Using keystrokes to predict social dynamics in dialogue

Adam Goodkind

August 29, 2021

Abstract

When we speak to a person face-to-face, we gain information not just from the words being spoken, but also from prosodic elements of speech such as tone of voice and rate of speech. However, when we interact with people in a text-based environment, all of the information gained from non-lexical sources is lost. The aim of my thesis, though, is to locate the information normally contained within speech prosody in the typing production patterns. These patterns are typing analogs of prosodic variations in spoken language production. This "silent" form of prosody (Implicit Prosody Hypothesis) has been studied extensively in activities such silent reading where the voice in our head influences the way we perceive the text being read or produced. An overarching goal of my thesis will be to detect silent prosody in language production, not just comprehension.

By measuring fine-grained variations in typing patterns, I aim to understand latent social motivations behind the overt lexical representations produced in text-based communication. Using a chat interface that records keystroke timing, I will ask participants to discuss movie recommendations, in order to engage them in a semi-structured but free-wheeling conversation. I will then study the interaction at three levels of granularity: Study 1 will look at the dialogic function of individual utterances, to investigate how this function impacts the way that the utterance is typed. Study 2 looks at two adjacent utterances, or dyads, to see how the sentiment of each utterance as well as the lexical similarity between the utterances influences the way they are typed. Study 3 looks at the overall conversation. Using keystroke patterns, Study 3 will use keystroke features to predict the self-reported level of rapport that the participant felt during the conversation.

By focusing on these three aspects of interaction, my hope is I can detect detect social motivations and dynamics normally limited to speech studies. Eventually these findings can be used to create explicit representations of information lost without spoken prosody, in hopes of making text-based communication a more rich and informative form of communication. This can be immediately applied to improving human-human text-based interactions, and perhaps in the future allow for more successful interactions between humans and computer agents, e.g. chatbots.

Contents

1 Introduction **4**

2 Motivation and Contributions **7**

2.1 Why it's important to better understand text-based CMC 7

2.1.1 Ubiquity of Computer-Mediated Communication 8

2.2 Uniqueness of keystroke dynamics 9

2.3 Parallels to cognitive science 13

2.4 Contributions 13

2.4.1 Human-Computer Interaction 13

2.4.2 Cognitive Science 14

2.4.3 Keystroke Research 14

3 Related Work **15**

3.1 Explicit and Implicit Prosody 15

3.1.1 Implicit Prosody Hypothesis 15

3.1.2 Implicit Prosody and Keystrokes 16

3.1.3 Spoken prosody and the current studies 16

3.2 Keystrokes 20

3.2.1 Advantages of keystroke-based analysis 21

3.2.2 Keystroke features 21

3.2.3 Consistent versus variable features 24

3.2.4 Emotion and keystrokes 25

3.2.5 Deception and keystrokes 25

3.2.6 Cognition and keystrokes 26

3.2.7 Linguistics and keystrokes 26

3.2.8 Keystrokes in chats 27

3.2.9 Keystrokes for authentication or verification 27

3.3 Ethical issues with keystroke collection 27

3.4 Dialogue 28

3.4.1 General Discourse Analysis 28

3.4.2 Conversation Analysis 29

3.4.3 Sentiment analysis in dialogue 30

3.4.4 Disfluencies in dialogue 30

3.5 Dialogue Acts 31

3.5.1 Types of dialogue acts 33

3.5.2 Dialogue Act Detection Via Spoken Prosody 34
3.6 Coordination of dialogue 34
3.6.1 Why coordinate? 36
3.7 Rapport 37
3.7.1 Defining rapport 37
3.7.2 Measuring rapport 38
3.7.3 The complexity of rapport 38
3.7.4 Rapport and speech prosody 39
3.7.5 Rapport and cognition 39
3.7.6 Rapport and keystrokes(?) 40

4 Studies Overview 41
4.1 Studies Summary 41
4.2 Overall Methodology 45
4.2.1 Exploratory study data 45
4.3 Thesis data collection 45
4.3.1 Proposed experimental setup 46
4.3.2 Post-test questionnaire 47
4.3.3 Participant information 48
4.4 Overall Features 48

5 Study 1: Effects of Dialogue Act function on typing patterns 50
5.1 Introduction 50
5.2 Research questions 50
5.3 Hypotheses 50
5.4 Exploratory study 52
5.4.1 Pauses before the first word of an utterance 53
5.4.2 Pauses after the first word of an utterance 54
5.4.3 What I learned 54
5.5 Expansion in thesis 55
5.6 Contributions 56
5.7 Progress 57

6 Study 2: Typing patterns are most accurately modeled by the interaction of semantic similarity and sentiment 58
6.1 Introduction 58
6.2 Research question 59
6.3 Hypotheses 59
6.4 Exploratory study 60
6.4.1 Effects of sentiment and similarity on initial pauses 61
6.4.2 Effects of sentiment and similarity on revision counts 61
6.4.3 What I learned 61
6.5 Expansion in thesis 62
6.6 Contributions 63
6.7 Progress 63

7 Study 3: Typing patterns are affected by feelings of rapport with a conversational partner 64
7.1 Introduction 64
7.2 Research questions 64
7.3 Hypotheses 65
7.4 Exploratory study 66
7.4.1 Distribution of rapport 66
7.4.2 Revisions and rapport 66
7.4.3 What I learned 68
7.5 Expansion in thesis 69
7.6 Contributions 69
7.7 Progress 70

8 Contributions and Publications 71
8.1 Overall Contributions 71
8.1.1 Possible Publications 71

9 Timeline 72

10 Concerns and Next Steps 73

11 Bibliography 74

Chapter 1

Introduction

Every day, we communicate through computers on projects ranging from a group lunch order to office presentations to critical medical decisions. And every day, we also *mis*communicate through computers: We don't pick up on an intentionally humorous response, or miss the criticality of a request. This is made more frustrating because if these responses or requests were made in a face-to-face setting, these underlying intentions would be easier to pick up through tone of voice or the rate of speech, i.e. spoken prosody. My thesis aims to use timing patterns in typing, called *keystroke dynamics*, to detect these underlying motivations, and make the information normally available only in face-to-face interactions also accessible in a text-based interaction, where prosodic information is assumed to be lost.

This thesis explores text-based computer-mediated collaboration through multiple lenses. My thesis uses keystroke patterns because the typing modality of language production combines the spontaneous and dynamic elements of spoken language production with the static elements of finalized written text.

These production patterns can illuminate cognitive processes and social dynamics that are not overtly evident from surface-level or visible word choice, but are present in the latent patterns, and can provide insight into why a user is taking a certain action. While the aim of my thesis is to make these patterns more immediately beneficial to person-to-person interactions through a computer, they can also be expanded and refined for person-to-computer interactions, e.g. talking to a chatbot.

My studies will strongly rely on the Implicit Prosody Hypothesis (e.g. Fodor, 2002b, and expanded on in Section 3.1). This theory posits that even when not speaking out loud, we still use prosody when reading silently or typing, where "hearing" a voice in our head helps dictate comprehension. I plan to explore how language production, specifically typing patterns, parallels spoken prosody, so that the information contained in spoken prosody, e.g. altered meaning from a different tone of voice, can be transferred to a text-based environment.

As an example, when we speak to a person we dislike, we tend to use a different tone-of-voice, a different speaking rate, and make different word choices. In response to a terrible idea, we may say *That's a great idea!* but the phrase is laden with irony and its facetiousness is readily evident. In text-based communication, while we can sometimes use fonts (Heath, 2021), emoticons (Yuasa et al., 2006), or punctuation (Gregory et al., 2004) to convey prosody, often this information is lost. As a result, our understanding of our relationship with an interlocutor may be inaccurate and uncertain. My studies aim to understand underlying functions and emotions, in order to make text-based communication more multi-dimensional and better represent the true intentions of speakers.[1]

Study 1: An utterance's functional purpose in a dialogue will change the way the words are typed - When engaged in a conversation, some utterances are intended to make reference to previous remarks, e.g. a clarifying question, while other utterances are intended to progress the conversation forward, such as a statement introducing a new topic. The former are referred to as "backward" functioning, while the latter are "forward" functioning. Sometimes the same or similar words can function in both of these capacities. This can be confusing in a text-based communication context, since an interlocutor only receives the static word, rather than any inflection or timing patterns, which can be informative as to whether utterance is referring to previous utterances or is being used to introduce a new topic.

This study explores how the temporal aspects of typing production change depending on dialogue function. I hypothesize that backward-functioning utterances take longer to initiate, because the speaker needs to take into account all of the previous conversation. Conversely, after a backward-functioning utterance has been initiated, I hypothesize that the rest of the utterance will be more easily produced because the speaker is utilizing language that has already been introduced and cognitively activated. On the other hand, for a forward-functioning utterance that intends to alter the direction of the conversation, the speaker will need to monitor whether or not they have gained the ability to shift the topic, requiring more vigilance and cognitive effort throughout the course of an utterance production. This will result in longer pauses and mistakes after the production of the initial word.

This would parallel findings in studies such as Dhillon (2008), which found similar pause differences in spoken language production. If I am able to observe these parallels, it would demonstrate the correspondence between spoken prosodic difference and type-written timing differences.

Study 2: Linguistic similarity modulated by sentiment change most accurately predicts typing behavior - Study 2 combines linguistic similarity with sentiment analysis, where similarity between dialogue partners' language production is not evaluated in isolation, but rather is modulated by the sentiment of the utterance itself as well as the sentiment of the previous utterance. While linguistic similarity is often taken as a sign of connection between partners, I hypothesize that similarity is only a sign of connectedness when sentiment is also matched as well.

Since pauses and revisions could be signs of strain or uncertainty (Stromqvist, 2007), a less harmonious relationship could produce longer pauses and more revisions. However, these underlying feelings towards a partner cannot accurately be predicted from only sentiment or similarity alone.

