get into into//into that waste paper basket then
b. B: really
c. A: yes
d. B: that's unusual

Here, the utterance (46-b) "really" is a feedback utterance, but not a backchannel in Yngve's sense. It interrupts the speaker and places an obligation on A to respond. A's "yes" is a response to that feedback, showing that B's contribution momentarily shifted the turn structure. Thus, although *feedback* refers to a communicative function, *backchannels* that take place in the back channel refer to its most minimal and spontaneous forms, and *the back channel* is the structural layer where such feedback may be embedded without disrupting the speaker's turn.

## 7.2 The Back Channel in Face-to-Face Communication

### 7.2.1 Multimodality and the Channel Concept

Face-to-face interaction is inherently multimodal. Participants in conversation rely not only on spoken language, but also on a wide range of nonverbal resources such as gesture, gaze, and facial expression (Clark, 1996). These modalities work together to allow interlocutors to construct meaning, manage turn-taking, and coordinate mutual understanding. From this perspective, dialogue is not a linear sequence of verbal turns, but a dynamic and layered system in which multiple channels cooperate to convey meaning (Kendon, 2004, ch. 7).

The idea that communicative behavior can be distributed across different channels was established early in communication theory. In 1948, Shannon defined a channel as the medium used to transmit the signal from transmitter to receiver. This foundational model conceptualized communication as the transfer of encoded information through a physical pathway, such as air for spoken language or electrical signals for telecommunication.

This concept is reflected in how audio and video streams are processed in media technology. These channels may be temporally and spatially coordinated, yet they remain functionally independent. For instance, the sound and video tracks of a film contribute distinct elements to the overall experience. Each channel can be edited or interpreted separately, but their combination results in a cohesive output. Crucially, these channels are not turn-based or sequential, but parallel and continuous.

Yngve's (1970) concept of the back channel draws on a similar structural intuition. Just as the sound channel does not need to be paused for the video channel to play, the speaker and the listener can simultaneously engage in different communicative channels without violating the structure of turn-taking. The existence of multiple channels allows for a layered interaction: one channel manages sequential turn-taking and topical progression, while another handles feedback and alignment. This architectural layering enables simultaneous communication without interference.

The multimodal view of human interaction provides crucial context for understanding the structural distinction between the main channel and the back channel. While the main channel is typically associated with the speaker's verbal contributions, the back channel consists of brief, non-floor-claiming cues produced by the listener (Yngve, 1970). These include brief verbal utterances (e.g., *mhm*, *yeah*), and other multimodal resources such as gesture, gaze, facial expression, body posture, and body movement (Mondada, 2016). In this sense, the back channel appears alongside the main channel without competing for the floor.

Understanding the back channel as a structurally distinct, multimodally realized channel of interaction sets the stage for examining how such mechanisms are transformed in digital communication. Before turning to this recontextualization, the next section will focus more closely on the forms and functions of back channel activity as it is realized in face-to-face dialogue.

### 7.2.2 Verbal and Nonverbal Back Channel Activity

According to Clark and Schaefer (1987), contributing to a conversation is a collaborative process involving both speaker and addressee. In this view, communication is not just about speaking and listening, but about jointly establishing a shared belief that the addressee has understood what the speaker intended. Clark and colleagues propose that speakers and listeners work together to reach mutual belief about comprehension. However, the speaker cannot determine whether they have been understood unless the listener provides some form of evidence. The most common type of such evidence is acknowledgments, brief responses that show understanding. As Clark and Brennan (1991) note, these acknowledgments often take the form of what are commonly called back channel responses.

Back channel activity is realized through a range of brief verbal and nonverbal signals that listeners produce while the speaker maintains the floor. Although minimal in form, they play a crucial role in maintaining the flow and coherence of face-to-face conversation (Schegloff, 1982; Gardner, 2001).

Verbal backchannels are frequently studied. Common examples include short responses like *mhm*, *uh-huh*, *yeah*, *right*, and *okay*. These expressions are often produced at points where the speaker signals the possibility of turn continuation or completion (Schegloff, 1982; Jefferson, 1984), which is also called transition relevance place (Sacks et al., 1974). Their primary function

is to indicate that the listener is following and the speaker may continue. These can be further categorized into *generic* and *specific* backchannels (Tolins and Fox Tree, 2014). Generic backchannels, such as *uh-huh* or *yeah*, are not content-specific but function to indicate continued attention or understanding. Specific backchannels respond to the content, often expressing surprise or evaluation, as in *oh wow*. Backchannels, in this sense, represent a formal category of communicative acts: minimal signals that maintain the interactional flow without disrupting speaker turns.

In face-to-face communication, back channel activities are not limited to verbal backchannels, but also include a rich variety of nonverbal behaviors that serve similar interactional functions. These include head nods, gaze shifts, eyebrow raises and other signals (Goodwin, 1980; Duncan, 1974; Allwood et al., 1992; Stivers, 2008; Heylen, 2006). These cues are crucial for maintaining conversation flow, as they allow listeners to provide feedback without taking the floor (Yngve, 1970). For instance, head nods are commonly used by listeners to signal understanding or agreement and are often temporally aligned with the speaker's utterances (Duncan, 1974; McClave, 2000; Stivers, 2008; Whitehead, 2011). Gaze behavior similarly coordinates turn-taking and listener engagement (Kendon, 1967; Goodwin, 1981; Rossano, 2013; Wagner et al., 2014). These cues are essential components of the multimodal nature of dialogue, allowing interlocutors to maintain interaction even when verbal feedback is minimal or absent (Goodwin, 1981).

## 7.3 Reframing the Back Channel for Instant Messaging Dialogue

The previous sections described how the back channel works in face-to-face interaction. There, listeners can show attention and understanding through short verbal cues, gestures, and other nonverbal signals while the speaker is still talking. These contributions are possible because face-to-face conversation happens in real time and across multiple communicative modes. In digital contexts such as IM, the realization of the back channel differs in significant ways. To understand how users provide feedback, react, or clarify without disrupting the main conversation, the concept of the back channel should be reconsidered. In this section, I propose a reframing of the back channel that adapts Yngve's original notion to the layered structure of IM platforms.

### 7.3.1 Why Redefine the Back Channel in IM

IM differs from face-to-face conversation in several important ways that affect how people give feedback or show understanding. In spoken interaction, back channel signals like "mhm", head nods, or facial expressions can be given while the speaker is still talking. This kind of overlap is possible because face-to-face interaction happens in real time and allows for simultaneous use of different communicative modes, such as speech, gesture, and gaze (Clark, 1996; Kendon, 2004; Mondada, 2016).

In IM, although voice messages are technically supported on many platforms, communication remains predominantly text-based. Even when both participants are typing at the same time, the system presents each message in the order of delivery, creating a visibly linear and sequential format. Minor timing differences, even on the scale of milliseconds, determine the spatial order in which messages appear. As a result, there

is no clear mechanism for displaying real-time feedback while someone else is “talking”. Any sense of simultaneity is lost at the point of display. A consequence is that the main conversational thread in IM is primarily organized as a spatial sequence rather than a temporal one. In face-to-face dialogue, conversation unfolds in time, whereas in IM it unfolds on screen, message by message. The results of Experiments I and II have already shown that readers rely strongly on this spatial ordering when resolving pronouns. This makes spatial organization central for understanding how back channel activity should be defined in IM.

Despite this constraint, IM platforms offer new tools that support functionally similar interactional practices. Users can react to earlier messages using emojis or make edits to previously sent utterances. These practices do not interrupt the linear message flow, but they still allow participants to give feedback or clarify meaning. In this sense, IM enables a form of back channel activity that occurs alongside the main thread of message exchange.

A difference in IM is that back channel behaviors often appear after the original message in the main channel, and sometimes several turns later. For example, a user might respond to an earlier message by adding a thumbs-up emoji or by editing the message using the platform’s edit function. These responses do not start a new topic or take over the conversation, instead, they refer back to something already said, offering clarification, feedback, or a reaction. In doing so, they create a layer of interaction that runs parallel to the main sequence of utterances. Importantly, these actions do not disrupt the visible message flow or break the spatial order of messages. Rather, they form a structurally separate but supportive layer of participation, which is similar in function to the back channel as originally defined by Yngve, but shaped by different spatial and temporal conditions. They reflect how IM users manage feedback and maintain coordination without interrupting or redirecting the ongoing dialogue.

Because of these differences, applying the concept of the back channel to IM requires redefinition. In spoken dialogue, what counts as a separate interactional "channel" is grounded in the possibility of real-time simultaneity. In IM, by contrast, the distinction is motivated by the visual organization of the interface: the main channel is the spatially ordered sequence of stand-alone messages, while the back channel consists of contributions that attach to this sequence. This reframing links the definition of the back channel directly to the spatial perspective that proved central in Experiments I and II.

### 7.3.2 Defining the Back Channel in IM

As discussed above, the back channel in IM needs to be understood differently from the back channel in face-to-face communication. Although the forms are different, the function remains comparable: to support the conversation without taking over the turn. In IM, users still give feedback and signal understanding, just in ways that fit the digital format.

In this context, the back channel in IM refers to a separate layer of interaction that allows participants to respond to earlier messages without changing the ongoing conversational sequence. Unlike the main channel, which consists of stand-alone utterances displayed as sequential turns in the message thread, back channel contributions are structurally subordinate. Because IM lacks real-time overlap and multimodal cues like nods or eye contact, back channel behavior is expressed through actions such as editing a message or reacting with an emoji.

Back channel activity in IM is realized through *participant-initiated actions* that:

- **Do not introduce new turns or topics**, but respond to, modify, or comment on existing ones.
- **Appear in a structurally layered position**, either visually attached to previous messages (e.g., as emoji reactions) or made after the message was already sent (e.g., message edits).

According to this definition, several consequences follow:

- Back channel activity in IM is **enabled by specific platform features**, including:
  - *Emoji reactions*, which provide affective or evaluative feedback without adding a new message;
  - *Message edits*, which revise previously sent content without initiating a new turn;
- Back channel activity in IM can **fulfill backchannel functions**, such as expressing understanding, agreement, or engagement, but it can also **serve functions not typical of traditional backchannels**, such as self-repair. Importantly, these functions are achieved without interrupting the speaker or taking over the speaker's role.

The distinction between the main channel and the back channel in IM is not based on the content of a message, but on how it is presented in the chat interface and what structural role it plays. Messages that appear as stand-alone utterances in the main thread constitute the main channel, as they participate in turn-taking.

In this sense, although *quoted replies* refer to previous content, they are presented as independent stand-alone messages and occupy a sequential position in the main thread. They initiate new turns and thus belong to the main channel. Following the proposed definition, one may further hypothesize that *emoji reactions* function as back channel cues when they appear as icons attached to a prior message, but that the same emoji, when sent as a stand-alone message, should be treated as a full turn in the main channel.

It is also important to distinguish between participant-initiated back channel activity and platform-generated technical signals. Automated features such as *read receipts* or the blue tick in WhatsApp (see Section 3.2.1) are system-controlled and do not always reflect communicative intent. For example, a message may appear as "seen" simply because it was displayed on the screen, even if the recipient is not actively attending, or the recipient may have disabled read receipts altogether. Although such technical cues can shape users' perceptions of availability, they are not under the direct control of participants and cannot be treated as back channel behavior in a strict sense.

Only user-initiated actions that are *visible to all participants* are counted as back channel contributions. Private actions, such as deleting a message for oneself, are not communicative moves in the shared conversational space and therefore fall outside the definition. [1]

[1] A related feature is *message recall*, which most IM platforms implement in two ways: *delete for everyone* and *delete for me*. The former removes the message from the conversation for all participants, while the latter removes it only from the sender's interface, leaving it visible to recipients. This function was first introduced in Chapter 3, Section 3.2.1. Whether message recall (delete for everyone) should be classified as back channel activity is not entirely clear. On the one hand, it does not introduce a new contribution and thus shares features with back channel activity. On the other hand, it alters the visibility of the main channel itself, making it a borderline case. For the purposes of this dissertation, I treat it as a potential candidate for back channel activity, while leaving its precise classification open for future research.

In sum, the back channel in IM is a user-driven, visually and temporally flexible layer of interaction that supports the ongoing conversation without disrupting it. While it lacks the real-time simultaneity of the back channel in face-to-face communication, it plays a comparable role in maintaining coordination, feedback, and mutual understanding in digital dialogue.

### 7.3.3 Contrasting the Main Channel in IM

To understand the back channel in IM, it is equally important to define what constitutes the main channel. In IM dialogue, the main channel refers to the visible sequence of stand-alone messages that appear in the central message thread. These messages form the primary flow of conversation, managing turn-taking, and contributing to the topic development of the dialogue.

Main channel utterances are typically composed as complete turns. They may initiate a topic, respond to a question, offer new information, or perform other discourse functions that move the discourse forward. Each message is displayed in its own message bubble, timestamped, and aligned with the speaker's position in the chat interface. Because IM is a turn-based medium, each message in the main channel occupies a clear position in the interactional sequence.

For example, sending a 😰 emoji as a stand-alone message constitutes a turn in the main channel, as shown in Figure 7.1 (previously introduced in Section 4.1.2). In contrast, attaching the same emoji as a reaction to a prior message functions as a back channel response, as illustrated in Figure 7.2. Similarly, a quoted reply belongs to the main channel because it appears as an independent message and takes part in the ongoing turn sequence, even though it refers to earlier content.

Importantly, the main channel in IM is linear and organised in a certain way in space: each message contributes to the growing textual record of

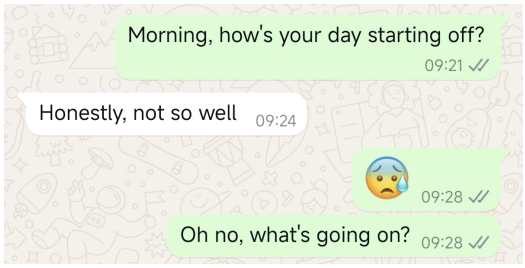



Figure 7.1: Emoji used as a stand-alone message in the main channel. The emoji appears in its own message bubble, contributing a new turn to the conversation.

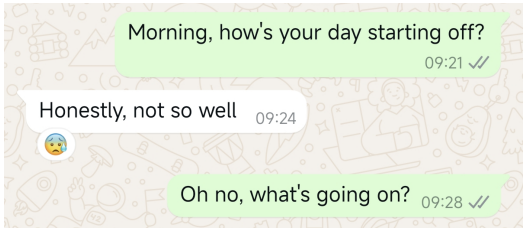



Figure 7.2: Emoji used as a comment in the back channel. The emoji appears visually attached to a previous message, providing feedback without initiating a new turn.

the conversation. This sequential structure differs sharply from the layered character of the back channel, which overlays the conversation without disrupting its order. While the main channel is responsible for moving the interaction forward, the back channel helps manage alignment and feedback.

In sum, the main channel in IM consists of user-generated, turn-taking contributions that appear in the main message thread. These contributions structure the interaction in terms of topical progression and response expectation. The back channel provides a parallel but subordinate layer of interaction that supports the main conversation.

The findings from Experiments I showed that, when temporal and spatial cues are put into competition, readers tend to prioritize spatial distance in resolving pronouns. The findings from Experiment II showed

that when repair is introduced not through message editing function but through quoting function in which the repairing utterance forms a new message, the repair is treated similarly to other utterances on the same layer. These observations point to a layered structure in IM dialogue. The present chapter has argued that IM allows for two structurally different layers of communication: the main channel and the back channel. Building on this account, the next question is whether channel structure itself shapes prominence, that is, whether utterances in the main channel are more likely to guide pronoun interpretation than those in the back channel. This question is addressed in the following chapter through two experiments testing whether pronouns preferentially refer to referents introduced in the main channel rather than those introduced in the back channel.

# Chapter 8

# Studies on Prominence in the Main and Back Channel in Instant Messaging Dialogue

Experiments I and II examined which type of order, spatial or temporal, is more relevant for recency-based prominence in IM dialogue. In Experiment I, self-repair via the message editing function was used to create misalignment between spatial and temporal distance. The results showed that spatial distance had a stronger effect on referent resolution than temporal distance, suggesting that readers rely more heavily on the spatial arrangement of utterances when interpreting pronouns in IM.

To isolate the role of repair itself, Experiment II manipulated the presentation format of the repair. Instead of replacing an earlier message via editing, the repair was displayed as a stand-alone utterance using a quoting function. The results indicated that repairs introduced as separate messages behave similarly to other stand-alone messages with respect to recency-based prominence. In contrast, when a repair was made through the message edit function, it was treated as less prominent. These findings suggest that the way in which utterances are introduced to the discourse and how visibly they appear in the conversation interface may influence

their prominence status.

This led to a new question: why does the same repair utterance behave differently depending on how it is introduced? The different effects pointed toward a deeper structural distinction in the IM dialogue, namely, the separation between the contributions of the main channel and the back channel. Unlike face-to-face conversation, IM introduces platform-specific features that allow users to modify or react to messages without generating a new sequential turn. These features add a layered structure to communication in IM, giving rise to what this dissertation terms the back channel.

As defined in Chapter 7, back channel activity in IM is characterized by the following features: it is initiated by the participant, and it does not enter the spatial sequence of the main channel but appears in a layered position attached to earlier content. Within this framework, repairs made using the edit function are classified as part of the back channel, while repairs presented as new stand-alone messages, for instance through quoting functions, are treated as contributions to the main channel.

This chapter presents two experiments that investigate how main vs. back channel contributions affect referent prominence in IM dialogue and whether emoji comments, as a form of back channel feedback, influence the resolution of ambiguous pronouns.

## 8.1 Research Question II

This is second main research questions for this dissertation:

> RQ II: *Does the channel in which an utterance appears (main channel vs. back channel) influence how likely the referent of a nominal expression in that utterance is to be referred to by subsequent pronouns in IM dialogue?*

Let me also illustrate this with an example. Consider the following exchange in Figure 8.1, with English translation provided in (47):


Figure 8.1: Example: Short exchange between Lea and Freya

(47) a. *Lea: Hello, how are you?*
b. *Freya: Greta sliced the ham.*
c. *Freya: After that, she thawed the bacon.*

In this short exchange, Lea initiates the conversation, and Freya responds by narrating what Greta has been doing: first slicing the ham and then thawing the bacon. Now, imagine that Freya realizes that the first message contains an error: Greta had actually sliced the roast, not the ham. Freya has two options: either edit her earlier message using the message editing function, as in Figure 8.2, or she can use the quote function to send a new message quoting the corrected utterance, as in Figure 8.3.


Figure 8.2: Example: Freya edits an utterance using the message editing function.


Figure 8.3: Example: Freya modifies an utterance using the quote function.

The reparing utterance remains the same, but the form it is introduced into the discourse is different. In Figure 8.2, according to the proposed definition, it appear in the *back channel* as an edit to a previously sent message. In Figure 8.3, it appear in the *main channel* as a new stand-alone message through a quoting function. Imagine after Lea expressed her comment, Freya continues the storytelling using a nonve verb: “Außerdem dorpelt sie ihn”. How, in this case, would the pronoun be resolved?