To find evidence for this, I will compare the accuracy of different regression models for predicting typing patterns. More specifically, I posit that the most accurate model of typing patterns such as pause length and revision frequency comes from using an interaction of

[1] As a note on terminology, I will often use the term *speaker* to refer to the person producing language, whether they are speaking out loud or producing messages in a text-based environment.

similarity and sentiment (`typing` $\sim$ `similarity * sentiment`) as independent variables, as compared to models that uses only similarity *or* sentiment to predict typing patterns. This finding would help improve existing models of a user's mental state, where sentiment or similarity are usually considered independently.

Study 3: How typing patterns change depending on rapport between speakers - Whether conversing with a friend, a customer, or a medical patient, establishing rapport is important in order to achieve successful communication. However, rapport is very difficult both to define and to measure. Since keystroke patterns are strongly connected to both cognition and emotion, I posit that keystroke patterns can be used to make accurate predictions about underlying rapport.

This study aims to predict levels of rapport assigned by participants after a text-based conversation. Using keystroke patterns that are sensitive to cognitive effort as well as social motivations, I will build a model that splits words by their psychological categories (using LIWC), then connects production patterns in each category to a speaker's impression of the rapport level between themselves and their partner.

While my study does not aim to more succinctly define rapport, its findings would be useful for measuring the rapport of a conversation, as well as ways to identify signals of high or low rapport. Since high rapport is critical for everything from better educational outcomes to increased marketing success, it is important to be able to monitor rapport and possibly make adjustments in order to raise low rapport.

Each study investigates dialogue at different levels of granularity: single utterances, dyads of utterances, and entire conversations. Since my goal is to ultimately make the underlying motivations of latent typing patterns salient to a partner, it is important to only extract the most relevant data that is actually connected to those motivations, rather than extracting every pattern in typing production.

For example, I may find that only initial pause times are correlated to the dialogic function of an utterance (Study 1), while the delay before hitting Send and transmitting a message only correlates to change in sentiment in a dyad (Study 2). If I aim to accurately display the "state" of the conversation at any moment, then knowing which typing features to extract will be necessary to make accurate representations.

By making information currently hidden in typing patterns more salient, I hope to make text-based communication more multi-dimensional, allowing the more true meanings of text-based communication to become readily observable.

Chapter 2

Motivation and Contributions

2.1 Why it's important to better understand text-based CMC

The COVID-19 pandemic, and its effects on remote working, have added a tragic emphasis to the need for a better understanding of computer-mediated communication, as text-based CMC has come to occupy an even more central role in our lives. In a Pew Future of Work study from December 2020, researchers reported that 57% of respondents often or sometimes use chat-based platforms such as Slack or Google Chat (Pew Research Center, 2020). Interestingly, when compared to video-based platforms such as Zoom, text-based platform usage was consistent across educational and income demographics. On the other hand, for video-based communication platforms there was a clear divide with better-educated and higher-income individuals using these platforms more often.

This shifting platform usage has a strong effect on the work environment. The 2021 Work Index Trends report from Microsoft notes that a manager's role is increasingly centered around keeping a team connected and monitoring employees' well-being and mindset (Microsoft, 2021). However, Muir et al. (2017, p. 526) points out that when power relationships are asymmetric, e.g. a (higher-powered) manager and a (lower-powered) employee, "managers can find maintaining positive working relationships and good levels of rapport with virtual team members particularly challenging when relying on instant messaging to communicate." While it is impossible to constantly monitor every employee continuously, keystrokes provide a unique way to do just that.

As I will show both in my own studies and in prior studies as well, keystroke patterns are sensitive to behavioral states such as levels of stress and amount of cognitive load. A manager could use a dashboard that provides updates on employees' mental states, and the manager could then intercede if an employee becomes especially stressed or overwhelmed. Importantly, though, in comparison to straightforward email monitoring, keystroke patterns also offer anonymity, in that keystroke patterns can be analyzed independently from the actual lexical content of what is being typed (see Section 3.3). This can be thought about with an analogy to spoken language: if we hear a close friend speaking nonsense words, or a friend is especially happy or stressed, we can usually still tell that it's our friend, because of the uniqueness of their voice, or detect that they are happy or stressed even if the actual words are difficult to discern. Similarly, keystroke patterns can offer this same level of uniqueness. At the same time, insights at this level of cognition raise serious ethical and privacy issues. These will be addressed regularly in my thesis, as well as in my IRB.

As of this writing (August 2021), while more workers are returning to their physical offices, it seems possible that while Zoom usage will be replaced by face-to-face meetings, text-based communication will still retain its more central role given its wider usage in work settings. (It will be very helpful to revisit this, to see what researchers find, but we have already seen the popularity of platforms such as Slack even before the pandemic made a significant portion of work remote.) Regardless of how these trends continue to change in the immediate future, text-based CMC plays an important role in our day-to-day routines. This points to the importance of my thesis, as I aim to better understand the latent emotions and thoughts that lie behind the salient text that interlocutors see in text-based communication.

2.1.1 Ubiquity of Computer-Mediated Communication

To understand society's changing work and lifestyle settings even more profoundly, it is best to reflect on the title of Gergle (2017), "Discourse processing in *technology*-mediated environments" (emphasis added). Although "computer"- and "technology"-mediated environments are synonymous in many respects, subtle and less-subtle differences exist. The most apparent difference is that computer-mediated environments seem to conjure pictures of a user sitting down and using a laptop or desktop computer to communicate via a chat-based client such as Instant Messaging. Technology-mediated environments, though, seem to expand this picture to scenes such as controlling a smart TV with your voice, or engaging in a video chat on a smartphone.[1]

The difficulty, though, arises from the sources of information missing in these (non face-to-face) contexts. For example when chatting online with a customer service agent we do not have physical knowledge about the agent, such as facial expressions, and we cannot even be sure of their name or age. We use all of these to help make more informed decisions about how to communicate more efficiently. We also use information from prosody, such as rising or falling tones, as well as hesitations or rewordings (e.g. Snow, 1994; Shriberg et al., 2000; Fodor, 2002b; Trott et al., 2019). Given this, many researchers have taken the view that CMC is an impoverished environment (Walther and Parks, 2002). Gergle (2017, *inter alia)* provides a broad overview of many of these arguments. One of the main goals of this thesis, then, is to show that the information traditionally considered to be limited to face-to-face interactions or even audio/visual online conversations, is actually available in latent information available in keystroke patterns. By making this information more overt for partners, we can enrich the CMC environment to allow for more successful interactions and collaborations.

[1] Although new technologies such as smart TVs may use mobile-based, touchscreen communication, the studies in my thesis will all be done on desktop or laptop computers, with full QWERTY keyboards. Nonetheless, the importance of these studies is that they are first steps in understanding much more broad phenomena such as mobile communication. Once fixed-keypad communication is more thoroughly understood, it can be extended to (often noisier) mobile-based communication.

2.2 Uniqueness of keystroke dynamics

Before jumping into a defense of keystroke research, as opposed to research into other language modalities, it is important to preface this with an admission: Research into spoken language research is exploding, with good reason (Sisman et al., 2021). More and more companies are turning to voice assistants for customer service, troubleshooting, and even solving complex issues (Rozumowski et al., 2020). In combination with these developments, more and more products are adding the capability to control devices with voice commands, from cars to televisions to toaster ovens. As such, there are many good reasons to be pushing ahead with research into spoken-language conversational research.

That being said, we still use, and specifically type on, computers *a lot*. Statistics show that we spend upwards of 10 hours a day using technology, and it is not uncommon for technology use to be described as "nearly constant" (Twenge and Farley, 2021). Especially with a move towards more remote work arrangements, what was formerly "water cooler" chat can easily be replaced by an SMS text message or a direct message on Slack. To play on an old workplace cliche, many meetings *are* becoming emails.

Focusing specifically on improving CMC, keystroke research presents many advantages over research using other modalities of language production. To gain a better sense of this, Tables 2.1a, 2.1b, and 2.1c below compare typing research to other modalities of language production, which should illuminate the relative advantages of this type of research. In the tables below, "text analytics" consists of fully written text in a final static form. For example, this could be an essay on a test or an email sent to a receiver. This excludes research into the process of writing, where content is being dynamically produced, possibly by hand, e.g. pen and paper. "Speech research" consists of studying vocal language production, whether a monologue or an interaction involving two or more participants.

	Keystroke research	Speech research	Text analytics
Collecting data	A straightforward keylogger can be used to collect keystroke data. Both the key being used, along with when the key was pressed and released (timestamps) are relatively trivial to collect. Different keyboards and different language-specific keyboards can cause differences in timing data, as these may collect samples at different rates. For the most part, though, data collection is not obtrusive, and users do not need to make adjustments to their regular habits in order to enable data collection.	Data collection *can be* non-obtrusive, where all that is necessary to collect data is a microphone embedded in a recorder or computer. That being said, as Svec and Granqvist (2010) shows, because different microphones often have different sampling rates and sensitivity levels, inter-microphone comparison is often difficult if not impossible. Often, voice data needs to be collected in laboratories using the same microphone, with the subject seated at a consistent distance away from the microphone.	Collecting written data is perhaps the easiest form of data collection. A researcher could use something as trivial as Google Forms to collect the final written answer to a prompt. Straightforward web scraping could also be used.

Table 2.1a

A comparison of collection methods for keystroke, speech, and written data. Speech data is usually the most difficult type of data to collect.

	Keystroke research	Speech research	Text analytics
Measuring production data	Extracting timing data from a log of keystrokes is relatively trivial. For the most part, the timing between keystrokes is simply the difference between timestamps of two keystroke events, e.g. key-release of keystroke 1 to key-press of keystroke 2. Importantly, if two users overlap in when they are typing, it is trivial to disentangle the two (or more) streams of typing.	Extracting timing data from speech can be extremely complex and labor-intensive. The lone exception might be measuring the onset of speech or the cessation of speech, although even these metrics can be ambiguous (Abramson and Whalen, 2017). Measuring the phonemes or phonological elements within a word, though, is time consuming and often subjective. The most common procedure for spoken language research is to utilize a trained human annotator to go through the phonetic output of a speech file, and manually mark where each sound begins and ends. The average time for this form of measuring is 4 hours for 1 hour of speech (Bazillon et al., 2008; Novotney and Callison-Burch, 2010). The task multiplies in complexity when voices are overlapping.	Timing data is not available for most written data. For research such as Kalman et al. (2013), the timestamps when a message was sent are sufficient.

Table 2.1b
A comparison of measurement methods for keystrokes, speech, and written data. Keystrokes have timing data readily available, and this data is trivial to measure.

	Keystroke research	Speech research	Text analytics
Available features	Timing, edits, final version	Timing, edits, final version	Final version

Table 2.1c
A comparison of available features from keystroke, speech, and written data.

2.3 Parallels to cognitive science

Studies such as Goodkind and Rosenberg (2015); Plank (2016) and my third Qualifying Paper show that phenomena present in speech data are also present in typing data. Fodor (2002a) describes this as “silent prosody” or the Implicit Prosody Hypothesis (see Section 3.1 for a full explanation of the theory and terminology). This is the phenomena experienced when we are producing or comprehending language silently while still being affected by non-explicit prosodic contours. For example, when we are reading silently, we often still hear a voice in our mind, which adds prosodic contours to static text, and is reflected in the speed at which we read different parts of the text.

The same phenomenon applies to typing, where even though we are *producing* language rather than comprehending it, we still hear a voice in our mind as we type, whether we are interacting with another agent or producing a thesis in solitude, and this alters the speed at which we type and the pauses we take. My thesis will be able to study to what extent this inner-voice affects the way that we type, by uncovering parallels between spoken prosody and typing patterns.

This is important because (explicit) prosody, i.e. prosody that is audible in a spoken language or visible in a signed language, signals everything from importance of the words being produced (Swerts and Geluykens, 1994), to how much the words being produced are part of shared knowledge or common ground (Mushin et al., 2003), to when a speaker is planning to yield the floor to their interlocutor(s) (Gravano and Hirschberg, 2009). By looking at similar phenomena in Implicit Prosody, it becomes possible to understand which timing-related choices exist only in the mind, compared to those timing adjustments that are explicitly produced.

Just as prior research on typing, such as Priva Cohen (2010); Logan and Crump (2011); Plank (2016), has shown that findings in typing research can apply to the larger domain of cognitive science, the studies in my thesis can have the same type of benefit.

2.4 Contributions

I hope to make contributions to three areas of inquiry: Human-Computer Interaction, Cognitive Science, and Keystroke Research.

2.4.1 Human-Computer Interaction

My thesis is concerned with human-to-human CMC, as opposed to human-to-computer CMC. While the findings from my thesis may one day be able to be applied to technologies such as automated assistants and chatbots, both of these introduce new variables, and thus fall outside the scope of my thesis.

Rather, my thesis will focus on human-to-human computer-mediated interaction, and study the information that is present in the latent patterns of keystrokes. I can then use this to make social information more visible to conversation partners, for example by displaying a visualization of typing patterns, so that one speaker can understand that their partner is excited or confused. In this way, I can add missing, but important, information to text-based communication.

These findings can also be extended to settings such as moderating an online discussion forum. While a moderator may not be able to review every single post, my findings will allow moderators to detect questionable postings. For example, if the typing patterns of the poster seem indicative of deception, or extreme anger, this could raise a flag. The moderator could then review that post, or that user, more closely.

2.4.2 Cognitive Science

A central goal of my thesis is to provide additional information about the Implicit Prosody Hypothesis (IPH). IPH has been extensively studied in language comprehension, but not production. This is because most studies of language production are limited to spoken language as opposed to written language production.

By studying IPH in text-based communication, I hope to show evidence for whether prosody is useful only for verbal interaction, where we tailor the way we speak to the person we're speaking to (*audience design*, Horton, 2017), or also is useful for structuring our own thoughts, regardless of whether this structuring is observed by others.

2.4.3 Keystroke Research

Keystroke production is still a relatively understudied domain of language production, when compared to speech analysis and text analysis. Moreover, very few studies exist that investigate keystrokes in dialogue: most study keystrokes in monologue, or isolated settings such as writing essays or entering passwords. Thus, these studies will help to expand the domain of keystroke research.

In addition, the web interface I am designing with my research assistant will be valuable for running task-based or interactive keystroke studies in the future.

Finally, after my data is sanitized and anonymized, I hope to make it available to the broader community, which is still in need of more data for conducting research.

Chapter 3

Related Work

This thesis will bring together three areas of research that have thus far not been fully integrated: dialogue analysis, keystroke pattern analysis, and speech prosody. I will begin by setting up prosody, especially *implicit* or *silent* prosody. This theory of prosody will be essential to the other studies in my thesis, as it will help to tie together seemingly disparate parts. I will then introduce keystroke dynamics, providing a background as well as the diversity and depth of keystroke research. Following this, I will present relevant aspects of sentiment analysis, semantic similarity in dialogue, and the notion of rapport with a partner. Throughout, I will highlight relevant prior studies that have preceded the studies in my thesis, and briefly show how these studies can be expanded by the studies proposed in my thesis.

3.1 Explicit and Implicit Prosody

Prosody is defined as the element of language that occurs above the individual sound or phoneme level, and instead focuses on longer segments of language. For example, prosody studies the different tone of voice we use when making a statement versus asking a question. It also studies the rate of speech, such as why we enunciate a complex word more slowly and precisely. Finally, prosody looks at why certain words are said more energetically than others, and how a speaker decides at what volume to produce a word or sentence (e.g. Pierrehumbert and Hirschberg, 1990; Selkirk, 1995). Taken together, these aspects of language production are called "prosodic contours of speech," as they "shape" the language being produced.

3.1.1 Implicit Prosody Hypothesis

Almost all studies of prosody, though, investigate *explicit* prosody, i.e. sounds that are audibly perceptible. This thesis, though, builds upon the concept of *silent* or *implicit* prosody (Fodor, 2002b; Lovric, 2003). The Implicit Prosody Hypothesis (IPH) investigates the prosodic contours projected silently onto stimulus, e.g. during silent reading or when typing on a computer. The IPH says that the prosodic contours and boundaries which are audibly apparent also impact the speed at which we comprehend and read language. Researchers use procedures such as eye-tracking to observe minute changes in reading time that take place at specific points in a sentence where spoken prosodic variations usually occur. For example, Ashby and Clifton Jr. (2005) found that words with two stressed syllables (e.g. *ULtiMAtum*) are read more slowly than words with one stressed syllable (e.g. *inSANity*). The researchers take this as evidence that readers routinely assign stress patterns to silently read words. For an overview of other empirical observations see Breen (2014).

3.1.2 Implicit Prosody and Keystrokes

Keystroke analysis is well-suited to picking up on a user's implicit voice. As pointed out in Galbraith and Baaijen (2019):

> [S]peaking prevents the monitoring of inner speech. By contrast, writing, partly because of its slower output, but mainly because it is produced manually rather than vocally, allows-—indeed encourages—-monitoring of inner speech.

Combining IPH with keystroke research provides a unique opportunity for both keystroke research and prosody. Previous studies of dialogue collect data of two types: spoken conversations where timing metrics are available but laborious to measure, or text-based CMC where entire messages are analyzed, sometimes along with the time when the message was transmitted. To my knowledge, with the exception of a handful of studies such as Bukeer et al. (2019) and Roffo et al. (2014), no studies of keystrokes in dialogue have been performed. Previous studies have also been primarily aimed at using keystrokes to predict features of the typist, rather than analyze the conversation itself, as I propose to do in my thesis.

By studying keystrokes in dialogue, I can gain insight into the use of silent prosody during interactions rather than in isolated activities such as silently reading. Moreover, we can use findings from speech science to investigate phenomena in typing, since the two activities (production and comprehension) likely share a common pathway (Pickering and Garrod, 2013). This also provides a significant opportunity for HCI and text-based CMC: It is commonly assumed that when speech is absent from a conversation, as in a text-based dialogue, that some source of information is lost, or at least significantly altered. By investigating *implicit* prosody in text-based chats, I hope to capture aspects of sentiment and thought that were previously thought to be limited to *explicit* prosody.[1] As an example, hesitancies and revisions (that a speaker produces when they are stressed or confused) are evident in explicit prosody, but are difficult to detect in a finalized textual message. The keystroke patterns that go into the creation of that message, though, may provide evidence of the unstable thought patterns underlying the message.

3.1.3 Spoken prosody and the current studies

Finally, spoken prosody is relevant to all of the studies in my thesis, and so insights from speech could be very instructive for my own investigations of typing. All of my studies will be explained in more detail in further sections, and so this is just a very quick summary to set the stage.

[1] Before drawing direct analogies between speech and typing phenomena, I feel it is necessary to also introduce features of keystrokes that I will be studying. A table with these parallels can be found in Tables 3.1a and 3.1b.

- Study 1 looks at *dialogue acts*, and the unique characteristics of producing different types of dialogue acts. Stolcke et al. (1998) and many studies since have shown that prosodic properties of speech improve the accuracy of predicting the type of dialogue act.

- Study 2 looks at how sentiment and similarity between conversation partners affects the typing patterns of those partners. Studies such as Gravano et al. (2011) show how prosodic characteristics of speech affect social perceptions, while Levitan et al. (2012) shows how as interlocutors styles become more similar (coordination), their prosody changes as well.

- Study 3 looks at how well typing patterns can predict the rapport that the typist feels towards their partner. Lubold and Pon-Barry (2014) found that elements of spoken prosody can be used to facilitate detection of rapport levels.

The evidence above points to how important prosodic information can be in collaboration. In fact, it is also well-established that prosodic contours are important for successful communicative outcomes, e.g. Pierrehumbert and Hirschberg (1990). For that reason, providing additional temporal-based information to a text-based conversation can aid all participants involved in a conversation.