- **H4**: When a repair in the main channel introduces a nominal expression, its referent is more likely to be referred to by a subsequent pronoun than when the same repair is presented in the back channel.

Likewise, in a similar situation, where Lea initiates the dialogue and Freya tells her what Greta has been doing (see Figure 8.1). This time, Lea wants to express her reaction. She is not happy with the fact that Freya sliced the ham. In this case, she has two options for expressing her feelings: she can either use the emoji reaction function (Figure 8.4) or send a new message quoting the utterance she wants to comment on (Figure 8.5).

Figure 8.4 illustrates an emoji reaction presented in the back channel, while Figure 8.5 shows the same reaction in the main channel. After Lea’s comment, Freya continues the story using a nonce verb: *Außerdem dorpelt sie ihn*. The question then arises: how would the pronoun be resolved in this case? This leads to another hypothesis:

- **H5**: When an emoji comment is presented in the main channel, the referent of a nominal expression in the utterance it comments on is more likely to be referred to by a subsequent pronoun than when the same emoji comment is presented in the back channel.


Figure 8.4: Example: Freya comments on an utterance using the quote function.


Figure 8.5: Example: Freya comments on an utterance using the comment with emoji function

## 8.2 Experiment III: Repair and Pronoun Resolution Across Channels

### 8.2.1 Aim and Hypothesis

The aim of Experiment III is to test whether the presentation of a repair, either as a separate message in the main channel or as an integrated element in the back channel, affects its impact on referent prominence. This question speaks to how specific design features of chat interfaces shape discourse processing. For clarity, the hypothesis is restated below:

- **H4**: When a repair in the main channel introduces a nominal expression, its referent is more likely to be referred to by a subsequent pronoun than when the same repair is presented in the back channel.

### 8.2.2 Method

#### Materials

The materials for Experiment III were adopted from the stimulus sets used in Experiments I and II. An example is provided in (48), where *den Schinken* ("the ham") is corrected to *den Braten* ("the roast"). Target-U contained the masculine pronoun *ihn* ("it"), which could refer to any of the referents introduced. Participants were asked to resolve the pronoun in a multiple-choice question, with each option corresponding to one of the potential antecedents.

(48) a. Lea: Hallo, wie geht's? (Opening question)
b. Freya: Greta hat den Schinken aufgeschnitten. (Context-U1)
c. Freya: Danach hat sie den Speck aufgetaut. (Context-U2)
d. Freya: Greta hat den Braten aufgeschnitten. (Repair-U)
e. Freya: Außerdem dorpelt sie ihn. (Target-U)

f. *FRAGE:* Was dorpelt Greta?
A: Schinken B: Speck C: Braten

(49) a. *Lea: Hello, how are you?* *(Opening question)*
b. *Freya: Greta sliced the ham.* *(Context-U1)*
c. *Freya: After that, she thawed the bacon.* *(Context-U2)*
d. *Freya: Greta sliced the roast.* *(Repair-U)*
e. *Freya: Besides, she is 'dorpeling' it.* *(Target-U)*
f. *QUESTION: What is Greta 'dorpeling'?*
*A: Ham B: Bacon C: Roast*

The stimuli followed the same four-part structure as in the first two experiments. Context-U1 and Context-U2 were perfect-tense utterances connected by a coordinating relation; they shared the same human subject but each introduced a different masculine accusative noun phrase. Repair-U functioned as a correction of Context-U1, replacing its object with a new, contextually plausible masculine object. Target-U appeared in the present tense and contained a subject pronoun consistent with the subjects in the context utterances, as well as the ambiguous pronoun *ihn*, which served as the critical referring expression. Filler items from Experiment I were also reused.

**Participants**

Data were collected from 35 participants. 5 participants were excluded from analysis due to failing a concentration check, resulting in a final sample of 30 participants (16 male, 14 female). The participants had a mean age of 34.5 years, with an age range of 20 to 64 years. All participants were native speakers of German, recruited through the Prolific platform. Prolific's pre-screening system was used to ensure that they had no history of language or cognitive disabilities.

### Design

Experiment III followed a similar within-subjects Latin Square design as in Experiment I and II. Each participant saw one version of each item, and the distribution of the two conditions: *Main Channel* and *Back Channel* was counterbalanced across participants.

Both experimental conditions included the same sequence of utterances: An opening question from one speaker and the other speaker replies by producing a series of utterances: Context-U1 (initial referent introduction), Context-U2 (second referent introduction), and Repair-U (repair of Context-U1). However, the presentation format of Repair-U varied across conditions. In the **Main Channel** condition, Repair-U appeared as a stand-alone message using the quoting function found in IM interfaces, clearly marked as a new contribution. In the **Back Channel** condition, Repair-U was displayed using the message edit function, visually replacing Context-U1 that it corrected, thereby creating a layered presentation consistent with the IM back channel.

### Procedure

Similar to Experiments I and II, this experiment was conducted online using Ibex Farm (Drummond, 2023). To become familiar with the interface and task, participants completed three practice trials before starting the main experiment.

In each trial, the dialogue began with an opening message from Speaker A (e.g., Lea, see Figure 8.6a), followed by two utterances (Context-U1 and Context-U2) from Speaker B (e.g., Freya, see Figures 8.6c and 8.7b), each introducing one referent. A dynamic typing indicator (Figure 8.6b) was displayed before each message to simulate real-time message construction. The duration of the typing indicator and reading times were based on the same rates used in Experiment II: 61 WPM for typing speed and 200 WPM for reading speed. These rates were retained because only one participant

in Experiment II reported that the waiting times were too long, compared to four participants in Experiment I.

The manipulation occurred at Repair-U, which contained a repair of the initial referent from Context-U1. The presentation of Repair-U differed by condition: In the **Main Channel** condition, it appeared as a stand-alone message using the quoting feature (Figure 8.8a). In the **Back Channel** condition, it was displayed using the repair function, visually replacing the Context-U1 (Figure 8.9a) and marked with a pencil symbol. This design ensured that Repair-U always followed Context-U2 temporally but differed in whether it appeared in the main channel or in the back channel.

Following Repair-U, the typing indicator appeared again before Target-U was displayed (see Figures 8.8b and 8.9b). Target-U contained a nonce verb and an ambiguous masculine pronoun (*ihn*), serving as the target for the pronoun resolution task (Figures 8.8c and 8.9c). After the dialogue was fully presented, participants had four seconds to review it before proceeding to the multiple-choice question screen (Figure 8.10).


(a) Opening message (b) Typing indicator before Context-U1 (c) Context-U1 appears

Figure 8.6: Exp III: Example screenshots for both conditions (1-3)


(a) Typing indicator before Context-U2 (b) Context-U2 appears (c) Typing indicator before Repair-U

Figure 8.7: Exp III: Example screenshots for both conditions (4-6)

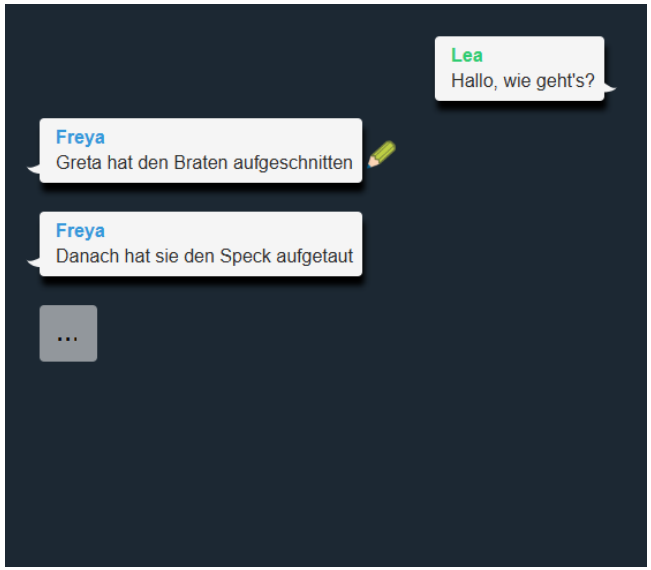

(a) Main Channel: Repair-U appears as a stand-alone repair message quoting Context-U1

(b) Main Channel: Typing indicator before Target-U

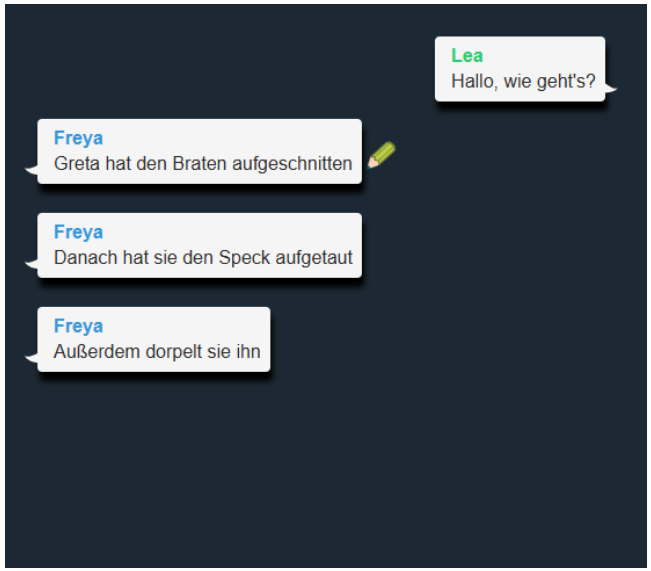
(c) Main Channel: Target-U appears

Figure 8.8: Exp III: Example screenshots from the Main Channel condition (7-9)


(a) Back Channel: Repair-U visually replacing Context-U1




(b) Back Channel: Typing indicator before Target-U




(c) Back Channel: Target-U appears

Figure 8.9: Exp III: Example screenshots from the Back Channel condition (7-9)

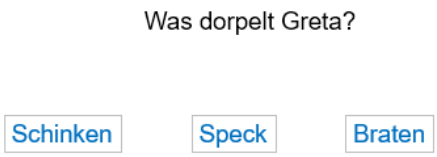



Figure 8.10: Exp III: Multiple-choice question following the dialogue

### 8.2.3 Results

Each condition included 450 trials, yielding a total of 900 responses from 30 participants. Figure 8.11 summarizes the proportional distribution of response types across the two experimental conditions. In the Back Channel condition, participants selected the referent in Repair-U in 18.2% of the trials, while the referent in Context-U2 accounted for 77.1% of responses. In contrast, in the Main Channel condition, participants chose the Repair-U substantially more often (85.3%). These data already suggest that the repaired referent was more prominent in the Main Channel condition.

To investigate whether repair introduced to the main channel lead to more possibilitys of the referent it contains being more likely to be rementioned in the subsequent pronoun than the same repair introduced in the back channel in IM, a Bayesian multinomial logistic regression model was fitted using the `brms` package (Bürkner, 2017) in R (version 4.4.2; R Core Team, 2024) within the RStudio environment (RStudio Team, 2025). The model predicted participants' referential choices: Ref1, Ref2, and RefRepair based on the experimental condition (Main Channel vs. Back Channel). The reference levels were set to Back Channel for Condition and Ref1 for referential choice. Random intercepts were included for participant ID and

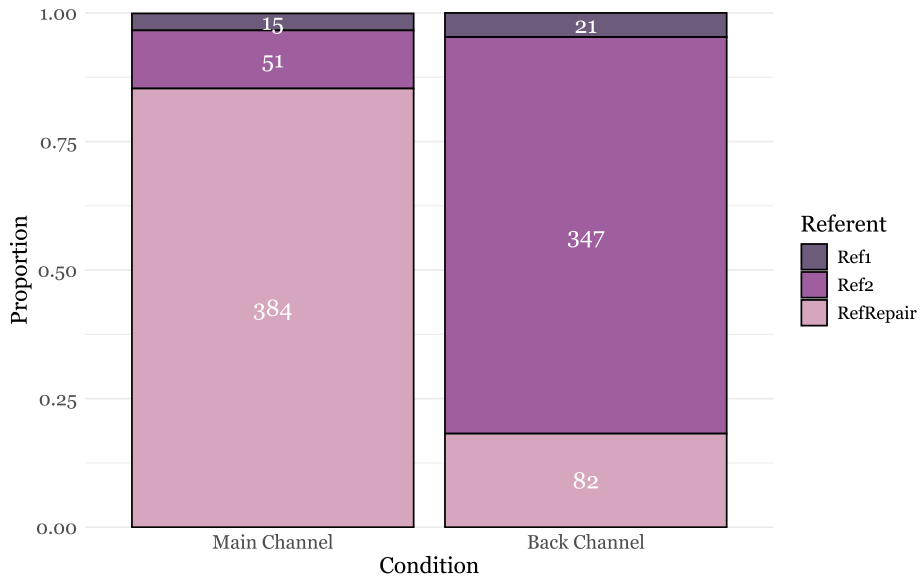



Figure 8.11: Proportional distribution of response types across experimental conditions. Each bar represents the proportion of response types within each condition. Exact response counts are shown above each bar.

item number to account for individual- and item-level variation. Following recommendations for multinominal logistic models (Gelman et al., 2008), weakly informative normal (0,2) prior was used for regression coefficient. Model estimation was carried out using four chains with 4000 iterations each.

Model diagnostics indicated excellent convergence (all R-hat values = 1.00; effective sample sizes > 1000). Posterior predictive checks confirmed that the model accurately captured the observed data distribution. The model returned strong evidence for condition-specific differences in referent choice. Posterior estimates of predicted response probabilities are presented in Figure 8.12.

To test the experimental predictions, a targeted hypothesis was tested using the `hypothesis()` function in `brms` package. H4 predicted that in the Main Channel condition, the repairing referent (RefRepair) would be more likely to be selected as the referent for the ambiguous pronoun “ihn” in the Main Channel condition compared with the Back Channel condition.

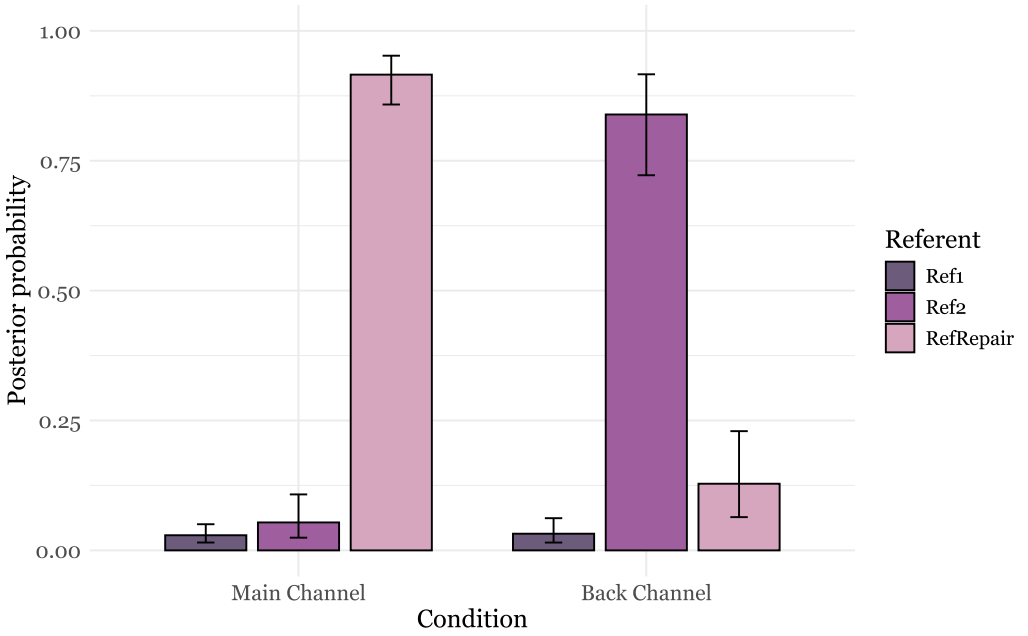



Figure 8.12: Posterior estimates of predicted response probabilities for referent choice across conditions in Experiment III. Estimates are derived from a Bayesian multinomial logistic regression model with `RefRepair` as the reference level for `Reference` and `Back Channel` as the reference level for `Condition`.

The contrast between RefRepair in the Main Channel condition and RefRepair in the Back Channel condition showed a strong preference for the Main Channel condition. The posterior estimate for this effect was 2.07 (Est. Error = 0.38), 90% CI [1.46, 2.70], and posterior probability = 1.00. This provides robust support for H4, indicating that referents introduced in repairs are more prominent when the repair appears in the main channel.

### 8.2.4 Discussion

The results of Experiment III provide strong evidence that channel influences pronoun resolution in IM. Repairs presented as stand-alone messages in the main channel were more likely to have their referent chosen as the antecedent of the ambiguous pronoun, whereas those presented via message editing in the back channel were selected much less frequently. This

supports H4 and indicates greater referent prominence for main channel repairs.

The design of Experiment III also allowed for an observation of the interplay between spatial and temporal recency, as all repair utterances occurred temporally after the introduction of Ref2. In the Back Channel condition, Repair-U was temporally close but remained spatially distant, whereas in the Main Channel condition, the repair was both temporally and spatially close. The predominant preference for Ref2 over RefRepair in the Back Channel condition is consistent with the findings of Experiment I, suggesting that spatial distance has a stronger influence on referent prominence than temporal distance. This further supports the finding that the spatial arrangement of utterances shapes how readers process and interpret discourse in IM.

So far, the evidence comes exclusively from repairs. A natural next question is whether the same main and back channel effect extends to other types of utterances. Experiment IV addresses this by focusing on emoji comments, which can either appear in the back channel as attached reactions or in the main channel as stand-alone messages.

## 8.3 Experiment IV: Emoji Comments and Channel-Based Prominence

### 8.3.1 Aim and Hypothesis

Findings from Experiment III support H4, showing that repairs in the main channel were more prominent than those in the back channel. Experiment IV extends this question to emoji reactions, testing whether their effect on prominence varies by channel. For clarity, the hypothesis is restated below:

- **H5**: When an emoji comment is presented in the main channel, the referent of a nominal expression in the utterance it comments on is more likely to be referred to by a subsequent pronoun than when the same emoji comment is presented in the back channel.

### 8.3.2 Method

**Materials**

Experiment IV tested whether the channel in which an emoji comment appears (main vs. back channel) affects the prominence of the utterance it comments on. Unlike Experiments I–III, which focused on repairs, this experiment used emoji comments as the manipulation. In the Back Channel condition, the emoji comment was attached to the original message using the emoji comment function, whereas in the Main Channel condition, it was presented as a stand-alone message using the quoting function.

The stimuli were modified based on those used in Experiments I–III. Each dialogue consisted of a short exchange between two interlocutors, presented in a simulated IM interface. Speaker A (e.g., Lea) initiated the conversation with an opening greeting (see Example (50-a)). Speaker B (e.g., Freya) then introduced a first referent (Context-U1) and a second referent (Context-U2) in two separate utterances ((50-b) and (50-c)). Subsequently, Speaker A provided an emoji comment on Context-U1. In the Main Channel condition, this emoji comment appeared as a stand-alone message, with Context-U1 quoted in the chat box; in the Back Channel condition, it was

added as a comment to the original message. Speaker B then produced a final utterance (Target-U) containing a pronoun that could refer to the referent from either Context-U1 or Context-U2. Participants were asked to resolve the pronoun by selecting one of two referents in a multiple-choice question.