Finally, keyboard typing presents a unique opportunity for language and HCI researchers. On the one hand, it makes available many of the same linguistic production features available both from static and finalized written text as well as dynamic and spontaneous elements of spoken language Ballier et al. (2019).

Tables 3.1a and 3.1b below outlines the main prosodic features of spoken language, which studies the ways that the same lexical item can be produced differently, e.g. the same word said quickly or slowly, and then describes speech prosody's analogs in keyboard typing. As is evident from the last column in this table, though, not all speech prosody features have an obvious or explicit parallel in text-based communication. Alternatively, some spoken features *may* have parallels in typing, or may be manifested overtly, but in a different guise.

Feature in speech	**Manifestation in speech**	**Manifestation in typing**
Pauses	Pauses are usually only measured between words. Pauses between phonemes within a word are often difficult to impossible to measure. In fact, some phoneme transitions do not have any delimiting characteristics; rather, a speaker produces them in contiguous succession.	Pauses are relatively trivial to measure. Pauses can be measured in many ways, as seen in Figure 3.1. Pauses between intra-word keystrokes are typically measured in the same way as pauses in between-word keystrokes.
Energy/ intensity/ loudness	Speakers consciously and unconsciously choose how much energy to use in producing speech. These choices are usually directly perceivable by the listener, and is meaningful.	A typist can choose to alter the visual attributes of their message, such as all capital letters or bold font. The process of physically typing with more energy/intensity would not usually be perceived by the message recipient. It is possible that if the typist is silently producing more intense language prosody, this could be manifested as increased typos, revisions, or longer keypresses (Lee et al., 2015).
Length of sound/ duration	Syllable lengthening in speech is learned early in development and implemented for many different reasons (Snow, 1994). Measuring or at least comparing syllable duration is relatively robust in speech science.	Repeated letters are employed frequently in typing. Kalman and Gergle (2009) finds evidence for many uses of letter repetition, which parallel uses in spoken prosody.

Table 3.1a

A comparison of parallel features in spoken prosody and keystroke dynamics (continued in Table 3.1b)

Feature in speech	**Manifestation in speech**	**Manifestation in typing**
Speech rate	Speakers speed up and slow down for a large variety of reasons. The production rate of language, per se, can encode significant information about the intended message.	If a message is only transmitted upon completion, then the typing rate within that message is not necessarily known. If a number of messages are transmitted rapidly, it can be inferred by the receiver that the language is being produced rapidly. A real-time typing environment, which is less common today, would also facilitate awareness of production rate.
Pitch/ fundamental frequency	Humans continuously alter the pitch of their voice, e.g. high tones and low tones. These alterations can convey significant amounts of information about the affective or emotional properties of the speaker.	Aside from inferences drawn by the receiver from altered language production, pitch cannot be conveyed in typing production.
Timbre	Timbre is difficult to define succinctly, but it represents the quality of sound that makes a particular voice have a different sound from another, even when producing the same phoneme.	In CMC, the messaging medium outputs uniform text styling. Hand-written communication, such as shaky or sloppy text, could possibly be considered a parallel for voice timbre.

Table 3.1b
A comparison of parallel features in spoken prosody and keystroke dynamics (continued from Table 3.1a)

As further evidence of a direct connection between spoken prosody and typing patterns, Kalman and Gergle (2009) finds that the same types of sounds elongated in spoken prosody are also produced as repeated keystroke characters in typed text. Moreover, the authors find that much like the same speech prosodic contour can be used for multiple dynamic effects, typists employ repeated letters for different effects, as well.

3.2 Keystrokes

Keystroke studies enjoy a long history, going back to at least the 1920s Coover (1923). In World War II, Allied forces analyzed the unique production timing patterns of telegraph operators, called the "Fist of the Sender". Since each operator had a unique temporal signature, and individual telegraph operators travelled with specific troop battalions, this analysis allowed Allied forces to track the movements of different Axis troop units Banerjee and Woodard (2012).

But while keystroke dynamics has its origin in identifying individuals by the timing of the dots and dashes they produced, modern studies of keystrokes have exploded in both the breadth of human behavioral traits that are studied, as well as the fine-grained level of detail at which these areas are studied. This section will begin with an overview of how keystroke timing is measured, and how these measures are built into higher-level features. Following this, because my thesis will study complex social processes, this section will then show the diversity of areas that keystroke analysis touches upon, to set up why keystroke analysis is an ideal method to measure multidimensional behavioral patterns.

It is also important to explicitly mention that producing language on a traditional keyboard is still a highly relevant phenomenon that requires more detailed study. While speech-based interactions continue to grow in popularity, keyboards are still used on a near-daily basis in many facets of everyday life. As stated by Conijn (2020), quoting Brandt (2014):

> Writing is omnipresent in our society and plays, more than ever, an important role in our daily communication, work, and learning (Brandt, 2014). As Deborah Brandt puts it, millions of people (including myself) spend more than half of their working day "with their hands on keyboards and their minds on audiences" (Brandt, 2014, cover).

Keystroke analysis has also moved beyond QWERTY keyboards, and can be applied to tablets and smartphones that use touchscreens and swiping across multiple "keys" at once (Saevanee et al., 2012; Villani et al., 2006). This is important for the applications of my research, because many computer interactions are not limited to users sitting down at desktops: they can take place in conference rooms, on the go, or with each participant using a different modality. As an example of this diversity, a recent report from the Pew Research center found that 60% of Americans prefer to get their news from mobile devices, while the other 40% prefer desktop computers or television (Pew Research Center, 2019).

Further, keystroke analysis is not intrusive in the way that attaching sensors for galvanic skin response or an iris scan require significant interruption in activities (Fairclough, 2009). Rather, keystroke analysis can repeatedly and continuously measure a typist's behavior without any incursion into the daily keyboarding habits of the user (Vizer and Sears, 2017; Locklear et al., 2014).

3.2.1 Advantages of keystroke-based analysis

The primary advantages of keystroke research are that it is relatively inexpensive and unobtrusive to collect data, and relatively easy to analyze. As an example, one recent study analyzed 136,000,000 keystrokes from 480,000 participants (Dhakal et al., 2018). In comparison to speech analysis, prior studies have estimated that to transcribe an hour of speech data, it takes a trained researcher or professional anywhere from 4-10 hours (Novotney and Callison-Burch, 2010; Bazillon et al., 2008).

On the other hand, accurately determining the final text of a typing session is trivial, and timing measures such as pauses or keypress duration are not dIfficult to accurately determine in typing data (Dahlmann and Adolphs, 2007). Even when different computers with different keyboard layouts are used within a single experiment the measured timing differences are negligible (Pinet et al., 2017; Bridges et al., 2020).

Another unique advantage of keystroke analysis is that the logs will keep revision data completely intact. For examples, a user's final text might be "A bee," but the keystroke log might look like the figure below:

```
[SHIFT] [A] [SPACE] [B] [U] [G] [DELETE] [DELETE] [DELETE] [B] [E] [E]
```

In this case, we can easily recover the revised text. On the other hand, in spoken language production, if a revised word or phoneme was not fully articulated, it would be difficult-to-impossible to retrieve.

Similarly, natural dialogue contains a significant amount of overlap, where one speaker begins talking before another speaker has stopped talking (Heldner and Edlund, 2010a). Overlapping speech is much more difficult to transcribe than single-speaker speech, and just as with incomplete or corrected speech, it is difficult to extract meaningful timing data, such as pause timing.

This is important information to be able to recover because revisions contain valuable information about a typist. Lindgren et al. (2019) points out that a revision that is immediate versus a revision after a pause indicates different underlying cognitive processes.

Similarly, small revisions such as typos versus larger revisions such as correcting an entire idea also implies meaningfully different cognitive processes. Further, and pertinent to my thesis, Lindgren et al. (2019) points out that revisions can occur because of how a writer initially perceives the reader, or changes their perception of the reader, analogous to audience design principles (Clark and Murphy, 1982; Horton and Gerrig, 2016).

3.2.2 Keystroke features

Before proceeding in any keystroke study, it is important to isolate the features being studied, and why they are being studied. To give an idea of the wealth of features available in keystroke analysis, below I have reproduced Figure 3.1 from Conijn (2020), which provides a succinct illustration of fundamental available features, which can be combined and expanded

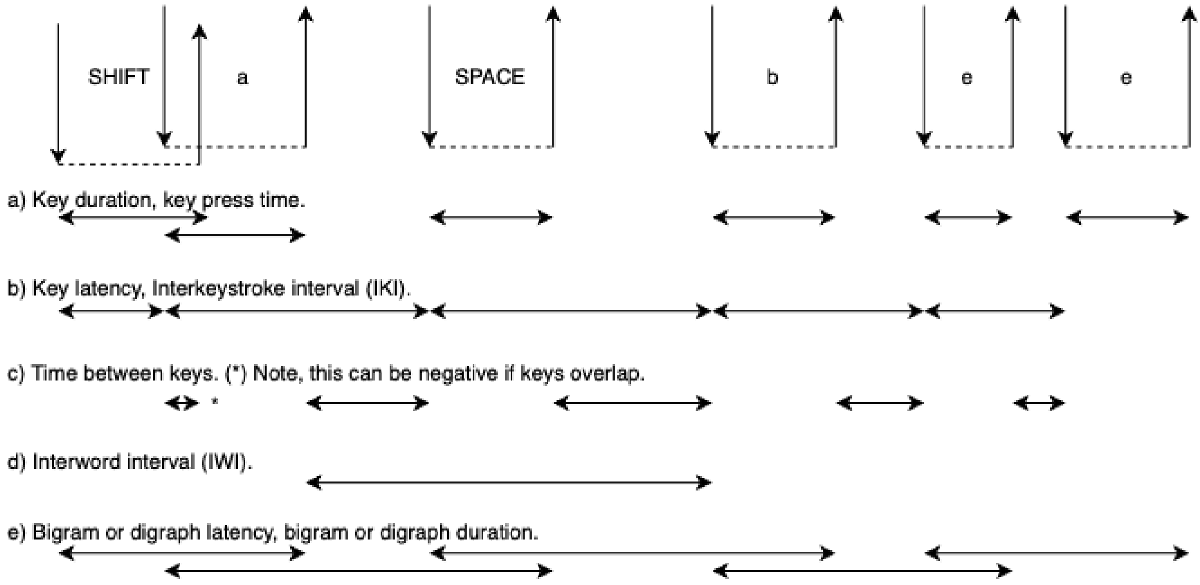

SHIFT
a
SPACE
b
e
e
a) Key duration, key press time.
b) Key latency, Interkeystroke interval (IKI).
c) Time between keys. (*) Note, this can be negative if keys overlap.
*
d) Interword interval (IWI).
e) Bigram or digraph latency, bigram or digraph duration.

Figure 3.1
A set of features available from extracted keystroke timing. Reproduced from Conijn (2020)

upon. There exist two primary dimensions that run through all of these features: the time interval *between* keystrokes, and the time duration for which a key was pressed. These features have analogs in speech data: the time taken to type a word is similar to the time taken to speak a word; the duration for which a typist holds down a key is similar to the intensity or loudness of speech.[2] Further details on these parallels, though, are unpacked in Figure 3.1.

Using the two dimensions of keystroke features (latency and duration), features of prior work can be organized into methodological categories. Table 3.2 is similar to a list of categories in Conijn et al. (2019); in addition I have also added a column with expected analogues in spoken language production.

[2] A more intuitive analog in speech to keypress duration would be phoneme or letter duration. However studies such as Lee et al. (2015) show that boredom and strong emotions affect keypress duration, because of the intensity of typing. Similarly, a highly emotional spoken phrase would be characterized by altered voice energy or volume.

Category	Examples in keystrokes	Analogues in speech
Pause timings or latencies	Interkeystroke intervals (IKI) between or within words (see e.g., Medimorec & Risko, 2017) or initial pause time (see e.g., Allen, Jacovina, et al., 2016)	Direct parallels in speech, where pauses between words and word duration are used. Pauses can be unfilled (silence) or filled (e.g. *um* and *uh*) (Clark and Fox Tree, 2002).
Keystroke duration	How long a key is held down for. The duration of a keypress is often associated with excitement and emotional response (e.g. Epp et al., 2011)	Energy in speech (manifested as loudness or intensity) is indicative of emotion and cognitive load (e.g. Mijic et al., 2017). It is important to note that keystroke duration does not parallel speech duration, e.g. elongated syllables. In typing, something like repeated letter, e.g. *hiiii* better parallels elongated syllables (Kalman and Gergle, 2009).
Revising behavior	The number of backspaces (see Deane (2013)), or time spent in revision (Goodkind et al., 2017).	Utterances are often repaired and restarted in the middle of a phrase. How often and where a repair exists is useful for inferring cognitive properties of a speaker Blacfkmer and Mitton (1991).
Fluency or written language bursts	Sequences of text production without interruptions, such as the number of words per burst after a pause or revision (see Baaijen et al. (2012); Van Waes and Leijten (2015))	Language learners, whether children or second language learners, often only speak a small sequence of words fluently, with a pause, and then a resumption of speech (Housen and Kuiken, 2009).
Verbosity	The number of words (see Allen et al. (2016)), or the number of unique words or lemmas (Goodkind et al., 2017)	The number of unique words used, and different *types* of words (e.g. nouns, verbs, etc.) are often measured in speech as a metric of cognitive development (Yu, 2010).

Table 3.2
Categories of keystroke features, along with possible parallels in speech production.

As mentioned above, one advantage of features such as those outlined in Table 3.2 is that they are also infinitely expandable. For example, rather than grouping all inter-keystroke intervals, a researcher can subdivide this feature into linguistically-delineated features, e.g., intervals-in-verbs, intervals-in-nouns, etc. Prior research has found this approach to be more accurate than approaches without subdivision (Goodkind et al., 2017; Brizan et al., 2015; Locklear et al., 2014). Further, a feature can be subdivided by its statistical properties, e.g. the mean of all measurements, the standard deviation, the minimum value, the top quantile, etc. (Kołakowska, 2018, 2015; Abadi and Hazan, 2020).

3.2.3 Consistent versus variable features

Keystroke production, like speech production, can be both consistent and varying, from speaking instance to speaking instance. For example, most speakers have a consistent and recognizable voice whenever they speak, just as typists have a consistent speed at which they type in all settings. But speakers also sound different when they are happy vs. sad, drunk vs. sober, or in the morning vs. the evening. The same differences apply to typing. As such, it is important to capture only meaningful differences in keystroke patterns that are induced by a change in stimuli, rather than normal variation.

Conijn et al. (2019) compared common keystroke dynamics features between tasks of varying cognitive complexity. By making this comparison, the researchers gained insight into which features are affected by the complexity of a task, versus features that stay constant within a subject but across tasks of varying complexity. The study found that features related to more broad properties of the writing composition, such as the number of words and total time, are sensitive to the task at hand. The researchers also found that pauses before words and revision features are sensitive to task complexity. Conversely, keystroke features such as the interkeystroke interval pauses within words varied less between tasks of different complexity, arguing that these represent innate behavioral characteristics of the typist

To complicate consistent versus variable features, though, the same observed behavior can have many different underlying motivations. As Schilperoord (2002) found, a typist might pause for a number of different reasons. A typist may pause because of cognitive overload, writing apprehension, or fatigue. Alternatively, a typist may pause because they were distracted by their phone or a bird outside. Medimorec and Risko (2017) then found that pause *location* is also important: pauses before individual words are indicative of ease of word retrieval, while pauses before sentences are due more to the complexity of the sentence being planned.

Because of all of the factors enumerated above, the data collection in my thesis will be semi-structured rather than open-ended, where users will produce spontaneous language, but they will be somewhat controlled by the discussion prompts and time limits. This approach is also more consistent with more modern methods of conversational studies (Clark, 1996), whereas the classic methodology in Conversation Analysis relied on “overhearing” naturally-occurring or spontaneous conversations (thereby without any experimental controls).

In my own experiments, the subjects will switch from being the recommender to the recommendee, and it will be interesting to see which which features change and which stay constant. This will help hint at the cognitive burden of each role.

3.2.4 Emotion and keystrokes

Prior research has shown that short pieces of text, such as blog entries, can be used to identify the emotions of a writer (Gill et al., 2008). However, just as emotion identification in spoken language is aided by a combination of text analysis and speech analysis, many studies have also shown that a combination of keystroke patterns alongside text analysis improves results (Kołakowska, 2013; López-Carral et al., 2019; Lee et al., 2015).

Emotion can be classified either discretely or continuously, and keystrokes seem to sensitive to both (Epp et al., 2011). In a discrete classification, emotions are categorical, e.g. happy, sad, neutral. In a continuous classification, emotions are evaluated on a two-dimensional spectrum: **valence**, or the degree of negativity or positivity, and **arousal**, the intensity of the evoked emotion.

From a discrete standpoint, a number of studies have looked at which keystroke features are most correlated with a change in valence and arousal. Lee et al. (2014) and Lee et al. (2015) used the same general experimental framework but presented emotional stimuli visually ad auditorily, respectfully. The studies found that emotional valence affected keystroke duration, where a more negative emotion led to longer keypresses, interpreted as less energetic responses. On the other hand, arousal affected typing speed, in that the more intense the emotion, the quicker a subject would type.

However, Lee et al. (2014) and Lee et al. (2015) had subjects simply type an identical sequence of numbers over and over again after being presented with stimuli. This does not seem to be a realistic or particularly meaningful response, and the authors even note that although the differences due to emotion were significant, the differences due to individual variability were more influential.

For a more realistic and open-ended response, López-Carral et al. (2019) presented subjects with emotional images, varying in emotional valence and arousal, and then had the subjects type captions for the images. Interestingly, the researchers found effects similar to Lee et al. (2014) and Lee et al. (2015), where valence was negatively correlated to keystroke duration and arousal was negatively correlated to typing speed. However, López-Carral et al. (2019) found much more significant influences.

The results of all of these studies seem to point to two takeaways. First, emotion, evoked in different ways, can affect keystroke patterns. Secondly, the more naturalistic a typing experience is, the more strongly emotion affects typing. This points to the utility of my experiments, in that a dialogue will present a more naturalistic setting than responding to individual stimuli or typing a sequence of numbers. It will be interesting to see if I find the same types of effects, but at an even stronger level. This could point to direct connections between emotion and language production.

3.2.5 Deception and keystrokes

Another important area of research that can utilize keystroke analysis involves deception detection. This has recently become more important due to “fake news” (Morales et al., 2020; Pisarevskaya, 2017; Conroy et al., 2015) and online harassment via deception (Zhou et al., 2008). These studies all use static, final text, though.

Advancements in deception detection from language production have been furthered by adding features related to speech prosody, such as tone of voice or pauses (Levitan et al., 2016). However, detection of deceptive language has been even further advanced using keystroke analysis, as it combines the notion of prosody with accurate and identifiable language (Borj and Bours, 2019; Monaro et al., 2018; Banerjee et al., 2014). Because keystroke patterns are sensitive to underlying cognitive processes, and a fabrication goes through a different cognitive pipeline than a truthful statement, these distinctions can be detected by measuring keystrokes.

3.2.6 Cognition and keystrokes

Some foundational models of cognition, such as Rumelhart and Norman (1982) actually used typing to create holistic models of the interaction between language production and motor control. These models have been refined over the years, and more recent models of cognition via typing are able to detect two distinct, hierarchical cognitive processes during typing production: an "outer" loop that controls cognition at the level of word retrieval, and an "inner" loop that controls intraword, letter-by-letter word execution (Yamaguchi et al., 2013; Logan and Crump, 2011).