(50) a. Lea: Hallo, wie geht's? (Opening question)
b. Freya: Greta hat den Schinken aufgeschnitten. (Context-U1)
c. Freya: Danach hat sie den Speck aufgetaut. (Context-U2)
d. Lea: 😠 (Emoji Comment on Context-U1)
e. Freya: Außerdem dorpelt sie ihn. (Target-U)
f. *FRAGE:* Was dorpelt Greta? A: Schinken B: Speck

(51) a. *Lea: Hello, how are you? (Opening question)*
b. *Freya: Greta sliced the ham. (Context-U1)*
c. *Freya: After that, she thawed the bacon. (Context-U2)*
d. *Lea:* 😠 *(Emoji Comment on Context-U1)*
e. *Freya: Besides, she is 'dorpeling' it. (Target-U)*
f. *QUESTION: What does Greta "dorpel"? A: Ham B: Bacon*

Due to the large number of available emojis, it was not feasible to include all emojis. Instead, the selected emojis were based on prior research showing that facial and people emojis are among the most frequently used in digital communication (Muhammad Hasyim, 2019). Additional studies informed the interpretation and usage of emojis in IM contexts (Annamalai and Salam, 2017; Li, 2021; Syahfitri et al., 2021; Fladrich and Imo, 2020). Based on these studies, I narrowed down the range of emojis to smileys and people categories and chose 10 emojis conveying concern, unsatisfied and skepticism, as defined by (Emojipedia, 2025). A complete list of the selected emojis, along with their Unicode names and meanings, is provided in Table 8.1.

| Emoji | Unicode Name | Meaning |
|---|---|---|
| 😲 | Astonished Face | May convey a wide range of emotions, including disbelief, excitement, concern or awe, amazement. |
| 🧐 | Face w. Monocle | May show that someone is pondering, considering, or questioning something |
| 😮 | Face w. Open Mouth | May convey such feelings as awe or disbelief |
| 🤨 | Face w. Raised Eyebrow | Conveys a wide variety of sentiments, including suspicion, skepticism, concern, consideration, disbelief, and disapproval |
| 🤔 | Thinking Face | Often used to question or scorn something or someone, as if saying *Hmm, I don't know about that* |
| 😠 | Angry Face | Conveys varying degrees of anger, from grumpiness and irritation to disgust and outrage |
| 🙄 | Face w. Rolling Eyes | As with the gesture of an eye-roll, commonly conveys moderate disdain, disapproval, frustration, or boredom |
| 🙁 | Slightly Frowning Face | Often conveys a mild degree of concern, disappointment, or sadness |
| 😒 | Unamused Face | May convey a variety of negative emotions, including irritation, displeasure, grumpiness, and skepticism |
| 👎 | Thumbs Down | A thumbs-down gesture indicating disapproval |

Table 8.1: Emojis Used in the Experimental Stimuli with Unicode Names and Descriptions

The reason for selecting emojis that expressed concern and skeptical emotions was guided by research on discourse processing. According to Clark and Schaefer (1989), during conversation, a speaker typically presents an utterance in the presentation phase, after which the addressee provides evidence of either understanding, meaning that they believe they understand the speaker's contribution; or of trouble in understanding, indicating that they did not fully comprehend what was said. When a listener signals trouble, the speaker is expected to repair or clarify the utterance (Clark and Schaefer, 1989; Schegloff et al., 1977; Dingemanse et al., 2015).

In this experiment, therefore, it is expected that the speaker would elaborate more on the utterance that was commented on with an emoji, which is Context-U1. By using emojis that convey trouble signals or negative emotions, the design aims to increase the likelihood that participants interpret Context-U1 as requiring elaboration, making it more likely to be referred to by the ambiguous pronoun in Target-U. If, despite this manipulation, participants in the Back Channel condition still prefer Context-U2 as the referent, this would suggest that the difference in channel influences referent prominence.

The visual appearance of an emoji can differ across platforms. The same emoji may look quite different depending on the platform or software, which could affect its interpretation. For example, the "Grimacing Face" emoji on Apple devices shows tightly clenched, individually outlined teeth and closely set eyes 😬; on Microsoft PCs, the mouth appears much wider, with two thick white rows separated by a red space 😬; on WhatsApp, the design shows more robotic eyes and a simplified mouth without individual teeth outlines 😬. The experiment used the Google Noto Color Emoji (Version: Android 11.0) set rather than relying on Unicode, which allows for platform-specific designs. This choice ensured that all participants viewed the same emoji design, as presented in Table 8.1. An exhaustive list of the stimuli is provided in Appendix C.

To reduce predictability and maintain participant engagement, the

experiment included 30 filler items designed specifically for this study. These fillers varied in conversational structure, including differences in turn-taking, comment channels, and spatial and temporal alignment or misalignment. Emojis in the fillers consistently expressed positive emotions, signaling agreement or understanding, and were appropriate for each given context. The 10 emojis used in the fillers included: Face Blowing a Kiss 😘, Grinning Squinting Face 😆, Smiling Face with Heart-Eyes 😍, Slightly Smiling Face 🙂, Smiling Face with Smiling Eyes 😊, Face with Tears of Joy 😂, OK Hand 👌, Thumbs Up 👍, Red Heart ❤️, and Clapping Hands 👏.

**Participants**

A total of 38 participants were recruited for the study. After applying a concentration check, 2 participants were excluded from the analysis, resulting in a final dataset of 36 participants (18 female, 17 male, and 1 diverse). The participants ranged in age from 22 to 72 years, with an average age of 41.7 years. All were native speakers of German and were recruited through the Prolific platform. Prolific's pre-screening system was used to ensure that they had no history of language or cognitive disabilities.

**Design**

Experiment IV followed a within-subjects Latin Square design. The experiment included two conditions: Main Channel and Back Channel. In the **Main Channel** condition, emoji comments were presented as stand-alone messages quoting the message the emoji gave a comment to, ensuring that the comment appeared as a new, separate contribution in the dialogue. In the **Back Channel** condition, the same emoji comments appeared directly on the original message. Importantly, the emoji itself is identical in size in both conditions; only the channel it appears in differs, either as an independent message in the main channel or as an attached reaction in the back channel.

The dependent variable was participants' responses to a multiple-choice question following each dialogue, in which they identified the referent of the ambiguous pronoun in Target-U. The response options corresponded to the potential antecedents introduced in the dialogue: the referents of Context-U1 and Context-U2.

**Procedure**

Similar to the previous experiments, this study was conducted online using the Ibex Farm platform (Drummond, 2023). To familiarize participants with the experimental interface and task requirements, each participant first completed three practice trials. The main experiment then proceeded with 30 experimental dialogues and 30 filler items, presented in a randomized order to control for potential order effects.

In each trial, the dialogue began with an opening message from Speaker A (e.g., Lea) (see Figure 8.13a), followed by two utterances from Speaker B (e.g., Freya), labeled Context-U1 and Context-U2 (see Figures 8.13c and 8.14b). Each of these utterances introduced a new referent, serving as a potential antecedent for the pronoun in the final utterance (Target-U). To simulate the dynamic nature of real-time IM conversation, a typing indicator (three looping dots; see Figures 8.13b and 8.14a) was displayed before each message. Typing and reading times were calculated individually for each utterance based on its length, using the same rates established in previous experiments: a typing speed of 61 WPM and a reading speed of 200 WPM.

In both conditions, the emoji comment was produced by Speaker A. After a delay of 1.5 seconds following Context-U2 (to mimic the natural pause for turn-taking in IM), a typing indicator appeared on Speaker A's side and remained visible for the duration of the reading time of Context-U2. This simulated the scenario where Speaker A, having read Context-U1 and Context-U2 and seeing no additional input from Speaker B, chose to comment on Context-U1 with an emoji.

In the **Main Channel** condition, the emoji comment was displayed as a separate, stand-alone message using the quoting function (see Figure 8.16a). This format made the comment visually and structurally independent from the original message.

In the **Back Channel** condition, the same emoji comment was presented as an integrated comment within the original message (see Figure 8.15a). The emoji was displayed at the bottom-right corner of Context-U1's chat box, in a green circle that matched the speaker A's name color. This design visually linked the emoji comment to the original message, signaling that the comment was from Speaker A. None of the participants reported having any difficulty understanding this format.

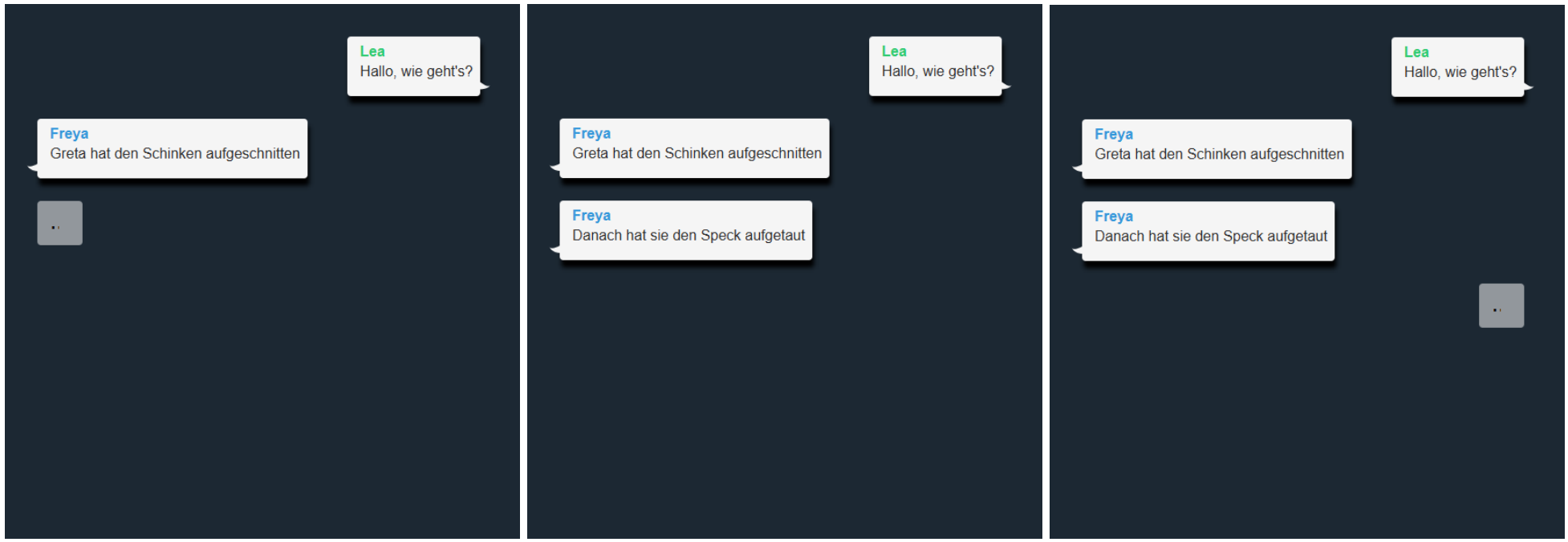
This manipulation ensured that the emoji comment always followed Context-U2 temporally but differed in spatial and structural presentation across conditions.

After the emoji comment, a typing indicator appeared on Speaker B's side, indicating that Speaker B was composing a message. Subsequently, Target-U was displayed, containing a nonce verb and an ambiguous masculine pronoun (*ihn*), which served as the target for the pronoun resolution task (see Figures 8.16c and 8.15c). After the entire dialogue was shown, participants had four seconds to review it before proceeding to a separate screen with the multiple-choice question (see Figure 8.17). Participants were asked to select one of the two referents as the antecedent of the pronoun.

As in the previous experiments, half of the 30 experimental dialogues featured a reversed speaker arrangement to balance potential biases. In these dialogues, Speaker B initiated the conversation, and Speaker A contributed Context-U1 and Context-U2 before Speaker B commented on Context-U1 with an emoji. In the Back Channel condition of these reversed dialogues, the emoji comment was attached to the bottom-left corner of Context-U1's chat box, with a colored circle matching Speaker B's name color (blue) to ensure participants could easily identify the source of the comment (see Figure 8.18 for Back Channel condition and Figure 8.19 for Main Channel condition).

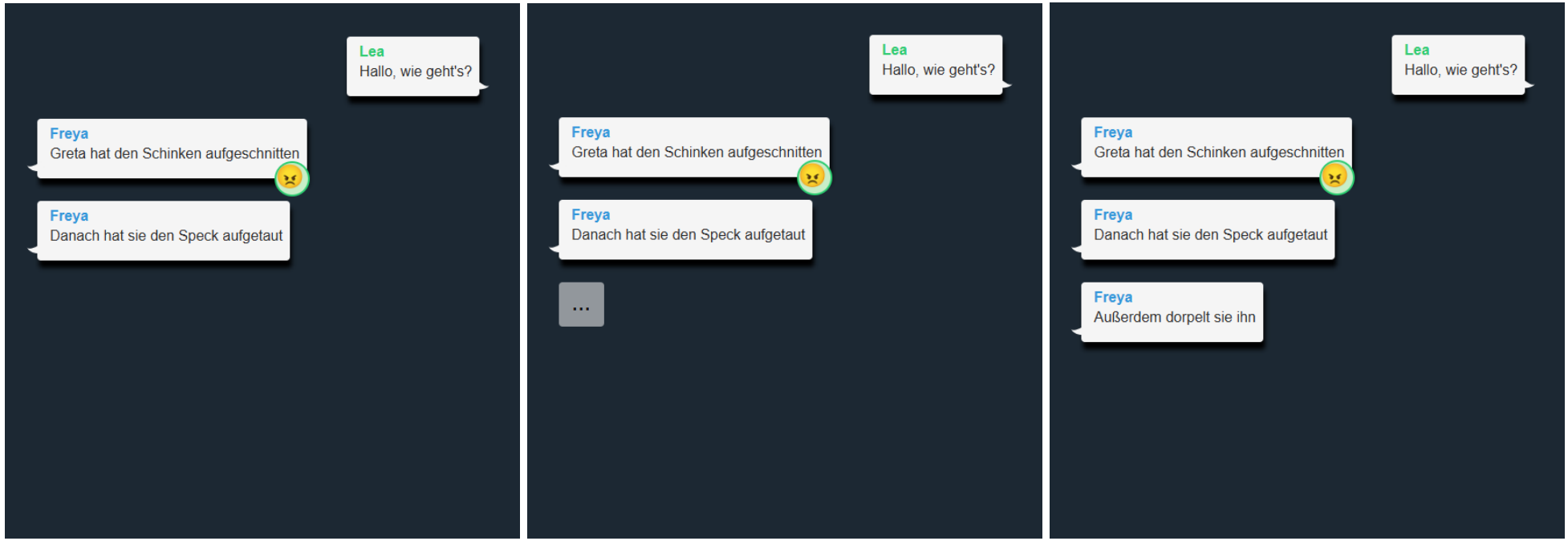

(a) Opening message (b) Typing indicator before Context-U1 (c) Context-U1 appears

Figure 8.13: Exp IV: Example screenshots for both conditions (1-3)




(a) Typing indicator before Context-U2 (b) Context-U2 appears (c) Typing indicator from speaker A

Figure 8.14: Exp IV: Example screenshots for both conditions (4-6)

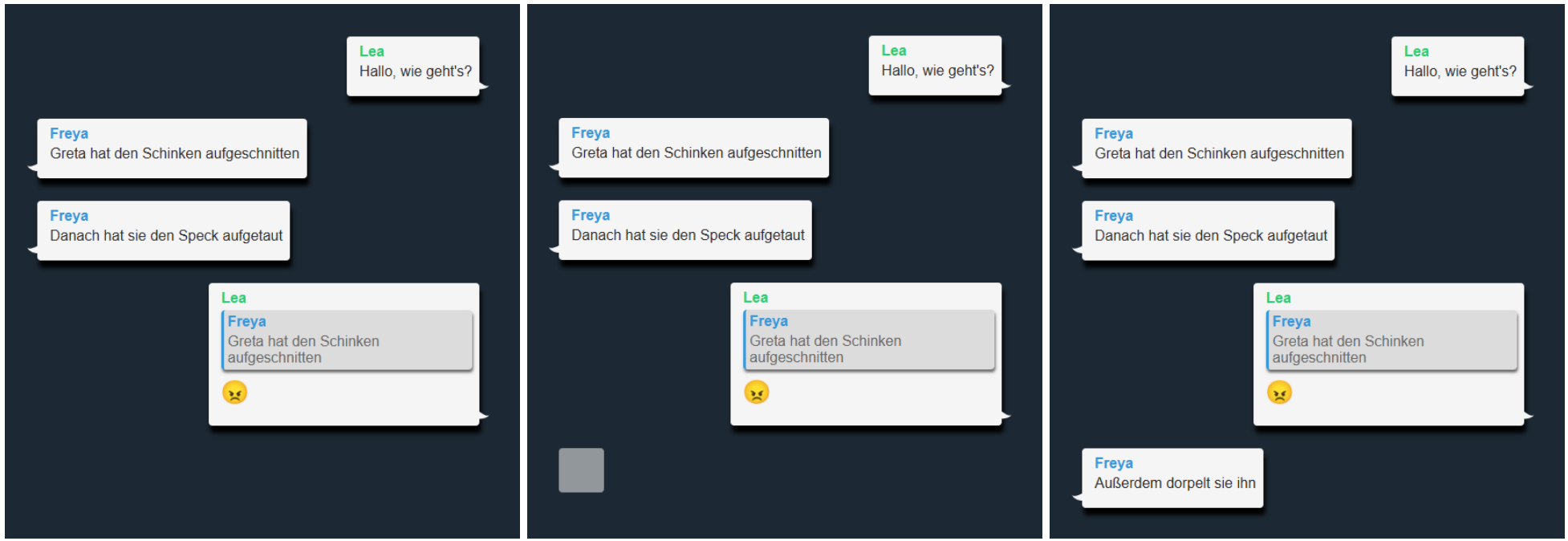



(a) Back Channel: Speaker A comments on Context-U1

(b) Typing indicator before Target-U

(c) Target-U appears

Figure 8.15: Exp IV: Example screenshots from Back Channel condition (7-9)




(a) Main Channel: Emoji comment appears as a stand-alone message quoting Context-U1

(b) Main Channel: Typing indicator before Target-U

(c) Main Channel: Target-U appears

Figure 8.16: Exp IV: Example screenshots from Main Channel condition (7-9)


Figure 8.17: Exp IV: For all conditions: Multiple-choice question following the dialogue

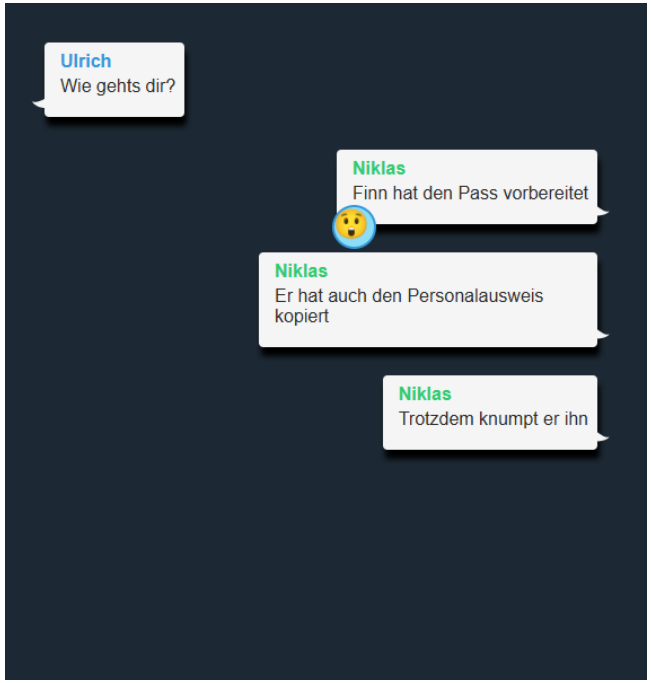



Figure 8.18: Exp IV: Back Channel: Example of the reversed speaker arrangement

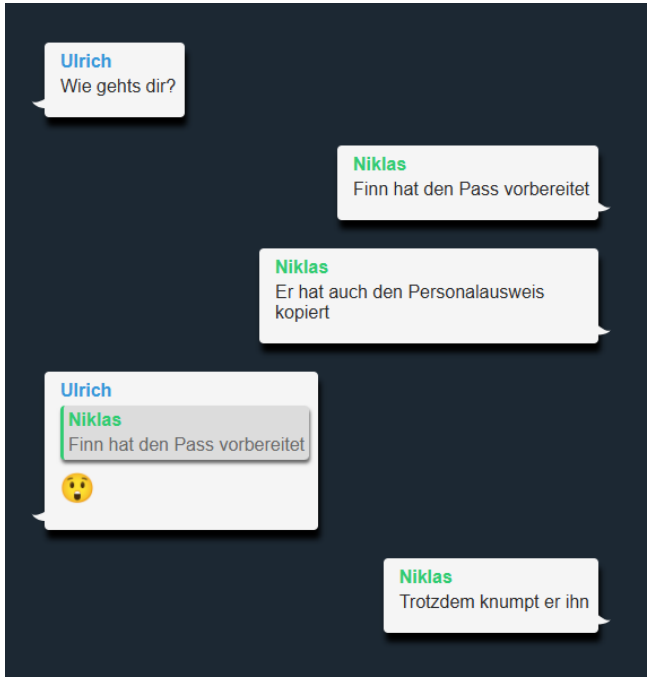



Figure 8.19: Exp IV: Main Channel: Example of the reversed speaker arrangement

### 8.3.3 Results

Each experimental condition comprised 540 trials, resulting in a total of 1080 responses from 36 participants. Figure 8.20 presents the proportional distribution of referent selections across the two conditions. In the Back Channel condition, participants selected Ref1 (the referent with the emoji comment) in 27% of the trials and Ref2 in 73%. In the Main Channel condition, Ref1 was selected in 45% of the trials, with Ref2 accounting for the remaining 55%. These results suggest that participants were more likely to select Ref1 as the referent when the emoji comment was presented as a separate, stand-alone message in the main channel than when it was embedded in the back channel.

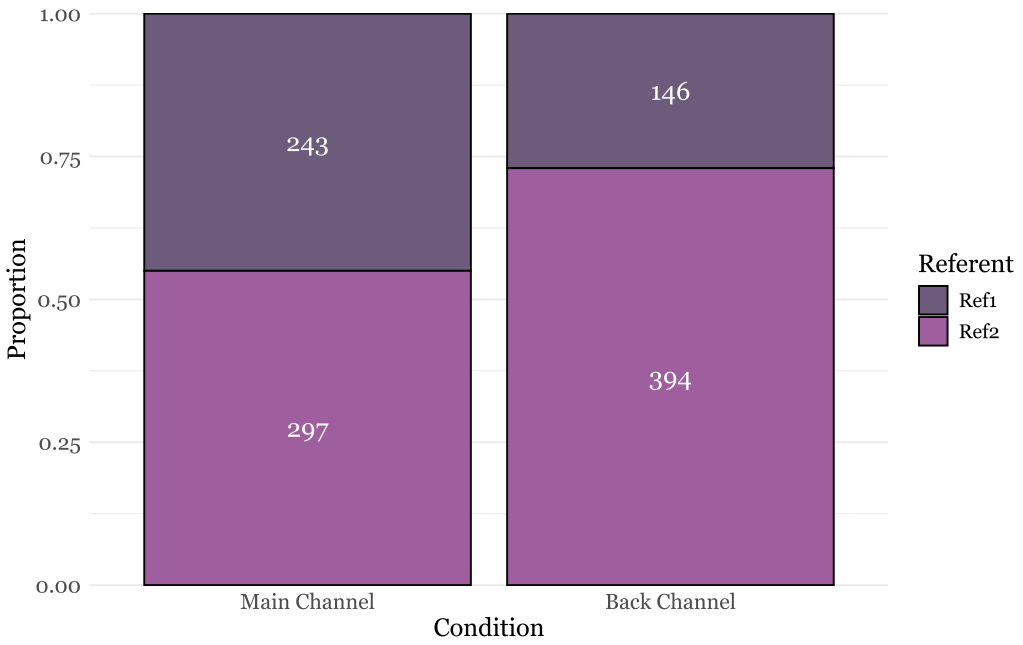



Figure 8.20: Proportional distribution of Ref1 and Ref2 selections across the two conditions. Each bar represents the proportion of referent selections within each condition. Exact response counts are shown above each bar.

To investigate whether emoji reactions presented in different channels in IM influence whether the referent introduced in the utterance it comments on is more likely to be referred to by a subsequent pronoun, A Bayesian logistic regression model was fitted in R (version 4.4.2; R Core Team, 2024)

using the `brms` package (Bürkner, 2017) within the RStudio environment (RStudio Team, 2025). The model predicted participants' referential choices: Ref1 (the referent of the nominal expression introduced in Context-U1, which is commented on with emoji) and Ref2 (the referent of the nominal expression introduced in Context-U2) based on the experimental conditions (Main Channel and Back Channel). The reference levels were set to Back Channel for Condition and Ref1 for referential choice. Random intercepts were specified for both participant ID and item number to account for variability across participants and items. Following the general approach used in the previous experiments, weakly informative priors were used: a normal(0, 2) prior was placed on the fixed effect coefficient and a Student-t(3, 0, 2.5) prior on the standard deviations of the random effects. The model was estimated using four MCMC chains with 4000 iterations per chain.

Convergence diagnostics indicated that the model fit the data well, with all R-hat values at 1.00 and effective sample sizes exceeding 1000. Posterior estimates of predicted response probabilities for referent choice are presented in Figure 8.21.

Directional hypothesis was tested using the `hypothesis()` function in `brms`. The hypothesis evaluated whether participants were more likely to select Ref1 in the Main Channel condition compared to the Back Channel condition, as predicted by H5.

It was hypothesized that the probability of selecting Ref1 would be higher in the Main Channel condition compared to the Back Channel condition. This prediction was confirmed by a hypothesis test on the model coefficient for the Main Channel condition. The estimated coefficient was -1.08 (90% Credible Interval: [-1.34, -0.82]), with a posterior probability of 1.00 for the hypothesis that ConditionMC < 0. This indicates an increase in the likelihood of choosing Ref1 under the Main Channel condition.

In addition to testing the main hypothesis, I also examined whether participants displayed an overall preference for Ref2 within each condition. Descriptive statistics suggested that Ref2 was chosen more frequently than Ref1 both in the BC condition (73% vs. 27%) and, to a lesser extent, in the

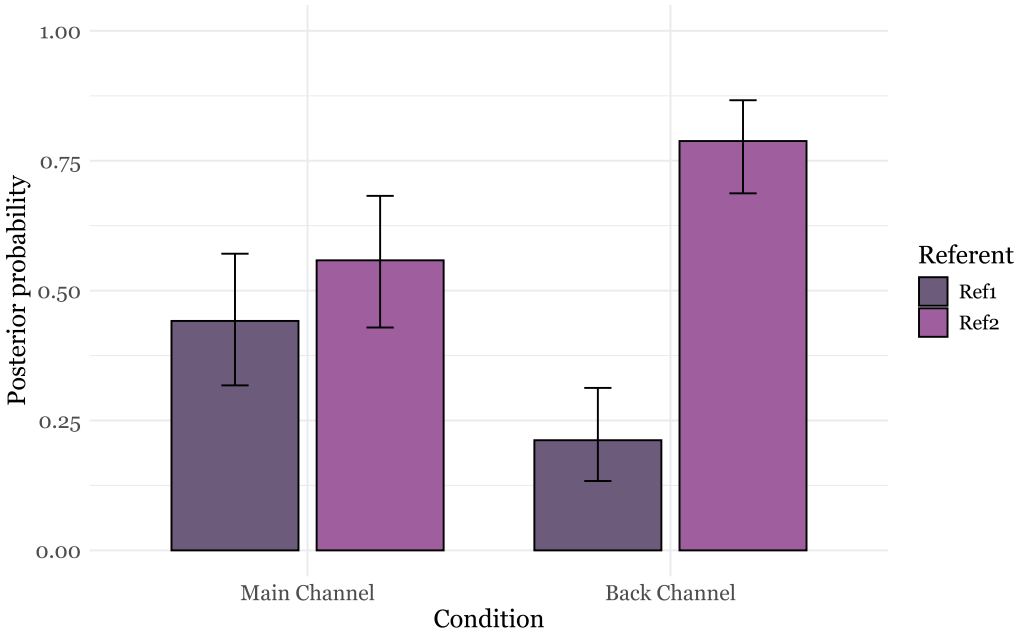



Figure 8.21: Posterior estimates of predicted response probabilities for referent choice across conditions in Experiment IV. Estimates are derived from a Bayesian logistic regression model with Ref2 and Back Channel condition serve as the reference level.

MC condition (55% vs. 45%).

Model-based exploratory tests confirmed this pattern. In the BC condition, the model strongly supported a preference for Ref2: the posterior estimate of the log odds for Ref2 (relative to Ref1) was 1.32 (90% CI [0.87, 1.78]). The posterior probability that Ref2 was more likely than Ref1 was effectively 1.00. This indicates a high likelihood of selecting Ref2 over Ref1 in the Back Channel condition.

In the Main Channel condition, the model also indicated a tendency toward Ref2, although the evidence was weaker. The estimated log odds for Ref2 was 0.24 (90% CI [−0.20, 0.69]). The posterior probability that Ref2 was more likely than Ref1 was 0.81, suggesting that while a descriptive preference exists, the data does not provide strong evidence of a statistically robust preference for Ref2 in the Main Channel condition.

### 8.3.4 Discussion

Experiment IV tested whether emoji comments presented as stand-alone messages in the main channel or integrated within earlier messages in the back channel affect the prominence of their target referents in IM dialogue. The results supported H5: participants were more likely to select Ref1 as the antecedent of the ambiguous pronoun when the emoji comment appeared in the main channel than in the back channel. This finding further supports the view that the main/back channel distinction shapes prominence in IM dialogue.

One explanation for this preference is that presenting the emoji comment as a stand-alone message in the main channel visually marks it as a new contribution. This likely draws more attention and emphasizes the comment's role as an evaluation, thereby increasing the prominence of the referent in Context-U1. By contrast, when the same emoji comment is displayed in the back channel as a reaction attached to the original message, it is visually tied to earlier content rather than occupying its own place in the message sequence. This reduces its impact on the reader's referent choices.

A feature of this experiment was the use of emojis expressing skepticism or disapproval, such as 🙄 and 🤨. Unlike backchannels that signal agreement or understanding (e.g., "uh-huh," "okay"), these emojis are more comparable to verbal feedback such as "huh?" or "hm?", which suggest confusion or disagreement.

In spoken conversation, such signals typically prompt clarification or elaboration from the interlocutor. However, when placed in the back channel in this experiment, these emojis appeared to lose this pragmatic force: despite their negative emotional tone, they had limited influence on how readers processed and selected referents. This suggests that although negative feedback in spoken interaction often prompts clarification or elaboration, the back channel placement of these emojis reduced their ability to shift readers' focus in the dialogue.

This finding is consistent with results from Experiment III, which showed that repairs presented in the main channel were more prominent

than repairs in the back channel. Together, these results suggest that contributions in the main channel, whether repairs or evaluative comments, are consistently treated as more prominent than those integrated into the back channel, regardless of content or emotional tone.

## 8.4 General Discussion

This chapter investigated how the channel in which utterances appear in IM dialogue affects referent prominence. Drawing on two experiments, the results provide consistent evidence that utterances presented as stand-alone contributions in the main channel lead to greater referent prominence than identical utterances integrated in the back channel.

A factor influencing this finding is the visual prominence of main channel messages in IM. Utterances in the main channel are visually marked as new contributions that may attract greater attention and shape subsequent interpretation. By contrast, back-channel elements are presented as attachments to earlier messages and do not constitute new turns. As a result, they are more likely to be interpreted as secondary or supportive, even when they introduce new referents (e.g., via message editing). This pattern is consistent with research on spoken dialogue, where back-channel responses typically do not claim the conversational floor but function as feedback or acknowledgment (Yngve, 1970; Schegloff et al., 1977).

In both experiments, the main channel condition ensured that the relevant discourse element (a repair in Experiment III, an emoji comment in Experiment IV) was both spatially and temporally recent. However, participants' referent choices differed: in Experiment III, participants preferred the referent introduced in the repair (Repair-U), while in Experiment IV, they often selected Ref2, even though Ref1 was made accessible again.

One possible explanation is that repairs inherently introduce new information by clarifying or correcting an earlier statement. As such, they may redirect attention to the corrected referent. The new referent, introduced explicitly in the repair utterance, becomes a focal point for

subsequent interpretation. By contrast, the emoji comments in Experiment IV did not introduce new referents; instead, they functioned as evaluations or feedback on a previously mentioned referent. Even though these emoji comments were presented in the main channel and thus appeared spatially and temporally recent, their communicative function might not have been strong enough to shift the narrative focus away from the most recent event.

A crucial factor distinguishing my findings from traditional spoken dialogue research is the role of speaker initiative. In spoken conversation, back channel utterances are typically initiated by the person who does not have the conversational floor. They serve as acknowledgments or continuers, offering supportive feedback without claiming the main speaking turn. In IM, however, experiments show that messages in the back channel can be initiated by both dialogue participants.

This distinction is key to understanding the different outcomes in Experiments III and IV. The repair in Experiment III, by its very nature, is an act of self-correction by the original speaker. This act of correcting one's own message may naturally shift attention to the correcting referent. By contrast, the emoji comments in Experiment IV, were not self-corrections but rather an evaluative response to a previous message by another chat participant, thus having a limited effect on the narrative plan of the speaker.

In sum, Experiments III and IV show that where an utterance appears in IM dialogue strongly affects referent prominence: Both experiments showed that utterances presented in the main channel are more prominent than the same utterance presented in the back channel. Experiment IV further indicates that although emoji comments presented as stand-alone messages can increase the prominence of their target referents, their evaluative nature limits their ability to override the ongoing discourse structure. Together, these findings show that both how an utterance is presented and what function it serves influence how readers interpret referents in digital communication.

# Chapter 9

# General Discussion and Conclusion

## 9.1 Research Questions Revisited

This research was motivated by an observation from everyday communication: in online chats, including private and group conversations, multiple topics are often addressed simultaneously. Participants frequently refer to people or things using pronouns or abbreviated expressions, sometimes leading to confusion or misunderstandings about what or whom they mean. My primary interest, therefore, was to understand what makes certain entities or utterances stand out and become more prominent in IM conversations, making them more likely to be easily referred to and understood by others.

IM dialogue, with its distinctive technical features, multimodal resources, and layered conversational structure, poses challenges not typically seen in other modes of communication. This motivated to my overarching question:

> How is prominence established and maintained in IM dialogue, and what features make a referent more likely to be interpreted as prominent?

In traditional spoken or written discourse, prominence is shaped by a variety of factors, including syntactic and grammatical features (Keenan and Dryer, 2007; Gundel et al., 1993; Gernsbacher and Hargreaves, 1988; Lambrecht, 1994; Prince, 1978; Thompson, 1987), as well as discourse and cognitive aspects such as information structure and rhetorical relations (Mann and Thompson, 1988; Lambrecht, 1994; Prince, 1981; Krifka, 2008). Concepts such as topicality, accessibility, and givenness have been central to understanding how antecedents are selected in language processing, with prominence often considered a critical factor that increases the likelihood of successful reference.

Among the many factors influencing prominence, I was particularly interested in the role of *recency*. In spoken discourse, recency is typically a temporal phenomenon: the number of clauses (or elapsed time) from the last occurrence in the preceding discourse (Givón, 1983). In written texts, although editing is possible before sending or publishing, what the reader encounters is usually a fixed, linear sequence. For the recipient, the most recent utterance they read is what appears last in the text and is generally the most prominent utterance.

However, IM communication introduces a range of new challenges and opportunities. Users can edit messages they have already sent, and in many platforms, the edited message appears in its original place in the sequence. This creates a misalignment: temporally, the edited message is the most recent communicative act, but spatially, it may appear higher up on the screen, not at the bottom where new messages are typically added. Similarly, IM platforms also allow for comments on earlier messages with emojis, which are feedback utterances that can be temporally close but spatially linked to previous utterances and do not introduce new turns in the chat flow. These technical features mean that recency in IM dialogue is not purely a matter of temporal order but also involves spatial arrangement.

**Temporal and Spatial Recency: Redefining Prominence Cues**

This led me to rethink the concept of recency within the context of IM. Unlike face-to-face conversation or phone calls, where speech is produced and perceived incrementally, in IM, it is necessary to distinguish between the temporal order of production and perception.

> The **temporal order of production** is the order in which utterances are *produced* by speakers. It is determined by the moment an utterance is sent to the system. For multi-utterance messages, the internal order follows the textual sequence of utterances within that message.
>
> The **temporal order of perception** is the order in which utterances are *perceived* by an addressee. This is the sequence determined by the moment an utterance is first taken in on screen.

Based on this, for the purpose of this dissertation, I defined **temporal distance** in IM as:

> The number of intervening utterances between the introduction of a referent and its subsequent mention, calculated according to the temporal order of perception. A referent is considered *temporally close* when it is mentioned in the most recent utterance, and *temporally distant* when several intervening utterances separate its introduction from the current point of reference.

On the other hand, **spatial distance** is understood as:

> The number of intervening utterances between the introduction of a referent and its subsequent mention, measured according to the spatial order of utterances. A referent is considered *spatially close* when it appears in the most recent position of the interface, and *spatially distant* when several intervening utterances separate it from the current point of reference.