One specific element of cognition that is often measured is "cognitive load," or the complexity of a task that determines the amount of cognitive processing necessary to undertake the task. Typing patterns can also be used to detect the amount of cognitive load a typist is under. Brizan et al. (2015) analyzed short answer typing sessions, where subjects were asked a range of questions, from simple recall questions to more complex analyses. The researchers were able to predict the difficulty of the question based on how the answer was typed, across a subject pool of more than 500 users. Similar to the discussion above about low-level keystroke features that stay consistent across tasks, Brizan et al. (2015) demonstrates that cognitive complexity seems to have a reliable effect, or create consistent differences, across users.

Similarly, Vizer and Sears (2017) created a *continuous* classification system to measure cognitive demand. Unlike many classification studies that only output a single classification at the end of a training instance, a continuous classification system is constantly updating and changing its predictions. In a situation such as a game or troubleshooting, cognitive demand will change, and so making predictions after all data has been collected is not necessarily useful or an accurate picture of changing demands.

3.2.7 Linguistics and keystrokes

Importantly for the studies in my thesis, keystroke analysis has also been shown to be sensitive to the same temporal and intensity patterns seen in spoken language. As noted in Ballier et al. (2019, p. 363), "It may not be the case that the variation of typing speed mirrors the variation of speech rhythm, but comparable grammars of chunking can be carried out for speech and keylog data."

This observation is important because it demonstrates that typing production also taps into the same cognitive processes manifested in speech or language comprehension. As an example, in psycholinguistics it has been repeatedly observed that more uncommon words, or words with lower frequency, are more difficult and take longer to comprehend and produce.

Along that line, Nottbusch et al. (2007) found that keystroke pause duration is correlated with both word frequency and word length.

In other studies, Plank (2016) found that pauses in typing correspond to boundaries in syntactic units (e.g. a noun phrase or verb phrase), and therefore can be used as a shallow syntactic parser. Similarly, Goodkind and Rosenberg (2015) found that typing patterns are sensitive to whether a word is part of a multiword expression or is a singleton. For example, the pauses around the phrase “muddying the water” would be more pronounced than the pauses around “sipping the water.”

3.2.8 Keystrokes in chats

Although the application of keystroke analysis to improve the conversational experience itself (in text-based CMC) is a novel contribution, this thesis will not be the first study to apply keystrokes to chats. Indeed, while the vast majority of keystroke studies test a typist in isolation, keystroke analysis of chats has proven useful for a handful of other goals.

Borj and Bours (2019) used keystroke analysis to identify liars in a chat. The central notion was that being deceitful required more deliberate effort and less natural thoughts, and this different mode of thinking was evident in different typing patterns. Bukeer et al. (2019) and Li et al. (2019) used keystrokes to identify gender in chats, while Roffo et al. (2014) found they could infer personality and identity in chats using keystrokes (although most of their features were based on lexical and stylographic textual features). What ties all of these studies, as well as my studies, together is the notion that typing patterns reflect innate features of a typist.

My studies also stand in contrast to these studies, in that keystroke patterns in chats will be informative about the cognitive state of each conversational partner, per se. To be more specific, keystroke patterns will reveal how each partner is thinking about, or viewing, their partner.

3.2.9 Keystrokes for authentication or verification

Finally, the most pervasive use of keystroke analysis has been as a “biometric” or personal identifier, to allow for securing a system (Monrose and Rubin, 1997a; Epp et al., 2011; Banerjee and Woodard, 2012; Locklear et al., 2014). Just as an iris pattern or fingerprint is unique to an individual, typing patterns are also unique identifiers for individuals. While my proposal will not go into detail concerning authentication, the main takeaway should be that keystroke production also represents an innate property of individuals, not unlike a voice.

3.3 Ethical issues with keystroke collection

This section should conclude with a discussion of the ethical issues surrounding keystroke analysis. As mentioned above, keystrokes can be a biometric marker, like a fingerprint or facial recognition. As such, collecting keystrokes without a participant’s knowledge would be ethically murky at best, but more likely strictly unethical. Moreover, it is relatively easy to collect keystrokes, as all major browsers allow extensions to keep keylogs without a user even giving explicit permission (Morales et al., 2020). Because keystroke patterns can reveal information such as gender, age, education level, and native language (e.g. Tsimperidis and Arampatzis, 2020; Goodkind et al., 2017), the information contained in our keystroke patterns should be protected.

All of the experiments in my thesis will obtain IRB approval, even though all participant identities are anonymized. It may be the case, though, that we notify participants only *after* the experiment that their keystrokes were logged, so that they are not self-conscious about their typing. This notification, though, would include the option to not share keystroke data if they object.

The importance of anonymization is especially relevant today, as technology firms devour enormous amounts of data and create massive open data sets. Because keystroke patterns can identify an individual, simply removing a proper name or email would be insufficient. This is specifically mentioned in Forsyth (2007), which was concerned with military-grade privacy masking. However, they acknowledge that names and usernames are often misspelled or abbreviated.

Nonetheless, recent advances in keystroke analysis have found success with "anonymizing" keystrokes, where the specific keys are unknown but only the typing rhythms overall are measured (Monaco and Tappert, 2017). As another attempt at further anonymization, Leinonen et al. (2017) instead seeks to automatically remove all traces of keystroke patterns, such as revisions and timestamps, in order to truly deidentify text.

The success of studies such as Monaco and Tappert (2017) also points to how powerful keystroke pattern analysis can be. Given that the verification of an individual can still be made from keystroke patterns alone, without the context of the actual keys or letters produced, this demonstrates the extent to which typing patterns and practices are an innate and reliable signal, similar to the vocal quality of each individual, where the timbre of a voice is consistent regardless of exactly which letter they are pronouncing.

3.4 Dialogue

🎮 Dr. Gwen, vaccinated! #BLM
@gweezlouise

Sentences are fake, utterances are real

8:51 AM · Apr 10, 2021 · Twitter Web App

3.4.1 General Discourse Analysis

The analysis of dialogue between multiple entities differs substantially from traditional linguistic analysis. By "dialogue," I mean spontaneous or quasi-spontaneous interactions between two or more entities, where the utterances of one entity bear some relationship to utterances of the other entities.[3]

In contrast, traditionally linguistics has concerned itself with planned, static written language that is independently motivated, with little-to-no interaction with other sentences. For example, a sentence such as "The cat the dog the man hit chased meowed." is of interest to those studying linguistic structure (namely center embedding), but would be very unlikely to occur in a spontaneous conversation, at least without significant pauses and pitch changes.

Before delving into dialogue at all, though, it is important to understand three key terms that constitute the hierarchical elements of a conversation: *turn*, *message*, and *utterance*. Ivanovic (2005, p. 80) provides succinct definitions of all three conversational constituents:

Utterance: This is the shortest unit we deal with and can be thought of as one complete semantic unit—something that has a meaning. This can be a complete sentence or as short as an emoticon (e.g. " `:-)`" to smile).

Message: A message is defined as a group of words that are sent from one dialogue participant to the other as a single unit. A single turn can span multiple messages, which sometimes leads to accidental interruptions.

Turn: Dialogue participants normally take turns writing/speaking, where a turn is made up of one or more messages.

Another interesting way to view this distinction is through the concatenation of two propositions. Kasher (1972) defines a sentence as "...a series of sounds that have a meaning." As a continuation, Krauss and Fussell (1996) shows that in a *dialogic* view of conversation, meaning emerges through the conversational process, rather than from a single sentence or single utterance, per se. Put another way, the utterances of interlocutors are tightly connected, and their meaning is shaped not only by the utterance itself, but also by utterances that were previously produced and may be produced in the future, including utterances from the speaker themself and from other speakers (Clark, 1996; Garrod, 1999).

In Section 3.2, the vast majority of keystroke studies are conducted with a single entity typing text in an isolated environment, whether engaged in free typing or fixed-text typing. While a handful of studies such as Roffo et al. (2014); Bukeer et al. (2019); Borj and Bours (2019) use keystroke patterns within conversations to identify deception or gender, my thesis will make a novel scholarly contribution in using keystroke patterns to analyze the dialogue itself, and the interactive process that emerges in a dialogue.

3.4.2 Conversation Analysis

The formal study of dialogue is known as Conversation Analysis (CA), and evaluates the unique dyadic nature of conversation. Conversation has been traditionally studied in naturalistic settings, such as a recording of an interaction between a telephone operator and an inquiring party (see Horton (2017) for an overview), rather than as a controlled experiment.

[3] This can of course cause problems, where it becomes difficult to disentangle the direction of influence. Niederhoffer and Pennebaker (2002, p. 347) describes the problem succinctly: "What Person A says at Time 1 influences what Person B says at Time 1. But what Person B says at Time 1 also directly influences what Person A says (in response) at Time 2."

In the studies proposed in my thesis, I aim to bridge this gap, where I will use controlled stimuli to prompt natural conversations.

Rather than studying individual utterances, conversations are studied at the pair-level, which contains an utterance from one participant and an adjacent response utterance from another participant (an "adjacency pair" or "dyad"). The unique feature of conversational analysis that is not available in monologic speed is "turn-taking." This is the notion that one participant speaks, and then another participant speaks. However, recent studies have shed light on the degree to which orderly turn-taking is an idealization, rather than a reflection of everyday conversation. Heldner and Edlund (2010b) and Levinson and Torreira (2015) have shown that as much as 30-40% of corpora contain overlap and prolonged pauses.

Because overlap is pervasive in conversation, typing analysis again provides a unique advantage. Whereas in speech research the process of disentangling overlapping speech is difficult, in typing data it is trivial to connect keystrokes to each interlocutor, and also measure the length of time that multiple speakers were simultaneously typing, or overlapping.

As may be gleaned from the above summaries of Conversation Analysis, there is a constant tension between studying speakers in dialogue at the dyadic level and the (individual) cognitive level. Recent analyses, though, have tried to synthesize these two levels of investigation. Brennan et al. (2010) looks at how language production and language comprehension are tightly coupled, and how having a conversation is the process of trying to understand the other while simultaneously adequately explaining one's own thoughts.