This distinction led to an empirical question: In cases where spatial and temporal cues point to different referents, which type of distance determines referent prominence? This led to the first research question:

> *RQI: Which has a stronger influence on pronoun resolution in IM dialogue: spatial distance or temporal distance between a referent's introduction and its subsequent mention?*

**Layering and the Main/Back Channel Distinction** Many platforms now allow users to edit their own messages after sending, or to comment with emojis that are visually attached to specific messages. Depending on the user's preference, these repairs or comments may appear as new messages in the chat, or layered alongside the original one (see Figure 5.2 and 5.3, introduced in Chapter 5, for an illustration of repair as a stand-alone message and as a modification to an existing message). These observations led to a reconceptualization of the IM interface in terms of two layers. The ongoing stream of messages was identified as the main channel, while back channel activity was defined as *participant-initiated actions* that meet two criteria:

- **Do not introduce new turns or topics**, but respond to, modify, or comment on existing ones.
- **Appear in a structurally layered position**, either visually attached to previous messages (e.g., as emoji reactions) or made after the message was already sent (e.g., message edits).

In the second part of this dissertation, I investigated this research question:

> RQ II: *Does the channel in which an utterance appears (main channel vs. back channel) influence how likely the referent of a nominal expression in that utterance is to be referred to by subsequent pronouns in IM dialogue?*

## 9.2 Summary of Experimental Results

### 9.2.1 Hypothesis Overview

Building on the motivations and research questions outlined above, in Experiment I and II, I tested the following hypotheses:

- **H1:** When temporal distance and spatial distance are aligned, the referent of the closer nominal expression is more likely to be referred to by a subsequent pronoun than the referent of the more distant nominal expression.
- **H2**: When temporal and spatial distance are misaligned, pronoun resolution is guided by temporal distance, such that a subsequent pronoun is more likely to refer to the referent of the temporally closer nominal expression.
- **H3:** When a repair is presented as a stand-alone message and introduces a referent, that referent is treated in the same way as referents introduced in other stand-alone utterances in terms of being selected as the antecedent of a subsequent pronoun in IM discourse.

In Experiment III and IV, the following hypotheses were tested:

- **H4:** When a repair in the main channel introduces a nominal expression, its referent is more likely to be referred to by a subsequent pronoun than when the same repair is presented in the back channel.
- **H5:** When an emoji comment is presented in the main channel, the referent of a nominal expression in the utterance it comments on is more likely to be referred to by a subsequent pronoun than when the same emoji comment is presented in the back channel.

### 9.2.2 Summary of Individual Experiment Results

**Experiment I: Spatial and Temporal Distance**

Experiment I investigated whether temporal distance or spatial distance plays a stronger role in guiding reference in IM dialogue. Participants were presented with dynamic chat scenarios in which temporal and spatial distance were either aligned or deliberately misaligned. Misalignment was created using the message editing function, so that an earlier, spatially distant message was updated, making it temporally close.

The results showed that when temporal and spatial distance were aligned, participants reliably preferred the temporally closer referent, consistent with traditional accounts of recency in discourse prominence. This supports H1: when both types of distance are aligned, the closer referent is more prominent than the more distant one. However, in cases of misalignment, participants overwhelmingly selected the referent that was spatially closer to the current utterance rather than the temporally closer one. This contradicts H2 and shows that spatial distance can override temporal cues in determining prominence.

A limitation of Experiment I is that it did not allow the influence of repair to be examined independently. It therefore remains unclear whether the preference for the spatially closer utterance reflects its position or a general reduction in prominence associated with repaired utterances, which would make unrepaired utterances (which were spatially close and temporally distant) the default referents.

**Experiment II: The Effect of Repair on Prominence**

To address the potential influence of repair, Experiment II examined whether repair utterances behave like other utterances in terms of referent prominence. Repair was implemented using a simulated message quoting function, and the conditions differed in the timing of the repair. In the Early Repair condition, the repair appeared immediately after the repaired utterance, whereas in the Late Repair condition, it appeared after an intervening utterance.

If repairs were generally less prominent, they should be disfavored regardless of their spatial position. The results, however, did not support this prediction. When repairs were presented as stand-alone messages, they were favored when spatially closer and disfavored when more distant. This pattern indicates that repair utterances were treated like other utterances in the main chat layer, with prominence determined by spatial distance. This finding supports H3, showing that stand-alone repairs contribute to referent prominence in the same way as other utterances in the discourse sequence.

Taken together, Experiments I and II addressed RQ I concerning the influence of spatial and temporal distance on prominence. The findings show that spatial distance is typically the dominant cue and that the prominence of a repair depends on how it is displayed in the chat interface rather than on its repair status alone.

**Experiment III: Repair in Main and Back Channel**

Building on the earlier findings, Experiment III focused on whether the layering of a repair: in the main channel or the back channel, affected referent prominence. Main channel repair was realized through the quoting function, whereby the repairing utterance quoted the original message and appeared as a stand-alone contribution. Back channel repair was introduced via message editing. The results indicated that it is not repair per se that reduces prominence, but rather in which channel the repair appears. Referents introduced through main channel repairs , were significantly more likely to be selected as the antecedent of a subsequent pronoun, while those introduced through back channel repairs were much less likely to be chosen. These findings supported H4: repairs presented in the main channel lead to greater referent prominence than equivalent repairs presented in the back channel.

**Experiment IV: Emoji Comments in Main and Back Channel**

Experiment IV extended this investigation by testing the effect of emoji comments appearing either as stand-alone messages in the main channel

or attached to earlier messages in the back channel. The results supported H5: emoji comments in the main channel increased the prominence of their target referents compared to the same comments in the back channel. Together, Experiments III and IV address RQ II by demonstrating that whether an utterance appears in the main channel or the back channel systematically influences referent prominence in IM dialogue: main channel utterances are more prominent than those in the back channel.

An additional finding from Experiment IV is that, even in the main channel condition, participants still showed a slight preference for the spatially closer referent, despite attempts to make a more distant referent accessible through quoting and emoji comments. This highlights the persistent influence of spatial distance, even when other prominence-lending cues are present.

### 9.2.3 Key Findings

To summarize, here are the key findings:

- *Spatial distance outweighs temporal distance in IM dialogues.*

In other words, people tended to interpret ambiguous pronouns or referring expressions as pointing to the entity mentioned in the message that was visually closest to the current point in the conversation, rather than to the message that was chronologically most recent.

This effect was most prominent in experimental conditions designed to "mismatch" spatial and temporal cues. When asked to resolve a pronoun, participants generally chose the referent in the bottom-most message: the spatially recent one, even though another message had been modified more recently in time. This preference for spatial recency was robust across experiments, even when temporal recency was visually highlighted (such as with a pencil emoji indicating repair).

- *Temporal closeness does not guarantee prominence in IM.*

Traditional models of discourse suggest that when other conditions being equal, the most recently mentioned entity in time should be the most prominent and thus most likely to serve as the referent for pronouns. However, the results from the current experiments challenge this view in the context of IM. When spatial and temporal recency are aligned, temporal recency still is a pominence lending factor, but it was not always the deciding factor as in other modes of communication such as face-to-face communication.

In particular, when a message had just been edited or received an emoji comment but was not the spatially most recent message in the sequence, its referent was not consistently treated as prominent. This suggests that temporal recency alone, in the form of a recent edit or reaction, is not sufficient to make a referent prominent for pronoun resolution. Rather, it is the spatial position of a message in the interface that more reliably guides readers' attention and shapes prominence assignment.

- *Two types of distances affect prominence: temporal and spatial.*

The findings from Experiment I and II collectively show that recency in IM is not a single, unified phenomenon, but rather consists of two distinct dimensions: temporal distance (the actual time since a message was posted, edited, or commented on) and spatial distance (the message's position within the chat window relative to the current utterance). Both types of distances can make a referent more prominent, but spatial distance is generally the stronger cue.

This distinction is important because it suggests that digital interfaces themselves shape the way people interpret and prioritize information. In IM, where users often scroll back and forth or interact with layered message structures, spatial distance becomes an especially powerful driver of communicative behavior.

- *Utterances in the main channel are more prominent than in the back channel.*

Across experiments III and IV, referents introduced in the main channel were consistently more likely to be selected as the antecedent of a subsequent pronoun. The main channel refers to the primary flow of chat messages, where new turns appear sequentially at the bottom of the screen, while the back channel refers to interactions such as message edits and emoji reactions that are visually attached to earlier messages without changing the ongoing conversational sequence. When a correction or comment was made to a message in the back channel, participants did not update their mental representation of the conversation to the same extent as when the same information was added in the main channel. This effect held true regardless of whether the back channel activity was a message edit or an emoji reaction. This suggests that the structure of IM utterances, not just the content or timing of messages, shapes how users interpret information.

In sum, the experiments presented in this dissertation reveal that referent prominence in IM dialogue is shaped by a combination of spatial, temporal, and channel factors, with spatial distance and the use of stand-alone messages in the main channel playing a stronger role. These findings challenge classic assumptions about how recency operates in conversation and provide a new empirical basis for understanding reference processing in digital communication.

## 9.3 Theoretical Implications

### 9.3.1 Revisiting Recency-Based Prominence

The concept of prominence in discourse has been associated with the accessibility of referents. Earlier research, such as Givón's work on *referntial distance*, proposed clause-based measures of recency to understand referent processing (Givón, 1983). Further experimental approaches, like that from Clark 1979, demonstrated that entities mentioned more recently within a

sentence are parsed more quickly. These findings established recency as a factor influencing referent accessibility. The results from Experiments I and II both confirm and challenge these established perspectives.

In line with classic theories, Experiments I and II found that when spatial and temporal distance are aligned, recency remains a robust cue for prominence: a closer referent is more likely to be selected as the antecedent of a subsequent pronoun than a more distant one. However, when spatial and temporal distance are misaligned, which is a situation made possible by IM's message editing and emoji reaction functions, spatial order overrides temporal distance.

This suggests that, in IM dialogue, physical layout and visual accessibility are more decisive for establishing prominence than the temporal sequence alone. Consequently, models of prominence that treat discourse as a purely linear or temporally ordered phenomenon may be insufficient for digital contexts. These findings highlight the need for revised theoretical accounts that integrate spatial arrangement and platform interface as determinants of prominence alongside temporal, grammatical and other prominence lending cues in digital contexts.

### 9.3.2 The Back Channel and Discourse Organization

The concept of "channel" originates in communication theory, referring to the physical pathway used to transmit signals from sender to receiver (Shannon, 1948). Building on this, Yngve (1970) introduced the notion of the "back channel", describing a parallel conversational layer in face-to-face interaction where listeners offer brief, supportive feedback while the main speaker continues talking. Over time, however, researchers increasingly used the term "backchannels" more narrowly to refer specifically to these short listener responses rather than to the structural layer itself (Schegloff, 1982; Jefferson, 1984).

In IM, if chat participants wish to give verbal backchannels such as "hmm" or "uh-huh," they must actively type and send these responses, causing them to appear as stand-alone messages within the main conversational

flow. Because most IM platforms do not support real-time overlapping speech, the parallel, supportive nature of face-to-face backchannels is not directly replicated. Instead, IM-specific features such as message edits and emoji reactions function as a distinct back channel layer, structurally separate from the main channel.

This result indicates a crucial difference from face-to-face communication. In spoken conversations, backchannels like "hmm" or "okay" typically provide simultaneous feedback without affecting the main speaker's contribution or significantly altering prominence structures. In contrast, IM requires explicit typing and sending of such responses, which necessarily become part of the main conversational sequence. Thus, back channel activity in IM is structurally and functionally distinct not only from backchannels in face-to-face conversation, but also from utterances that perform a similar feedback function in the main channel.

These results call for an update to existing theories of conversational structure, suggesting that digital, multimodal environments require new conceptualizations of discourse that explicitly account for the main/back channel distinction as a central factor in how users process conversations, assign prominence, and track references. The findings also refine classic theories of conversational repair (Schegloff et al., 1977): repair in IM reliably shifts prominence only when it appears in the main channel, while back channel repairs have comparatively little effect. More broadly, this dissertation demonstrates that temporal, spatial, and channel-based factors must be considered together to account for prominence and reference resolution in digital discourse.

### 9.3.3 Emoji as Feedback in IM

A particularly interesting finding from Experiment IV is that even when an emoji comment in the main channel drew attention back to a spatially distant referent, participants still showed a descriptive preference for the referent introduced by the spatially closer message from the main speaker. This suggests that while main channel emoji comments can

partially restore the accessibility of an earlier referent, they do not fully override the prominence of the most recently introduced one.

This pattern resonates with theories of back channel feedback in spoken conversation, where short responses such as “okay”, “yeah”, or “uh-huh” signal that the listener is following along without claiming the floor or changing the topic (Clark and Schaefer, 1989; Clark and Brennan, 1991; Schegloff, 1982; Gardner, 2001). Even when an emoji comment appears in the main channel, it may retain something of this evaluative, non-redirecting character, limiting its capacity to shift referential focus away from the ongoing narrative.

A likely explanation is that emoji comments differ fundamentally from repairs: whereas repairs in Experiment III were initiated by the speaker themselves and thereby integrated into their unfolding discourse plan, the emoji comments in Experiment IV were listener-initiated. Because they did not originate from the speaker currently driving the narrative, readers may have expected the speaker to continue with their original plan rather than reorient to the earlier utterance being commented on. In this sense, emoji comments in the main channel drew attention to the earlier referent but did not override the broader expectation of narrative continuity.

## 9.4 Practical Implications

### 9.4.1 Implications for IM Platform Design

One finding from this research is that spatial distance plays a stronger role than temporal distance in determining how users assign prominence and interpret references in IM conversations. Messages that appear closer to the bottom of the chat interface, the point where new messages occur, tend to be treated as more prominent, even if they are not the most recent in terms of when they were sent or edited. This result has important implications for how IM platforms should manage features such as message edits, withdraw, reactions, and comments.

Currently, most IM platforms implement the message editing function by allowing edited messages to remain in their original spatial position. Depending on the platform, users may edit messages within a limited time frame (e.g., 15 minutes after sending) or no time limit. However, this design often leads to a misalignment between temporal and spatial distance, potentially causing confusion about which message should be considered most relevant at any given moment. For example, if a participant edits a message sent temporally earlier that is no longer visible on the recipient's current screen, the update may go unnoticed entirely. Moreover, although many platforms use a timestamp on messages to indicate when they were originally sent, edited messages are typically only marked with an "edited" tag or a pencil icon, without clarifying when the message was modified. This lack of explicit timestamping for edits further complicates message sequence interpretation and makes discourse processing more challenging for users.

The results from Experiments I and II suggest that participants strongly depend on spatial cues to determine which information is most important or relevant. Therefore, platform designers might consider ways to make edited content more visually prominent. For instance, platforms could highlight edited messages more distinctly, place notifications or reminders at the bottom of the chat window when edits occur, or temporarily place edited messages in a highly visible position near the newest messages. These adjustments would help ensure that important modifications or corrections are not overlooked due to their spatial position.

Similarly, Experiment IV showed that emoji reactions to previous messages, which typically appear in the back channel, have a limited impact on how users interpret references. To address this, IM platforms could enhance the visibility and integration of these reactions into the ongoing conversation. One possible approach would be allowing users to easily post emoji reactions as standalone messages that quote the original message directly, thus clearly placing the feedback into the main channel.

Another strategy could involve a more deliberate and prominent notification system for these reactions. Some platforms like WhatsApp already offer basic notifications for emoji reactions, but another more efficient strategy could be the introduction of a notification system summarizing reactions or comments in the main channel. For example, a small banner stating, "New reactions were added, tap to view", could prompt users to revisit content that might otherwise remain unnoticed.

By incorporating such design adjustments, IM platforms could better align their interface features with the natural cognitive and communicative strategies users employ. These improvements would enhance clarity, reduce confusion, and support coherent digital interactions.

### 9.4.2 Applications in Natural Language Processing

This research also carries implications for computational models of dialogue, particularly in the area of natural language processing (NLP) systems that aim to resolve references, track discourse structure, or summarize chat-based conversations. Many current models are trained on traditional forms of written or spoken dialogue, where messages unfold in a linear sequence over time. However, the findings of this dissertation show that in IM contexts, users rely heavily on visual layout and channels rather than just on when a message was sent.

These findings have direct implications for the design of NLP systems for tasks such as chat summarization, chatbot response generation, and dialogue state tracking. A model that relies solely on timestamp order may fail to identify the most prominent referents in a conversation if it does not account for where messages appear on the screen or whether they belong to the main or back channel. The results of this dissertation suggest that spatial recency is a stronger cue to prominence than temporal recency alone, and that NLP systems aimed at understanding or generating coherent IM dialogue should therefore incorporate features such as spatial positioning and channel distinction into their discourse models.

These findings also support broader efforts to develop conversational AI systems that better align with human patterns of digital communication. Understanding how users interpret feedback, repair, and visual organization in IM can help NLP technologies become more context-aware. As digital messaging develops, NLP models that involve both linguistic content and the distinction between main and back channel will be essential for more natural and accurate machine interpretation of dialogue.

## 9.5 Methodological Considerations

The experiments reported in this dissertation used controlled manipulations to directly test how spatial and temporal distance, and the distinction between main and back channel influence referent choice in IM dialogue. One of the main strengths of the experiments lies in the use of dynamic chat stimuli that closely simulate the real-time structure of IM conversations. While many existing chat corpora provide message text along with metadata such as timestamps and sender identity, they generally do not preserve the interface layout as it is presented to users. The experimental stimuli reproduce the sequential display of messages in a chat window, making it possible to directly manipulate variables such as the spatial position of a repair. This allows the observed effects to be more confidently attributed to specific interface-related manipulations rather than to uncontrolled variation in the content or structure of the conversations.

Another methodological feature was the use of nonce verbs, which have long been employed to control for semantic influences on referential choice and argument interpretation (e.g., Naigles, 1990; Fisher et al., 1994; Kako, 2006). Because nonce verbs carry no real-world associations, participants cannot draw on prior lexical knowledge and are likely to rely primarily on structural and contextual cues when interpreting the discourse. This ensures that any observed effects on referent choice reflect discourse-level factors rather than verb semantics. The use of pronoun resolution as a measure of prominence is well supported in the literature (e.g., Brennan et al., 1987;

Arnold et al., 2000; Patterson, 2013; Patterson and Schumacher, 2020), where prominence is closely tied to how likely an entity is to be picked up by a referring expression.

Together, these design choices made it possible to directly test the main hypotheses of this dissertation. By intentionally misaligning spatial and temporal distance in the chat interface, the experiments could isolate and compare the effects of these two types of distance. Similarly, presenting the same utterance in either the main or the back channel allowed for a direct test of how channel affects referent prominence, while holding message content constant. In each case, the experimental design was intended to address the research questions in a way that reflects both real-world IM usage and concerns in discourse processing theory.

## 9.6 Limitations and Open Questions

While the experimental design of this dissertation offers several strengths, several limitations should be acknowledged, each of which points to opportunities for future work.

An important limitation concerns the use of nonce-verbs. Although these verbs were carefully selected in a pretest to minimize semantic bias and the influence of world knowledge, their use introduced a degree of artificiality, as real conversations rarely include invented words. Participants may have perceived the dialogues as unnatural, potentially affecting how deeply they engaged with the content or how intuitively they interpreted the referents.

A further limitation concerns the simulated nature of the experimental task. Although the setup closely mimicked real IM interfaces, participants read pre-written conversations rather than engaging in live, interactive exchanges. Additionally, although the forced-choice pronoun resolution task clearly measured participants' interpretations, it did not fully capture the real-time dynamics involved in processing references. Consequently, the findings reflect comprehension and interpretation more than spontaneous production. To better capture how prominence and reference resolution

function in actual usage, future studies might use interactive tasks such as semi-scripted chat interactions with chatbots or integrate real-time measures such as eye-tracking or event-related potentials, which would provide more direct evidence of when and how factors such as spatial distance and channel placement influence attention and cognitive effort during discourse processing.

The experimental design also made certain simplifications to maintain control. All dialogues used coordinating discourse relations, which ensured clarity in referent tracking but does not reflect the full range of relations found in natural conversation, such as elaboration, evaluation, background, or contrast. It remains an open question whether a more complex discourse structure interacts differently with distances and channels. Future studies should examine how subordinating or contrastive relations affect reference interpretation in IM dialogue.

Another important limitation lies in the range of emojis used. In Experiment IV, only a small set of emojis, chosen to reflect skeptical or disapproving reactions, were included. These emojis were intended to simulate feedback like “huh?” or “hm?”. Emojis are, however, highly varied in form, tone, and function. They can signal agreement, sarcasm, or humor. Different emoji types would likely have different effects on referent prominence. Future work should test a wider variety of emojis and examine how their communicative function interacts with their placement in the main or back channel.

Another area not explored is individual variation among participants. While the experiments revealed general trends, users may differ in how they respond to temporal, spatial, or channel cues. Variables such as age, reading habits, or familiarity with digital communication could influence how users resolve references. These differences were not analyzed here, but future studies could use individual-level processing measures to investigate variation in discourse strategies.

The dissertation also leaves open questions about other forms of back channel activity. While message editing and emoji reactions were examined, other common functions such as message withdrawal were

not tested. Unlike editing or reacting, withdrawal removes a message from the dialogue entirely, which may have distinct effects on reference tracking and coherence. For instance, does withdrawing a message reduce the prominence of the referents it introduced? Does it prompt readers to reinterpret earlier turns? These remain open empirical questions, particularly as message withdrawal becomes an increasingly common feature of modern IM platforms.

## 9.7 Concluding Remarks

This dissertation set out to investigate how prominence is established and maintained in IM dialogue, with a particular focus on its consequences for the resolution of referring expressions. Across four experiments, the role of temporal and spatial distance and the distinction between main and back channel were examined. Together, these studies contribute to a better understanding of how the technical features of digital communication platforms shape the way users process and interpret discourse.

One of the central findings is that spatial distance consistently outweighs temporal distance in determining referent prominence in IM dialogue. Even when a sent message is edited or commented on and thus being temporally close, participants tend to select the spatially most recent referent as the intended antecedent. This challenges traditional recency-based accounts of prominence, which assume temporal order to be the primary factor, and suggests instead that visual layout plays a crucial role in discourse interpretation in digital settings.

The experiments also reveal that not all utterances are treated equally, even when they carry similar content. Messages in the main channel are more prominent than those in the back channel. This demonstrates that the organization of IM dialogue into distinct channels matters for how users process discourse: prominence is shaped not only by content or timing, but also by where and how information is presented in the interface.

These findings have implications for both linguistic theory and practical application. They suggest that existing models of reference and discourse coherence need to be extended to account for multimodal, layered forms of communication such as IM, and they point to opportunities for platform design and NLP systems to better support user comprehension and reduce referential ambiguity in digital conversations.

Of course, this work is only a starting point. The limitations discussed leave many questions open. Nevertheless, by offering a structured empirical investigation into how the features of digital interfaces affect discourse prominence, this dissertation lays the groundwork for future research into reference processing in digitally mediated communication. Even small features of communication technology, as the findings show, can have a measurable impact on how language is processed and how reference is managed in interaction. Understanding these effects matters not only for linguistic theory, but also for the design of the tools through which so much of human communication now takes place.

# Appendix A

# Selected Nonce Verbs

The following table lists the 54 nonce verbs that were unanimously judged by all participants in the pretest as not conveying any identifiable meaning. Thirty of these verbs were used as stimuli in the experiments; the remaining 24 were used in fillers.

- poffert
- repault
- blagmet
- trocht
- paucht
- truntelt
- schorpt
- knumpt
- kolpt
- krisst
- gandelt
- heppt
- speugelt
- zirft
- wöfft
- tummt
- trehmt
- strapt
- kohbt
- daupt
- tahft
- blohbt
- gereupt
- stropt
- knaffelt
- krunkt
- kneukt
- dorpelt
- frinnt
- wopft
- flürt
- korchelt
- knolkt

- mochelt
- flaugt
- humpft
- krapeliert
- straperiehrt
- kohabusiert
- rapulliert
- weidomiert
- urgelt
- zworbelt
- quokt
- stippert
- knorfelt
- fläht
- porchelt
- frambt
- quibst
- grimbst
- kludert
- quibbt
- glarpt

# Appendix B

# Stimuli Used in Experiments I–III

This appendix provides the full set of dialogue stimuli used in Experiments I – III. Each stimulus consists of a short IM-style exchange between two fictional speakers, followed by a multiple-choice question about pronoun resolution. The dialogues were presented according to the manipulate of each experiment.

Repairs are indicated using a forward slash (“/”) to show the original noun phrase (NP) and the correcting NP, as used in the Late Repair and Early Repair conditions. For instance, in (52-b), “den Hamster” in the initial utterance (Context-U1) was corrected to “den Hasen” in the Repair-U, presented here as *den Hamster / den Hasen*.

In half of the stimuli, the chat interface was horizontally mirrored: speaker A’s messages appeared on the left and speaker B’s on the right, while in the other half, this layout was reversed. Figures B.1 illustrate examples in which the speaker positions are reversed relative to the examples shown in the main body of this dissertation.

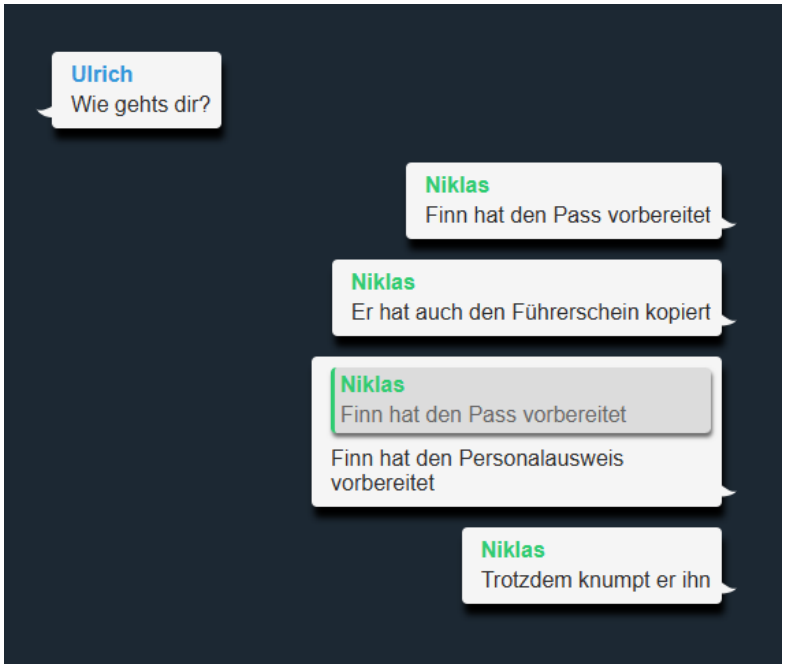



Figure B.1: Example of mirrored chat interface

(52) a. Alexandra: Na?
b. Lena: Frieda hat den Hamster / den Hasen gefüttert (U1/RU)
c. Lena: Sie hat auch den Hund geduscht
d. Lena: Und sie stapt ihn
e. **Question:** Was stapt Frieda?
**A**: Hamster **B**: Hasen **C**: Hund

(53) a. Lea: Hallo, wie geht's?
b. Freya: Greta hat den Schinken / den Braten aufgeschnitten
c. Freya: Danach hat sie den Speck aufgetaut
d. Freya: Außerdem dorpelt sie ihn
e. **Question:** Was dorpelt Greta?
**A**: Schinken **B**: Braten **C**: Speck

(54) a. Alina: Alles klar bei dir?
b. Franziska: Hanna hat einen Computer / einen Beamer ausgeliehen
c. Franziska: Dann hat sie einen Lautsprecher angeschlossen
d. Franziska: Deswegen blagmet sie ihn
e. **Question:** Was blagmet Hanna?
**A**: Computer **B**: Beamer **C**: Lautsprecher

(55) a. Amelie: Und, alles klar?
b. Fiona: Helene hat den Rosenkohl / den Blumenkohl gewaschen
c. Fiona: Und sie hat den Schellfisch zubereitet
d. Fiona: Deshalb wopft sie ihn
e. **Question:** Was wopft Helene?
**A**: Rosenkohl **B**: Blumenkohl **C**: Schellfisch

(56) a. Larissa: Hi, alles okay soweit?
b. Finnja: Konstantin hat den Knoblauch / den Schnittlauch klein geschnitten.
c. Finnja: Daraufhin hat er auch den Kürbis gewürfelt
d. Finnja: Allerdings tummt er ihn
e. **Question:** Was tummt Konstantin?
**A**: Knoblauch **B**: Schnittlauch **C**: Kürbis

(57) a. Lara: Alles gut?
b. Eva: Julius hat einen Pudding / einen Jogurt bestellt
c. Eva: Er hat noch einen Kuchen mitgenommen
d. Eva: Aber er knaffelt ihn
e. **Question:** Was knaffelt Julius?
**A**: Pudding **B**: Jogurt **C**: Kuchen

(58) a. Amira: Wie gehts dir?
b. Ella: Julian hat einen Hamburger / einen Windbeutel gemacht
c. Ella: Danach hat er den Toast getoastet
d. Ella: Trotzdem daupt er ihn
e. **Question:** Was daupt Julian?

**A**: Hamburger **B**: Windbeutel **C**: Toast

(59) a. Klara: Na, wie läufts bei dir?
b. Lukas: Ida hat einen Fußball / einen Basketball gekauft
c. Lukas: Sie hat aber schon einen Baseball besorgt
d. Lukas: Jedoch quibbt sie ihn
e. **Question:** Was quibbt Ida?
**A**: Fußball **B**: Basketball **C**: Baseball

(60) a. Anastasia: Wie läufts so?
b. Marc: Isabella hat für diesen Winter einen Heizofen / einen Heizlüfter besorgt
c. Marc: Außerdem hat sie einen extra Heizkörper installiert
d. Marc: Denn sie wöfft ihn
e. **Question:** Was wöfft Isabella?
**A**: Heizofen **B**: Heizlüfter **C**: Heizkörper

(61) a. Kathrin: Hey, was gibts Neues?
b. Marius: Jana hat einen Kleiderschrank / einen Bücherschrank lackiert
c. Marius: Sie hat auch den Couchtisch hingestellt
d. Marius: Also repault sie ihn
e. **Question:** Was repault Jana?
**A**: Kleiderschrank **B**: Bücherschrank **C**: Couchtisch

(62) a. Anna: Na?
b. Mattis: Jasmine hat einen Rucksack / einen Schalfsack ausgepackt
c. Mattis: Sie hat dann einen Pullover herausgenommen
d. Mattis: Und sie kneukt ihn
e. **Question:** Was kneukt Jasmine?
**A**: Rucksack **B**: Schlafsack **C**: Pullover

(63) a. Juna: Hallo, wie geht's?
b. Maximilian: Henry hat einen Kochtopf / einen Wok hingestellt
c. Maximilian: Henry hat einen Messbecher weggenommen

d. Maximilian: Außerdem grimbst er ihn
e. **Question:** Was grimbst Henry?
**A**: Kochtopf **B**: Wok **C**: Messbecher

(64) a. Juliane: Alles klar bei dir?
b. Michael: Henrik hat den CD-Spieler / den MP3-Player verkauft
c. Michael: Er hat auch den Camcorder angestrichen
d. Michael: Deswegen glarpt er ihn
e. **Question:** Was glarpt Henrik?
**A**: CD-Spieler **B**: MP3-Player **C**: Camcorder

(65) a. Annabelle: Und, alles klar?
b. Miran: Fritz hat einen Blattkäfer / einen Marienkäfer gefangen
c. Miran: Er hat wieder einen Schmetterling gefunden
d. Miran: Deshalb tahft er ihn
e. **Question:** Was tahft Fritz?
**A**: Blattkäfer **B**: Marienkäfer **C**: Schmetterling

(66) a. Annika: Hi, alles okay soweit?
b. Moritz: Friedrich hat einen Pfirsich / einen Apfel genommen
c. Moritz: Und er hat schon einen Obstkorb dekoriert
d. Moritz: Allerdings truntelt er ihn
e. **Question:** Was truntelt Friedrich?
**A**: Pfirsich **B**: Apfel **C**: Obstkorb

(67) a. Vincent: Alles gut?
b. Nick: Franz hat einen Hobel / einen Rohrabschneider benutzt
c. Nick: Er hat einen Fuchsschwanz kaputt gemacht
d. Nick: Aber er flaugt ihn
e. **Question:** Was flaugt Franz?
**A**: Hobel **B**: Rohrabschneider **C**: Fuchsschwanz

(68) a. Ulrich: Wie gehts dir?
b. Niklas: Finn hat den Pass / den Personalausweis vorbereitet
c. Niklas: Er hat auch den Führerschein kopiert
d. Niklas: Trotzdem knumpt er ihn

e. **Question:** Was knumpt Finn?
**A**: Pass **B**: Personalausweis **C**: Führerschein

(69) a. Tony: Na, wie läufts bei dir?
b. Noah: Felix hat einen Schal / einen Handschuh gestrickt
c. Noah: Danach hat er einen Pullover zerlegt
d. Noah: Jedoch zworbelt er ihn
e. **Question:** Was zworbelt Felix?
**A**: Schal **B**: Handschuh **C**: Pullover

(70) a. Tom: Wie läufts so?
b. Noel: Fabian hat den Koffer / den Kasten geöffnet
c. Noel: Danach hat er den Coupon rausgenommen
d. Noel: Denn er stippert ihn
e. **Question:** Was stippert Fabian?
**A**: Koffer **B**: Kasten **C**: Coupon

(71) a. Tobias: Hey, was gibts Neues?
b. Oskar: Josephina hat einen Schlüssel / einen Transponderchip gefunden
c. Oskar: Daneben hat sie auch einen Schlüsselanhänger aufgehoben
d. Oskar: Also quibst sie ihn
e. **Question:** Was quibst Josephina?
**A**: Schlüssel **B**: Transponderchip **C**: Schlüsselanhänger

(72) a. Timo: Na?
b. Paul: Julia hat den Porzellanteller / den Glasteller zerbrochen
c. Paul: Dann hat sie noch den Tisch getreten
d. Paul: Und sie gereupt ihn
e. **Question:** Was gereupt Julia?
**A**: Porzellanteller **B**: Glasteller **C**: Tisch

(73) a. Till: Hallo, wie geht's?
b. Philipp: Leonie hat den Whisky endlich / den Rum endlich gekauft

c. Philipp: Sie hat auch den Gin mitgebracht
d. Philipp: Außerdem korchelt sie ihn
e. **Question:** Was korchelt Leonie?
**A**: Whisky **B**: Rum **C**: Gin

(74) a. Theodor: Alles klar bei dir?
b. Elisabeth: Aaron hat sich einen Bleistift / sich einen Füller ausgeliehen
c. Elisabeth: Er hat aber einen Spitzer gestohlen
d. Elisabeth: Deswegen schorpt er ihn
e. **Question:** Was schorpt Aaron?
**A**: Bleistift **B**: Füller **C**: Spitzer

(75) a. Stephan: Und, alles klar?
b. Elisa: Adam hat den Bettbezug / den Kissenbezug gewaschen
c. Elisa: Er hat den Schlafanzug gebügelt
d. Elisa: Deshalb weidomiert er ihn
e. **Question:** Was weidomiert Adam?
**A**: Bettbezug **B**: Kissenbezug **C**: Schlafanzug

(76) a. Simon: Hi, alles okay soweit?
b. Elena: Adrian hat einen Roller / einen Spachtel mitgebracht
c. Elena: Er hat auch einen Farbtopf geöffnet
d. Elena: Allerdings krisst er ihn
e. **Question:** Was krisst Adrian?
**A**: Roller **B**: Spachtel **C**: Farbtopf

(77) a. Sebastian: Alles gut?
b. Diana: Alexander hat einen Rauchmelder / einen Feuermelder installiert
c. Diana: Er hat auch einen Feuerlöscher hingestellt
d. Diana: Aber er quokt ihn
e. **Question:** Was quokt Alexander?
**A**: Rauchmelder **B**: Feuermelder **C**: Feuerlöscher

(78) a. Samuel: Wie gehts dir?

b. Clara: Lina hat einen Döner / einen Donut verschlungen
c. Clara: Sie hat aber vorher schon einen Lutscher gegessen
d. Clara: Trotzdem knorfelt sie ihn
e. **Question:** Was knorfelt Lina?
**A**: Döner **B**: Donut **C**: Lutscher

(79) a. Robin: Na, wie läufts bei dir?
b. Charlotte: Lisa hat den Konferenzsaal / den Tanzsaal geschmückt
c. Charlotte: Dann hat sie noch den Weihnachtsbaum verschoben
d. Charlotte: Jedoch rapulliert sie ihn
e. **Question:** Was rapulliert Lisa?
**A**: Konferenzraum **B**: Tanzsaal **C**: Weihnachtsbaum

(80) a. Richard: Wie läufts so?
b. Ariana: Livia hat einen Mahnbescheid / einen Steuerbescheid bekommen
c. Ariana: Dann hat sie noch einen Aktenordner gekauft
d. Ariana: Denn sie stropt ihn
e. **Question:** Was stropt Livia?
**A**: Mahnbescheid **B**: Steuerbescheid **C**: Aktenordner

(81) a. Rafael: Hey, was gibts Neues?
b. Antonia: Lotte hat einen Tippfehler / einen Rechtschreibfehler gemacht
c. Antonia: Danach hat sie einen Absatz abgebrochen
d. Antonia: Also heppt sie ihn
e. **Question:** Was heppt Lotte?
**A**: Tippfehler **B**: Rechtschreibfehler **C**: Absatz

# Appendix C

# Stimuli: Experiment IV

(82) a. Alexandra: Na?
b. Lena: Frieda hat den Hamster gefüttert
c. Lena: Sie hat auch den Hund geduscht
d. Alexandra: 😕
e. Lena: Und sie strapt ihn
f. **Question:** Was strapt Frieda?
**A:** Hamster **B:** Hund

(83) a. Lea: Hallo, wie geht's?
b. Freya: Greta hat den Schinken aufgeschnitten
c. Freya: Danach hat sie den Speck aufgetaut
d. Lea: 😠
e. Freya: Außerdem dorpelt sie ihn
f. **Question:** Was dorpelt Greta?
**A:** Schinken **B:** Speck

(84) a. Alina: Alles klar bei dir?
b. Franziska: Hanna hat einen Computer ausgeliehen
c. Franziska: Dann hat sie einen Lautsprecher angeschlossen
d. Alina: 🧐
e. Franziska: Deswegen blagmet sie ihn
f. **Question:** Was blagmet Hanna?