3.4.3 Sentiment analysis in dialogue

Most sentiment identification in dialogue has been performed on audio data (Yeh et al., 2019; Shon et al., 2021). The prosodic features used in these studies, though, do have analogs in typing patterns. Yeh et al. (2019) used loudness, pitch and duration, while Shon et al. (2021) found that classifying emotion using "semi-labeled" input from both speech and text, separately, improved the accuracy of their system. This is helpful for my own studies, since I will access to both timing information and textual information.

A key difference between sentiment analysis in self-written text and sentiment analysis in dialogue is that text analytics lacks context, in that there is little to no moment-to-moment coordination between the producer and their audience. Dialogue studies, though, highlight the "joint action" of language use (Clark, 1996). Bothe et al. (2017) points out that in capturing a more accurate picture of a dialogue, for instance more accurately labeling the function of a turn, not only is the sentiment of the turn itself important, but the sentiment *change* between turns is also important. This amounts to a key distinction that I plan on investigating: not only how the sentiment of an utterance affects typing patterns, but also how sentiment change affects typing patterns.

3.4.4 Disfluencies in dialogue

A disfluency in language production includes everything from (silent) pauses, to filled pauses (e.g. *uh* or *um*), to restarts where a word or phrase is replaced by another word or phrase (e.g. "I went to the mall, I mean, the store.") (Stolcke and Shriberg, 1996; Ferreira and Bailey, 2004). While these disfluencies may seem to be an exception, many studies have concluded that they are rather frequent. For example, Bortfeld et al. (2001, inter alia) estimates that 6 out of every 100 words are affected by disfluencies. Measured in another fashion, [CITE] estimates that $30 - 40\%$ of all utterances are affected by at least one disfluency.

The above statistics point to one reason why keystroke analysis is so useful when studying dialogue. In keystroke analysis, disfluencies are easy to detect and measure. Pauses are measured as the duration between keypresses or between words, while repairs are completely intact in a keystroke record, with the replaced text being unambiguously preserved. This can be contrasted to seminal studies of disfluencies in speech such as Fox Tree (1995), which did not measure occurrences of phenomenon like silent pauses, because they felt the instances were too difficult to identify consistently.

From a cognitive perspective, one can look to studies such as Dammalapati et al. (2021) for spoken language or Conijn et al. (2019) for keystroke production. Dammalapati et al. (2021) looked at the distributions of disfluencies relative to word likelihood and dependency distance, and found that more cognitively demanding words, i.e. those with lower transition likelihood or those with longer dependency distances, resulted in a higher proportion of disfluencies. Similarly, Conijn et al. (2019) also found that cognitive requirements affected disfluencies in typing.

Dammalapati et al. (2021) also looked at the duration and semantic type of the words preceding disfluencies. They found that words preceding disfluencies were usually drawn out, and were more often function words (as opposed to content words). Similar to Bell et al. (2009), they found that speakers use drawn out words or (inserted) function words as a crutch while trying to process or retrieve difficult words.

Building on all of these findings, my studies will also measure disfluencies rates and the rate at which they are repaired. Taking these measurements will provide additional insight into the cognitive demands on a speaker, and possibly show how that cognitive burden relates to conversational success.

3.5 Dialogue Acts

A dialogue act (DA) is a description of the pragmatic function of an utterance in a conversation (Dhillon, 2008), and must have a *sender* and *receiver*, or addressee (Bunt, 2005). A dialogue act goes beyond merely the syntactic and semantic content of an utterance, and focuses on what purpose the utterance serves as a means to achieving a successful conversational interaction. The "purpose" of an utterance might be to acknowledge what was previously expressed by a conversational partner, to achieve mutual agreement, or the converse, expressing disagreement in an attempt to establish a different understanding from what was previously uttered. A DA may also not reference prior utterances, but rather serve the purpose of advancing the conversation, by shifting topics or asking a question.

My studies are motivated by the idea that when two speakers are involved in a conversation, it is critical to not only understand *what* a person is saying, but also *why* they are saying something, i.e. the pragmatic function of the utterance (aka the "dialogic" function). This level of meaning is crucial because different utterances can include identical words, or lexical content, but can serve very different conversational purposes. For example, in (1), the conversation can be continued in two different ways, **B1** or **B2** below.

(1) **A**: We shouldn't go.
B: Yeh.
B1: But what about Sam?
B2: I agree with you.

In this example, "Yeh" can serve different functions: In continuation **B1**, "yeh" served as a means to acknowledge **A**'s statement, but after gaining control of the conversation, a qualifying question is posed. On the other hand, continuation **B2** shows that **B** used "yeh" as a means to express agreement. In a conversation where **B** does not continue their utterance, but rather only produces "yeh," it becomes very difficult to disambiguate meaning.[4]

This raises the question of how much cognitive effort is required to produce different dialogue acts. If a backward-functioning DA like **B1** takes into account the previous conversation, will a backward-functioning DA become harder to produce as a conversation precedes and more previous context accumulates? Or is only the previous turn taken into account? Ivanovic (2005), studying dialogue acts in instant messages, found that dialogue acts occur as adjacency pairs. This would imply that the effort required to produce a backward-functioning dialogue act stays constant throughout a conversation. On the other, Bothe et al. (2018b) found that only taking into account local context, e.g. only the previous two utterances, decreases the accuracy of dialogue act identification. Study 1 will help shed additional light on this question.

I will also investigate if the inverse holds for forward-functioning DAs, where the effort required to change topics increases or decreases throughout a conversation. Using the same observation from Ivanovic (2005) as above, I will investigate whether shifting topics, as a forward-functioning DA does, also require a consistent amount of cognitive effort, regardless of how much context has built up?

These questions were investigated for ACKNOWLEDGEMENTS (backward-functioning) and FLOOR GRABBERS (forward-functioning) by Bhagat et al. (2003) for a broad set of (spoken) prosodic features, and then by Dhillon (2008) specifically for pause duration. As mentioned in Section 3.4, for spoken conversations, after a forward-functioning dialogue act is produced, the speaker needs to expend cognitive effort to ensure that the conversation can actually proceed to a new topic. My plan is to first investigate if parallel phenomena are seen in text-based dialogue, and broaden the study to other backward- and forward-facing dialogue acts.

While studies such as Dhillon (2008) focus on pauses from a temporal aspect, it is also important to consider the cognitive reasons behind increased pause length. For example, if a speaker is pausing longer to confirm the status of the dialogue and confirm that they can now speak, this also implies that there is more cognitive effort involved as compared to a simpler acknowledgment of what has been said by an interlocutor. Given this increased use of cognitive resources, it seems reasonable to expect that other cognitive lapses such as typos are also more likely to occur.

Previous studies of the phenomena above have investigated spoken language production. But scenarios such as Example 1, with ambiguously functioning words, are even more problematic in an online, text-based dialogue environment. Since text-based communication

[4] It should be noted, though, that this example is overly simplistic. Most likely "yeh" would be followed very quickly by a continuation, thus reducing the need to evaluate utterance **B** on its own.

lacks the information traditionally only available in spoken prosody, e.g. tone of voice or drawn-out pauses, it is even more difficult to discriminate different meanings from identical lexical content (Walther, 2011; Gergle, 2017). This could present a communication obstacle in many online human-to-human and human-to-computer online environments, such as the following scenarios: An online customer service representative would want to know whether a customer is in agreement or simply trying to regain control of a conversation; this would determine how and when the representative should proceed. Although subsequent utterances in (1) may resolve pragmatic ambiguity, studies such as Gravano and Hirschberg (2011) demonstrate how precisely the timing of turn-taking is synchronized, and so an interjection a moment too late or too early can feel very unnatural.

3.5.1 Types of dialogue acts

The DA labels discussed in my thesis were created for the Switchboard Dialogue Act Corpus (SwDA) (Jurafsky, 1997; Stolcke et al., 1998). This corpus was created by hand-annotating the dialogue acts of a significant proportion of the conversations in the original Switchboard Corpus (Godfrey et al., 1992). The Switchboard Corpus consists of the audio and text transcription of approximately 2,500 conversations between sets of two partners, who were given a minimal topic prompt for discussion. In addition, the corpus is time-aligned at the word-level, where the timing of the ending of each word was mapped using supervised phone-based recognition.[5]

The tagset of dialogue acts was created based on the DAMSL (Dialog Act Markup in Several Layers) annotation scheme (Core and Allen, 1997). This tagset includes information about whether an utterance is forward- or backwards-functioning: forward-functioning is similar to traditional Speech Act Theory, in that the DA is defined by the effect the speaker hopes to have on the addressee; a backward-functioning DA denotes how an utterance relates to previous utterances, such as an answer to a question (Bunt and Romary, 2009; Keizer et al., 2002). The DAMSL tagset also includes information about the utterance's form and content, such as whether it relates to the larger topic under discussion or to the communicative process itself (Core and Allen, 1997).

The modified DAMSL annotation schema used for SwDA was created with the intention that it can be applied to purely text-based dialogue, without the need to listen to the spoken language (Stolcke et al., 2000, p. 344). However, the modified scheme can still be directly mapped back to the original DAMSL annotation, allowing for comparisons to other studies.

As an example of how an utterance's pragmatic function can differ from its lexical content, one can look to two different SwDA labels: ANSWER and AGREEMENT. These two labels could be assigned to two different utterances of identical lexical content, such as "Yes" or "Mm-hmm". However, the differentiating factor is what type of utterance these *follow*. A declarative statement following an assertion could be an AGREEMENT, e.g. "A: It's cold out. B: "Mm-hmm", whereas the same declarative statement following a question might be an ANSWER, e.g. "A: Are you cold? B: Mm-hmm".

[5] Of course an annotation style of this sort lacks pre-word and intra-word timing, as well as revision timing. My thesis posits that these durations are critical for understanding a speaker's intent.