**A:** Computer **B:** Lautsprecher

(85) a. Amelie: Und, alles klar?
b. Fiona: Helene hat den Rosenkohl gewaschen
c. Fiona: Und sie hat den Schellfisch zubereitet
d. Amelie: 🤨
e. Fiona: Deshalb wopft sie ihn
f. **Question:** Was wopft Helene?
**A:** Rosenkohl **B:** Schellfisch

(86) a. Larissa: Hi, alles okay soweit?
b. Finnja: Konstantin hat den Knoblauch klein geschnitten
c. Finnja: Daraufhin hat er auch den Kürbis gewürfelt
d. Larissa: 🤔
e. Finnja: Allerdings tummt er ihn
f. **Question:** Was tummt Konstantin?
**A:** Knoblauch **B:** Kürbis

(87) a. Lara: Alles gut?
b. Eva: Julius hat einen Pudding bestellt
c. Eva: Er hat noch einen Kuchen mitgenommen
d. Lara: 😮
e. Eva: Aber er knaffelt ihn
f. **Question:** Was knaffelt Julius?
**A:** Pudding **B:** Kuchen

(88) a. Amira: Wie gehts dir?
b. Ella: Julian hat einen Windbeutel gemacht
c. Ella: Danach hat er den Toast getoastet
d. Amira: 😲
e. Ella: Trotzdem daupt er ihn
f. **Question:** Was daupt Julian?
**A:** Windbeutel **B:** Toast

(89) a. Klara: Na, wie läufts bei dir?

b. Lukas: Ida hat einen Fußball gekauft
c. Lukas: Sie hat aber schon einen Baseball besorgt
d. Klara: 🙄
e. Lukas: Jedoch quibbt sie ihn
f. **Question:** Was quibbt Ida?
**A:** Fußball **B:** Baseball

(90) a. Anastasia: Wie läufts so?
b. Marc: Isabella hat für diesen Winter einen Heizofen besorgt
c. Marc: Außerdem hat sie einen extra Heizkörper installiert
d. Anastasia: 👎
e. Marc: Denn sie wöfft ihn
f. **Question:** Was wöfft Isabella?
**A:** Heizofen **B:** Heizkörper

(91) a. Kathrin: Hey, was gibts Neues?
b. Marius: Jana hat einen Kleiderschrank lackiert
c. Marius: Sie hat auch den Couchtisch hingestellt
d. Kathrin: 😒
e. Marius: Also repault sie ihn
f. **Question:** Was repault Jana?
**A:** Kleiderschrank **B:** Couchtisch

(92) a. Anna: Na?
b. Mattis: Jasmine hat einen Rucksack ausgepackt
c. Mattis: Sie hat dann einen Pullover herausgenommen
d. Anna: 😕
e. Mattis: Und sie kneukt ihn
f. **Question:** Was kneukt Jasmine?
**A:** Rucksack **B:** Pullover

(93) a. Juna: Hallo, wie geht's?
b. Maximilian: Henry hat einen Wok hingestellt
c. Maximilian: Henry hat einen Messbecher weggenommen

d. Juna: 😠
e. Maximilian: Außerdem grimbst er ihn
f. **Question:** Was grimbst Henry?
**A:** Wok **B:** Messbecher

(94) a. Juliane: Alles klar bei dir?
b. Michael: Henrik hat den CD-Spieler verkauft
c. Michael: Er hat auch den Camcorder angestrichen
d. Juliane: 🧐
e. Michael: Deswegen glarpt er ihn
f. **Question:** Was glarpt Henrik?
**A:** CD-Spieler **B:** Camcorder

(95) a. Annabelle: Und, alles klar?
b. Miran: Fritz hat einen Marienkäfer gefangen
c. Miran: Er hat wieder einen Schmetterling gefunden
d. Annabelle: 🤨
e. Miran: Deshalb tahft er ihn
f. **Question:** Was tahft Fritz?
**A:** Marienkäfer **B:** Schmetterling

(96) a. Annika: Hi, alles okay soweit?
b. Moritz: Friedrich hat einen Apfel genommen
c. Moritz: Und er hat schon einen Obstkorb dekoriert
d. Annika: 🤔
e. Moritz: Allerdings truntelt er ihn
f. **Question:** Was truntelt Friedrich?
**A:** Apfel **B:** Obstkorb

(97) a. Vincent: Alles gut?
b. Nick: Franz hat einen Hobel benutzt
c. Nick: Er hat einen Fuchsschwanz kaputt gemacht
d. Vincent: 😮
e. Nick: Aber er flaugt ihn

f. **Question:** Was flaugt Franz?
**A:** Hobel **B:** Fuchsschwanz

(98) a. Ulrich: Wie gehts dir?
b. Niklas: Finn hat den Pass vorbereitet
c. Niklas: Er hat auch den Personalausweis kopiert
d. Ulrich: 😲
e. Niklas: Trotzdem knumpt er ihn
f. **Question:** Was knumpt Finn?
**A:** Pass **B:** Personalausweis

(99) a. Tony: Na, wie läufts bei dir?
b. Noah: Felix hat einen Schal gestrickt
c. Noah: Danach hat er einen Pullover zerlegt
d. Tony: 🙄
e. Noah: Jedoch zworbelt er ihn
f. **Question:** Was zworbelt Felix?
**A:** Schal **B:** Pullover

(100) a. Tom: Wie läufts so?
b. Noel: Fabian hat den Koffer geöffnet
c. Noel: Danach hat er den Coupon rausgenommen
d. Tom: 👎
e. Noel: Denn er stippert ihn
f. **Question:** Was stippert Fabian?
**A:** Koffer **B:** Coupon

(101) a. Tobias: Hey, was gibts Neues?
b. Oskar: Josephina hat einen Schlüssel gefunden
c. Oskar: Daneben hat sie auch einen Schlüsselanhänger aufgehoben
d. Tobias: 😒
e. Oskar: Also quibst sie ihn
f. **Question:** Was quibst Josephina?
**A:** Schlüssel **B:** Schlüsselanhänger

(102) a. Timo: Na?
b. Paul: Julia hat den Porzellanteller zerbrochen
c. Paul: Dann hat sie noch den Tisch getreten
d. Timo: 😕
e. Paul: Und sie gereupt ihn
f. **Question:** Was gereupt Julia?
**A:** Porzellanteller **B:** Tisch

(103) a. Till: Hallo, wie geht's?
b. Philipp: Leonie hat den Whisky endlich gekauft
c. Philipp: Sie hat auch den Gin mitgebracht
d. Till: 😠
e. Philipp: Außerdem korchelt sie ihn
f. **Question:** Was korchelt Leonie?
**A:** Whisky **B:** Gin

(104) a. Theodor: Alles klar bei dir?
b. Elisabeth: Aaron hat sich einen Bleistift ausgeliehen
c. Elisabeth: Er hat aber einen Spitzer gestohlen
d. Theodor: 🧐
e. Elisabeth: Deswegen schorpt er ihn
f. **Question:** Was schorpt Aaron?
**A:** Bleistift **B:** Spitzer

(105) a. Stephan: Und, alles klar?
b. Elisa: Adam hat den Bettbezug gewaschen
c. Elisa: Er hat den Schlafanzug gebügelt
d. Stephan: 🤨
e. Elisa: Deshalb weidomiert er ihn
f. **Question:** Was weidomiert Adam?
**A:** Bettbezug **B:** Schlafanzug

(106) a. Simon: Hi, alles okay soweit?
b. Elena: Adrian hat einen Roller mitgebracht
c. Elena: Er hat auch einen Farbtopf geöffnet

d. Simon: 🤔
e. Elena: Allerdings krisst er ihn
f. **Question:** Was krisst Adrian?
**A:** Roller **B:** Farbtopf

(107) a. Sebastian: Alles gut?
b. Diana: Alexander hat einen Feuermelder installiert
c. Diana: Er hat auch einen Feuerlöscher hingestellt
d. Sebastian: 😮
e. Diana: Aber er quokt ihn
f. **Question:** Was quokt Alexander?
**A:** Feuermelder **B:** Feuerlöscher

(108) a. Samuel: Wie gehts dir?
b. Clara: Lina hat einen Donut verschlungen
c. Clara: Sie hat aber vorher schon einen Lutscher gegessen
d. Samuel: 😲
e. Clara: Trotzdem knorfelt sie ihn
f. **Question:** Was knorfelt Lina?
**A:** Donut **B:** Lutscher

(109) a. Robin: Na, wie läufts bei dir?
b. Charlotte: Lisa hat den Tanzsaal geschmückt
c. Charlotte: Dann hat sie noch den Weihnachtsbaum verschoben
d. Robin: 🙄
e. Charlotte: Jedoch rapulliert sie ihn
f. **Question:** Was rapulliert Lisa?
**A:** Tanzsaal **B:** Weihnachtsbaum

(110) a. Richard: Wie läufts so?
b. Ariana: Livia hat einen Mahnbescheid bekommen
c. Ariana: Dann hat sie noch einen Aktenordner gekauft
d. Richard: 👎

e. Ariana: Denn sie stropt ihn
f. **Question:** Was stropt Livia?
**A:** Mahnbescheid **B:** Aktenordner

(111) a. Rafael: Hey, was gibts Neues?
b. Antonia: Lotte hat einen Tippfehler gemacht
c. Antonia: Danach hat sie einen Absatz abgebrochen
d. Rafael: 😒
e. Antonia: Also heppt sie ihn
f. **Question:** Was heppt Lotte?
**A:** Tippfehler **B:** Absatz

# Bibliography


Abrams, Zsuzsanna Ittzes. 2001. Computer-mediated communication and group journals: Expanding the repertoire of participant roles. *System*, 29(4):489–503.

AirDroid. 2024. Turn off read receipts in whatsapp: What happens. Accessed: 2025-09-06.

AirDroid. 2025. Wie man erkennt, ob jemand ihre whatsapp-nachricht gelesen hat. https://www.airdroid.com/de/parent-control/know-if-someone-read-your-message-whatsapp/. Accessed: 10 September 2025.

Albert, Georg. 2015. Semiotik und syntax von emoticons. *Zeitschrift für angewandte Linguistik*, 2015(62):3–22.

Allwood, Jens, Nivre, Joakim, and Ahlsén, Elisabeth. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, 9(1):1–26.

Annamalai, Subashini and Salam, Sobihatun Nur Abdul. 2017. Undergraduates' interpretation on whatsapp smiley emoji. *Jurnal Komunikasi, Malaysian Journal of Communication*, 33(4):89–103.

Arens, Katja. 2014. Whatsapp: Kommunikation 2.0. eine qualitative betrachtung der multimedialen möglichkeiten. In *SMS, WhatsApp & Co. Gattungsanalytische, kontrastive und variationslinguistische Perspektiven zur Analyse mobiler Kommunikation*, pages 81–106. Monsenstein und Vannerdat.

Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. Routledge, New York.

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

Arnold, Jennifer E. 2025. Why does recency guide pronoun comprehension? it's not just topicality, attention, or predictability. *Discourse Processes*, 0(0):1–23.

Arnold, Jennifer E, Eisenband, Janet G, Brown-Schmidt, Sarah, and Trueswell, John C. 2000. The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76(1):B13–B26.

Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse*, volume 50 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers, Dordrecht, Boston, London.

Asher, Nicholas and Lascarides, Alex. 2003. *Logics of Conversation*. Cambridge University Press.

Asher, Nicholas and Vieu, Laure. 2005. Subordinating and coordinating discourse relations. *Lingua*, 115(4):591–610.

Baron, Naomi S. 2004. See you online: Gender issues in college student use of instant messaging. *Journal of language and social psychology*, 23(4):397–423.

Baron, Naomi S. 2013. Instant messaging. *Pragmatics of computer-mediated communication*, 9:135.

Barzilay, Regina and Lapata, Mirella. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

Bavelas, Janet B., Coates, Linda, and Johnson, Trudy. 2000. Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6):941–952.

BBC News. 2022. Technology advancements in 2022. Accessed: 2025-03-05.

Berglund, Therese Örnberg. 2009. Disrupted turn adjacency and coherence maintenance in instant messaging conversations. *Language@ Internet*, 6.

Böhlke, Olaf. 2003. A comparison of student participation levels by group size and language stages during chatroom and face-to-face discussions in german. *CALICO journal*, pages 67–87.

Brennan, Susan E, Friedman, Marilyn W, and Pollard, Carl. 1987. A centering approach to pronouns. In *25th annual meeting of the association for computational linguistics*, pages 155–162.

Brennan, Susan E and Hulteen, Eric A. 1995. Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-based systems*, 8(2-3):143–151.

Burmester, Juliane, Sauermann, Antje, Spalek, Katharina, and and, Isabell Wartenburger. 2018. Sensitivity to salience: linguistic vs. visual cues affect sentence processing and pronoun resolution. *Language, Cognition and Neuroscience*, 33(6):784–801.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28.

Cacciagrano, Diletta and Corradini, Flavio. 2001. On synchronous and asynchronous communication paradigms. In *Theoretical Computer Science. ICTCS 2001*, volume 2202 of *Lecture Notes in Computer Science*, pages 256–270, Berlin, Heidelberg. Springer.

Chafe, Wallace L. 1974. Language and consciousness. *Language*, 50(1):111–133.

Chen, Yihua, Yang, Xingchen, Howman, Hannah, and Filik, Ruth. 2024. Individual differences in emoji comprehension: Gender, age, and culture. *Plos one*, 19(2):e0297379.

Clark, Herbert H. 1994. Managing problems in speaking. *Speech communication*, 15(3-4):243–250.

Clark, Herbert H. 1996. *Using Language*. Cambridge University Press, Cambridge.

Clark, Herbert H. and Brennan, Susan E. 1991. Grounding in communication. In Lauren Resnick, Levine B., M. John, Stephanie Teasley, and D., editors, *Perspectives on Socially Shared Cognition*, pages 13–1991. American Psychological Association.

Clark, Herbert H. and Fox Tree, Jean E. 2002. Using uh and um in spontaneous speaking. *Cognition*, 84(1):73–111.

Clark, Herbert H and Schaefer, Edward F. 1987. Collaborating on contributions to conversations. *Language and cognitive processes*, 2(1):19–41.

Clark, Herbert H and Schaefer, Edward F. 1989. Contributing to discourse. *Cognitive science*, 13(2):259–294.

Clark, Herbert H. and Sengul, C. J. 1979. In search of referents for nouns and pronouns. *Memory & Cognition*, 7:35–41. Accepted: 14 November 1978.

Clark, Herbert H and Wilkes-Gibbs, Deanna. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Cohen, Henri, Douaire, Josée, and Elsabbagh, Mayada. 2001. The role of prosody in discourse processing. *Brain and Cognition*, 46(1):73–82.

Collister, Lauren Brittany. 2011. *-repair in online discourse. *Journal of Pragmatics*, 43(3):918–921.

Colman, Marcus and Healey, Patrick G. T. 2011. The distribution of repair in dialogue. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, pages 1563–1568.

Craig, David. 2003. Instant messaging: The language of youth literacy. *The Boothe Prize Essays*, 1:116–133.

Cramer, Henriette, De Juan, Paloma, and Tetreault, Joel. 2016. Sender-intended functions of emojis in us messaging. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services*, pages 504–509.

Crawley, Rosalind A, Stevenson, Rosemary J, and Kleinman, David. 1990. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19:245–264.

Crystal, David. 2006. *Language and the Internet*, 2nd edition. Cambridge University Press. Accessed: 2023-10-01.

Danesi, Marcel. 2017. *The semiotics of emoji: The rise of visual language in the age of the internet*. Bloomsbury Publishing.

Darhower, Mark. 2002. Instructional features of synchronous computer-mediated communication in the intermediate l2 class: A sociocultural case study. *CALICO journal*, 19(2):249–278.

DataReportal. 2025. Global digital overview. Accessed: 2025-03-05.

De Seta, Gabriele. 2018. Biaoqing: The circulation of emoticons, emoji, stickers, and custom images on chinese digital media platforms. *First Monday*, 23(9).

Derriks, Barbara and Willems, Dominique. 1998. Negative feedback in information dialogues: identification, classification and problem-solving procedures. *International journal of human-computer studies*, 48(5):577–604.

Dhakal, Vivek, Feit, Anna Maria, Kristensson, Per Ola, and Oulasvirta, Antti. 2018. Observations on typing from 136 million keystrokes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–12, New York, NY, USA. Association for Computing Machinery.

Dingemanse, Mark, Roberts, Seán G, Baranova, Julija, Blythe, Joe, Drew, Paul, Floyd, Simeon, Gisladottir, Rosa S, Kendrick, Kobin H, Levinson, Stephen C, Manrique, Elizabeth, et al. 2015. Universal principles in the repair of communication problems. *PLoS ONE*, 10(9):e0136100.

Drummond, Alex. 2023. Ibex. https://korpling.german.hu-berlin.de/ibex. Accessed July 20, 2023.

Drummond, Kent and and, Robert Hopper. 1993. Back channels revisited: Acknowledgment tokens and speakership incipiency. *Research on Language and Social Interaction*, 26(2):157–177.

Dudenredaktion. 2023. Häufigkeit. https://www.duden.de/hilfe/haeufigkeit. Accessed April 6, 2023.

Duncan, Starkey. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2):283.

Duncan, Starkey. 1974. On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 3(2):161–180.

Dürscheid, Christa and Frick, Karina. 2014. Keyboard-to-screen-kommunikation gestern und heute: Sms und whatsapp im vergleich. *Sprachen*, pages 149–181.

Dürscheid, Christa and Siever, Christina Margrit. 2017. Jenseits des alphabets–kommunikation mit emojis. *Zeitschrift für germanistische Linguistik*, 45(2):256–285.

Emojipedia. 2025. Emojipedia. Accessed: 2025-03-11.

Esaulova, Yulia, Dolscheid, Sarah, Reuters, Sabine, and Penke, Martina. 2021. The Alignment of Agent-First Preferences with Visual Event Representations: Contrasting German and Arabic. *Journal of Psycholinguistic Research*, 50(4):843–861.

Faulkner, Xristine and Culwin, Fintan. 2005. When fingers do the talking: a study of text messaging. *Interacting with computers*, 17(2):167–185.

Fisher, Cynthia, Hall, D Geoffrey, Rakowitz, Susan, and Gleitman, Lila. 1994. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92:333–375.

Fladrich, Marcel and Imo, Wolfgang. 2020. □ j = □ j? oder: Das gelächter der geschlechter 2.0: Emojigebrauch in der whatsapp-kommunikation. In Konstanze Marx, Henning Lobin, and Axel Schmidt, editors, *Deutsch in Sozialen Medien: Interaktiv – multimodal – vielfältig*, pages 95–122. De Gruyter, Berlin, Boston.

Franco, Courtny L and Fugate, Jennifer MB. 2020. Emoji face renderings: Exploring the role emoji platform differences have on emotional interpretation. *Journal of Nonverbal Behavior*, 44(2):301–328.

Fraundorf, Scott H, Benjamin, Aaron S, and Watson, Duane G. 2013. What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of memory and language*, 69(3):196–227.

Frege, Gottlob. 1892/1948. Sense and reference. *The Philosophical Review*, 57(3):209–230. Originally published in 1892 as *Über Sinn und Bedeutung*, translated by Max Black.

Gardner, Rod. 1997. The conversation object mm: A weak and variable acknowledging token. *Research on Language and Social Interaction*, 30(2):131–156.

Gardner, Rod. 1998. Between speaking and listening: The vocalisation of understandings1. *Applied linguistics*, 19(2):204–224.

Gardner, Rod. 2001. *When Listeners Talk: Response Tokens and Listener Stance*, volume 92 of *Pragmatics & Beyond New Series*. John Benjamins, Amsterdam.

Gelman, Andrew, Jakulin, Aleks, Pittau, Maria Grazia, and Su, Yu-Sung. 2008. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, 2(4):1360–1383.

Gernsbacher, Morton Ann and Hargreaves, David J. 1988. Accessing sentence participants: The advantage of first mention. *Journal of memory and language*, 27(6):699–717.

Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón, editor, *Topic continuity in discourse: A quantitative cross-language study*, pages 1–42. John Benjamins Publishing Company.

Gleitman, Lila R., January, David, Nappa, Rebecca, and Trueswell, John C. 2007. On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, 57(4):544–569. Language-Vision Interaction.

Goodwin, Charles. 1980. Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological Inquiry*, 50(3-4):272–302.

Goodwin, Charles. 1981. Achieving mutual orientation at turn beginning. In *Conversational Organization: Interaction Between Speakers and Hearers*, chapter 2, pages 55–94. Academic Press, New York.

Grosz, Barbara, Joshi, Aravind, and Weinstein, Scott. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

Grosz, Barbara J., Joshi, Aravind K., and Weinstein, Scott. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Grosz, Patrick G., Greenberg, Gabriel, Leon, Carson De, McCready, Elin, and Solt, Stephanie. 2023. A semantics of face emoji in discourse. *Linguistics and Philosophy*, 46(5):905–957.

Gundel, Jeanette K, Hedberg, Nancy, and Zacharski, Ron. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, pages 274–307.

Guntuku, Sharath Chandra, Li, Mingyang, Tay, Louis, and Ungar, Lyle H. 2019. Studying cultural differences in emoji usage across the east and the west. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 226–235.

Haas, Gabriel, Gugenheimer, Jan, Rixen, Jan Ole, Schaub, Florian, and Rukzio, Enrico. 2020. "they like to hear my voice": Exploring usage behavior in speech-based mobile instant messaging. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–10.

Halliday, Michael AK. 1967. Notes on transitivity and theme in english: Part 2. *Journal of linguistics*, 3(2):199–244.

Halliday, Michael Alexander Kirkwood and Hasan, Ruqaiya. 1976. *Cohesion in English*. Routledge.

Hasyim, Muhammad. 2019. Linguistic functions of emoji in social media communication. *Opcion*, 35.

Heritage, John. 1985. A change-of-state token and aspects of its sequential placement. In J. MaxwellEditor Atkinson, editor, *Structures of Social Action*, Studies in Emotion and Social Interaction, page 299–345. Cambridge University Press.

Herring, Susan C. 1996. *Computer-mediated communication*. John Benjamins publishing company.

Herring, Susan C. 1999. Interactional coherence in cmc. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, pages 13–pp. IEEE.

Herring, Susan C. 2013. 10. relevance in computer-mediated conversation. *Pragmatics of computer-mediated communication*, 9:245.

Herring, Susan C. and Dainas, Ashley R. 2017. ”nice picture comment!” graphicons in facebook comment threads. In *Hawaii International Conference on System Sciences*.

Heylen, Dirk. 2006. Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics*, 3(03):241–267.

Holvikivi, Jaana. 2019. New visual trends and language in internet communication-emoji and beyond. In *Nordic Science and Technology Studies conference in Tampere*, pages 1–9.

Jasinskaja, Katja, Chiriacescu, Sofiana, Donazzan, Marta, von Heusinger, Klaus, and Hinterwimmer, Stefan. 2015. Prominence in discourse. In *Prominence in Linguistics: Proceedings of the pS-prominenceS International Conference*, pages 134–153, Viterbo, Italy. DISUCOM Press, University of Tuscia.

Jasinskaja, Katja and Karagjosova, Elena. 2020. Rhetorical relations. *The Wiley Blackwell companion to semantics*, 4:2645–2673.

Jasinskaja, Katja, Li, Yuting, Same, Fahime, and Uerlings, David. 2024. Reference and discourse structure annotation of elicited chat continuations in german. In *Proceedings of The 18th Linguistic Annotation Workshop (LAW-XVIII)*, pages 176–187.

Jefferson, Gail. 1984. Notes on a systematic deployment of the acknowledgement tokens “yeah”; and “mm hm”;. *Paper in Linguistics*, 17(2):197–216.

Jewitt, Carey. 2017. *Routledge handbook of multimodal analysis*, second edition, first published in paperback edition. Routledge Taylor & Francis Group, London ; New York.

Jewitt, Carey, Bezemer, Jeff, and O’Halloran, Kay. 2016. *Introducing multimodality*. Routledge.

Johnson, Genevieve. 2006. Synchronous and asynchronous text-based cmc in educational contexts: A review of recent research. *TechTrends: Linking Research & Practice to Improve Learning*, 50(4).

Jurafsky, Daniel and Martin, James H. 2025. Discourse coherence. In *Speech and Language Processing*, chapter 24. Pearson. Draft of January 12, 2025.

Kako, Edward. 2006. Thematic role properties of subjects and objects. *Cognition*, 101(1):1–42.

Kayan, Shipra, Fussell, Susan R, and Setlock, Leslie D. 2006. Cultural differences in the use of instant messaging in asia and north america. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 525–528.

Keenan, Edward L. and Comrie, Bernard. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8(1):63–99.

Keenan, Edward L. and Dryer, Matthew S. 2007. Passive in the world's languages. In TimothyEditor Shopen, editor, *Language Typology and Syntactic Description*, page 325–361. Cambridge University Press.

Kelly, Ryan and Watts, Leon. 2015. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. In *Experiences of technology appropriation: Unanticipated users, usage, circumstances, and design*.

Kendon, Adam. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

Kitzinger, Celia. 2012. Repair. In Jack Sidnell and Tanya Stivers, editors, *The Handbook of Conversation Analysis*, pages 229–256. Wiley-Blackwell. First published: 28 August 2012.

König, Katharina. 2015. Dialogkonstitution und sequenzmuster in der sms- und whatsapp-kommunikation. *Travaux neuchâtelois de linguistique*, 63:87–107.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4):243–276.

Kuno, Susumu. 1972. Functional sentence perspective: A case study from japanese and english. *Linguistic inquiry*, 3(3):269–320.

König, Katharina. 2024. Transmodal messenger interaction–analysing the sequentiality of text and audio postings in whatsapp chats. *Discourse, Context & Media*, 62:100818.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge Studies in Linguistics. Cambridge University Press.

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics*, 39(3):463–516.

Levelt, Willem JM. 1983. Monitoring and self-repair in speech. *Cognition*, 14(1):41–104.

Li, Li and Yang, Yue. 2018. Pragmatic functions of emoji in internet-based communication—a corpus-based study. *Asian-Pacific Journal of Second and Foreign Language Education*, 3:1–12.

Li, Yuting. 2021. Empirical study on emoji usage as feedback strategy in dialogue. Master's thesis, Universität zu Köln. Master's thesis.

Ling, Rich and Baron, Naomi S. 2007. Text messaging and im: Linguistic comparison of american college data. *Journal of language and social psychology*, 26(3):291–298.

Maia, J. 2018. *Towards a Typographical Linguistics: The Semantics-Pragmatics of Typographic Emphasis in Discourse*. Doctoral dissertation, University of South Carolina. Retrieved from Scholar Commons.

Mann, William C and Thompson, Sandra A. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

McClave, Evelyn Z. 2000. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, 32(7):855–878.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Meredith, Joanne and Stokoe, Elizabeth. 2014. Repair: Comparing facebook 'chat'with spoken interaction. *Discourse & communication*, 8(2):181–207.

Miller, Hannah, Kluver, Daniel, Thebault-Spieker, Jacob, Terveen, Loren, and Hecht, Brent. 2017. Understanding emoji ambiguity in context: The role of text in emoji-related miscommunication. In *Proceedings of the 11th International AAAI Conference on Web and Social Media*, pages 152–161.

Miller, Hannah, Thebault-Spieker, Jacob, Chang, Shuo, Johnson, Isaac, Terveen, Loren, and Hecht, Brent. 2016. "blissfully happy" or "ready to fight": Varying interpretations of emoji. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, pages 259–268.

Mondada, Lorenza. 2016. 17. multimodal resources and the organization of social interaction. In Andrea Rocci and Louis de Saussure, editors, *Verbal Communication*, pages 329–350. De Gruyter Mouton, Berlin, Boston.

Muhammad Hasyim, Burhanuddin Arafah. 2019. Linguistic functions of emoji in social media communication. *Opción*, 35:558–574.

Naigles, Letitia. 1990. Children use syntax to learn verb meanings. *Journal of child language*, 17(2):357–374.

Oxford Online Dictionary. 2025. Emoticon. Retrieved: 2025-03-11.

Patterson, Clare. 2013. Discourse coherence in pronoun resolution. *Discours*, 12. Mis en ligne le 10 juillet 2013.

Patterson, Clare and Schumacher, Petra B. 2020. The timing of prominence information during the resolution of german personal and demonstrative pronouns. *Dialogue & Discourse*, 11(1):1–39.

Peslak, Alan, Ceccucci, Wendy, and Sendall, Patricia. 2010. An empirical study of instant messaging (im) behavior using theory of reasoned action. *Journal of Behavioral and Applied Management*, 11(3):263–278.

Pohl, Henning, Domin, Christian, and Rohs, Michael. 2017. Beyond just text: semantic emoji similarity modeling to support expressive communication👫📲😀. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(1):1–42.

Polanyi, Livia. 1988. A formal model of the structure of discourse. *Journal of Pragmatics*, 12(5):601–638.

Primus, Nikolaus P Himmelmann& Beatrice. 2013. Prominence beyond prosody–a first approximation. *pS-prominenceS: Prominences in Linguistics.*, page 38.

Prince, Ellen F. 1978. A comparison of wh-clefts and it-clefts in discourse. *Language*, pages 883–906.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Syntax and semantics: Vol. 14. Radical Pragmatics*, pages 223–255. Academic Press, New York.

R Core Team. 2024. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rodríguez, Kepa Joseba and Schlangen, David. 2004. Form, intonation and function of clarification requests in german task-oriented spoken dialogues. In *Proceedings of Catalog (the 8th workshop on the semantics and pragmatics of dialogue; SemDial04)*.

Romiszowski, Alexander and Mason, Robin. 2013. Computer-mediated communication. In *Handbook of research on educational communications and technology*, pages 402–436. Routledge.

Rossano, Federico. 2013. Gaze in conversation. In Jack Sidnell and Tanya Stivers, editors, *The Handbook of Conversation Analysis*, chapter 15, page 308–329. Wiley-Blackwell, Oxford, U.K.

RStudio Team. 2025. *RStudio: Integrated Development Environment for R*. Posit, PBC.

Sacks, Harvey, Schegloff, Emanuel A, and Jefferson, Gail. 1974. A simplest systematics for the organization of turn-taking for conversation. *language*, 50(4):696–735.

Sampietro, Agnese and König, Katharina. 2024. The medium is accountable: Metacommunication and media ideologies about voice messages in whatsapp chats. *Discourse & Communication*, 18(1):51–71.

Schegloff, Emanuel A. 1982. Discourse as an interactional achievement: Some uses of 'uh huh'and other things that come between sentences. *Analyzing discourse: Text and talk*, 71(93).

Schegloff, Emanuel A. 1992. Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American journal of sociology*, 97(5):1295–1345.

Schegloff, Emanuel A. 1997. Third turn repair. In Gregory R. Guy, Crawford Feagin, Deborah Schiffrin, and John Baugh, editors, *Towards a Social Science of Language: Papers in Honor of William Labov*, volume 128 of *Current Issues in Linguistic Theory*, pages 31–40. John Benjamins. Published online: 23 May 1997.

Schegloff, Emanuel A. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1):1–63.

Schegloff, Emanuel A. 2007. *Sequence organization in interaction: A primer in conversation analysis I*, volume 1. Cambridge university press.

Schegloff, Emanuel A. 2013. Ten operations in self-initiated, same-turn repair. In Makoto Hayashi, Geoffrey Raymond, and JackEditors Sidnell,

editors, *Conversational Repair and Human Understanding*, Studies in Interactional Sociolinguistics, page 41–70. Cambridge University Press.

Schegloff, Emanuel A., Jefferson, Gail, and Sacks, Harvey. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53(2):361–382.

Schiano, Diane J., Chen, Coreena P., Isaacs, Ellen, Ginsberg, Jeremy, Gretarsdottir, Unnur, and Huddleston, Megan. 2002. Teen use of messaging media. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, page 594–595, New York, NY, USA. Association for Computing Machinery.

Schiffrin, Deborah. 1980. Meta-talk: Organizational and evaluative brackets in discourse. *Sociological Inquiry*, 50(3-4):199–236. First published: July 1980.

Schlobinski, Peter, Fortmann, Nadine, Groß, Olivia, Hogg, Florian, Horstmann, Frauke, and Theel, Rena. 2001. Simsen. eine pilotstudie zu sprachlichen und kommunikativen aspekten in der sms-kommunikation. *Networx Nr. 22*, 22.

Schlöder, Julian J and Fernández, Raquel. 2015. Clarifying intentions in dialogue: A corpus study. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 46–51.

Schneebeli, Célia. 2019. Gifs in online interaction: embodied cues and beyond. *Cahiers de l'ILSL*, 59.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423. First published: July 1948.

Shifman, Limor. 2013. *Memes in digital culture*. MIT press.

Siebenhaar, Beat. 2018a. Funktionen von emojis und altersabhängigkeit ihres gebrauchs in der whatsapp-kommunikation. In Arne Ziegler, editor,

*Jugendsprachen/Youth Languages*, pages 749–772. De Gruyter, Berlin, Boston.

Siebenhaar, Beat. 2018b. Sprachgeschichtliche aspekte der verwendung von bildzeichen im chat. *Sprachwandel im Deutschen*, pages 307–318.

Siegenthaler, Eva, Wurtz, Pascal, Bergamin, Per, and Groner, Rudolf. 2011. Comparing reading processes on e-ink displays and print. *Displays*, 32(5):268–273.

Sotillo, Susana M. 2000. Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*.

Spilker, Jörg, Klarner, Martin, and Görz, Günther. 2000. Processing self-corrections in a speech-to-speech system. In *Verbmobil: Foundations of speech-to-speech translation*, pages 131–140. Springer.

Spivey, Michael J, Tanenhaus, Michael K, Eberhard, Kathleen M, and Sedivy, Julie C. 2002. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45(4):447–481.

Statista. 2025a. Number of available emojis worldwide. Accessed: 2025-03-11.

Statista. 2025b. Whatsapp - statistics & facts. Accessed: 2025-03-07.

Stivers, Tanya. 2008. Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, 41(1):31–57.

Syahfitri, Rahma Yulia, Natsir, Muhammad, et al. 2021. Emoji in whatsapp group conversation. In *6th Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2021)*, pages 296–300. Atlantis Press.

TechTarget. 2021. What is video conferencing? Accessed: 2025-03-06.

Thompson, Sandra A. 1987. The passive in english: A discourse perspective. In Robert Channon and Linda Shockey, editors, *In Honor of Ilse Lehiste*, pages 497–512. De Gruyter Mouton, Berlin, New York.

Thurlow, Crispin. 2003. Generation txt? the sociolinguistics of young people's text-messaging. *Discourse analysis online*, 1(1):30.

Tolins, Jackson and Fox Tree, Jean E. 2014. Addressee backchannels steer narrative development. *Journal of Pragmatics*, 70:152–164.

Tree, Jean E.Fox. 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34(6):709–738.

Tschirner, Erwin and Jones, Randall L. 2006. *A Frequency Dictionary of German: Core Vocabulary for Learners*. Routledge Frequency Dictionaries. Routledge, London and New York.

Varnhagen, Connie K, McFall, G Peggy, Pugh, Nicole, Routledge, Lisa, Sumida-MacDonald, Heather, and Kwong, Trudy E. 2010. Lol: New language and spelling in instant messaging. *Reading and writing*, 23:719–733.

Voida, Amy, Newstetter, Wendy C, and Mynatt, Elizabeth D. 2002. When conventions collide: the tensions of instant messaging attributed. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 187–194.

Von Heusinger, Klaus and Schumacher, Petra B. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics*, 154:117–127.

Wagner, Petra, Malisz, Zofia, and Kopp, Stefan. 2014. Gesture and speech in interaction: An overview. *Speech Communication*, 57:209–232.

Wang, Weiquan, Zhao, Yi, Qiu, Lingyun, and Zhu, Yan. 2014. Effects of emoticons on the acceptance of negative feedback in computer-mediated communication. *Journal of the Association for Information Systems*, 15(8):3.

Warschauer, Mark. 1995. Comparing face-to-face and electronic discussion in the second language classroom. *CALICO journal*, pages 7–26.

Weiß, Martin, Bille, Dariana, Rodrigues, Johannes, and Hewig, Johannes. 2020. Age-related differences in emoji evaluation. *Experimental Aging Research*, 46(5):416–432.

Werry, Christopher C. 1996. Linguistic and interactional features of internet relay chat. *Pragmatics and beyond new series*, pages 47–64.

Whitehead, Kevin A. 2011. Some uses of head nods in "third position" in talk-in-interaction. *Gesture*, 11(2):103–122.

Yngve, Victor H. 1970. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567–578, Chicago, IL. Chicago Linguistic Society. CLS 6.