3.5.2 Dialogue Act Detection Via Spoken Prosody

As is mentioned many times throughout this thesis, a number of parallels exist between spoken prosody and keystroke dynamics (see Table 3.2).

Many studies have been conducted that use lexical, syntactic, prosodic, speaker, and time-based features to detect dialogue act boundaries (Shriberg et al., 2009, inter alia). Importantly, though, Shriberg et al. (p. 214 2009) mentions, “In many such studies, [spoken] prosodic features have been shown to improve performance over lexical features alone...”

Further, earlier studies of prosody in dialogue act detection found that pauses and word duration were the most informative prosodic features (Shriberg et al., 1998). Similarly, Laskowski and Shriberg (2010) used lexical features and prosodic features to recognize dialogue acts. While the researchers found that prosodic features on their own were not very helpful, they did find that adding prosodic features to lexical features improved the F1 score of dialogue act recognition.

An interesting observation comes from Bhagat et al. (2003), which studied the prosody of utterances with the same lexical content (“yeah”) but belonging to different dialogue acts. Because they found significant prosodic differences, they conclude “. . . any prosodic features across dialogue acts cannot be attributed to differences in the distribution of specific words.”

In my studies I will look at the typing-based analogs to speech prosodic features, and show how these are unique to the dialogue act within which they occur. While some features of speech may be harder to parallel in keystroke analysis, features such as pauses and word duration are straightforward to measure.

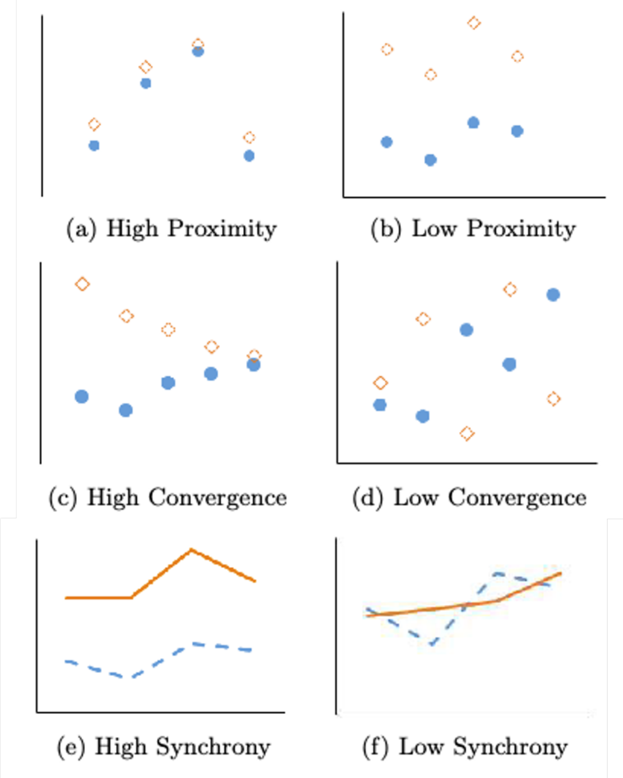
Dhillon (2008) conducted a study in which the same lexical item was measured in different dialogue acts, to see if the prosody was different in different types of dialogue acts. In this study, the pauses preceding and following identical lexical items were measured, where despite the lexical consistency, the dialogue act label of the utterances differed. Specifically, Dhillon (2008) looked at the terms “right” and “good”, when each was used as either a “floor grabber” to gain back control of a conversation, or to express “acceptance and agreement.” Overall, the study found that the pauses after a floor grabber are significantly longer than pauses after expressing agreement. These results speak to the importance of dialogue act recognition in (partially) determining cognitive strain, as pauses can be the result of mental processes requiring additional time or resources to process input and output, both in speech and typing (Chukharev-Hudilainen, 2014; Alves et al., 2007).

Finally, the types of models built from DA information also vary widely. Early efforts found great success in employing Hidden Markov Models to the task (Stolcke et al., 1998; Jurafsky et al., 1997). Other approaches have used Bayesian modeling of word predictability (Grau et al., 2005) and Latent Semantic Analysis (LSA) (Serafin and Di Eugenio, 2004). Recent approaches have used neural networks such as recurrent neural networks (RNNs) to classify dialogue acts (Khanpour et al., 2016; Bothe et al., 2018a).

3.6 Coordination of dialogue

The phenomenon of coordination between interlocutors, where a conversation becomes increasingly similar as it progresses, has been extensively studied at almost every level of language production. This ranges from low-level phenomena such as phonetic articulation, to syntactic structure, to broader semantic concepts (but see (Horton, 2017, p. 40) for an extensive list of studies). As a note on terminology, this broad aspect of conversational analysis is also known as "accommodation," "entrainment," and "convergence." To further complicate matters, some studies use one of these terms to refer to a specific type of similarity (as seen below). For the sake of consistency, I will refer to the phenomenon as "coordination," and clarify when a specific type of coordination is being discussed or measured.

Lubold and Pon-Barry (2014) makes a distinction between three types of coordination: proximity, convergence, and synchrony (see Figure 3.2).


(a) High Proximity

(b) Low Proximity

(c) High Convergence

(d) Low Convergence

(e) High Synchrony

(f) Low Synchrony

Figure 3.2
Three different types of speaker coordination as defined by Lubold and Pon-Barry (2014).

Each dot or line in Figure 3.2 represents one of two speakers at certain points in time.

- **Proximity** is the similarity between speakers at each point in time.
- **Convergence** is the degree to which speakers become more similar over time.
- **Synchrony** is how "in sync" speakers are throughout a conversation. Even though their language production at any one point might be very different from their partner, they still change at the same rate and in the same direction as their partner.

3.6.1 Why coordinate?

Beyond different types of coordination, there also exist different explanations as to why coordination happens at all. Explanations of coordination largely fall into two categories: *unmediated* and *mediated* (Branigan et al., 2011). Mediation here reflects the extent to which conversational partners influence, or mediate, the amount of coordination that occurs.

In an unmediated account, coordination is a product of the "linguistic environment," or the words and structures that have been produced and comprehended recently and further back in a conversation. According to (Garrod and Pickering, 2009), the process of coordination represents a bottom-up activation of representations during comprehension. More recent linguistic representations (e.g. words or syntactic structures) create stronger activations, and stronger activations directly shapes subsequent word choice and structure.

Most importantly for my thesis, (Garrod and Pickering, 2009) also points out that this account of coordination is cognitively economical and is automatic and effortless. In other words, coordination occurs independently of intention, wherein a speaker generally is not *trying* to coordinate. As similar words are used over and over again, they become more highly activated, and naturally synchronize the word choices in a conversation.

This points to a unique advantage of keystroke analysis of dialogue, where *effort* can be measured along with word choice. If a keypress duration or an interval before a word is longer, this strongly suggests that greater cognitive effort has been exerted. Whereas a static final transcript cannot reveal this, and it is laborious to perform the same analysis on spoken language, keystroke analysis is both informative about this and relatively straightforward to measure.

A mediated account, on the other hand, proposes that the addressee(s) strongly shape how a speaker produces an utterance, and will determine the rate of coordination (Branigan et al., 2011). This is also known as *audience design* (see Schmader and Horton, 2019, for an overview). Clark (1996), Clark and Brennan (1991), and Clark and Schaefer (1989) advance the notion that utterances are planned specifically to be best understood by interlocutors, which means that the beliefs and capabilities of a partner are taken into account during the planning phase. This process is more effortful, especially if a speaker determines that a partner has limited capabilities. In that case, a speaker will need to make significant adjustments to their planned utterance, and expend more effort, in order to accommodate the addressee (Branigan et al., 2011).

While coordination in a conversation is traditionally viewed as a positive signal, i.e. the interlocutors use coordination to express affinity, and therefore are feeling more strongly aligned with one another, it is important to point out that this is not always the case. As an example of how different types of coordination imply different signals, Niederhoffer and Pennebaker (2002) notes that semantic coordination is an indication that speakers are more engaged in a conversation, as this coordination is a demonstration of interest in the same topic. This points to the usefulness of analyzing language production patterns, such as pause times or edits, in addition to the linguistic context within which they occur. For example, a long pause when referencing the same topic as the dialogue partner may signal that a great deal of effort was dedicated to choosing language, and therefore that simply using the same word choice does not imply that the partners are in sync.

Going further, Scissors et al. (2009) demonstrates that not all coordination signals positive relationships in a conversation. Their methodology derives from James Pennebaker's seminal psycholinguistic studies (Pennebaker et al., 2003; Chung and Pennebaker, 2014), which found that certain understudied aspects of language also provide meaningful psychological information. These studies highlighted linguistic aspects such as function words (e.g. determiners and pronouns) instead of the more frequently studied content words (e.g. nouns and adverbs), as well as the psychological motivations for verb tense.

Scissors et al. (2009) points out that verb tense may reflect the period of time the speaker is focusing on, e.g. past tense implies focusing on the past, and that a positive sign of speaker affiliation is if they are both focusing on the same period of time. They find that when speakers coordinate their choice of verb tense as well as positive content words, this is usually a sign of a positive interaction. However, when speakers coordinate their use of negative emotion words, even though "coordination" is taking place, this does not signal a positive interaction.

Throughout my thesis I will repeatedly come back to Pennebaker's word classifications, as they are extremely useful for my studies. By connecting differing amounts of cognitive effort to different psychological categories, I hope to show that beyond just matching linguistic style, it is important to match the amount of cognitive effort in psychological categories.

3.7 Rapport

3.7.1 Defining rapport

Rapport is essential to establishing a good relationship, but is notoriously difficult to define. Tickle-Degnen and Rosenthal (1990), a seminal study on the conceptualization of rapport, defines it as "...an individual's experience of harmonious interaction with another person, often described as 'clicking' or 'having chemistry."' While this definition is comprehensive, and has been adopted by subsequent researchers, it still resorts to phrases such as "harmonious interaction" and "having chemistry," which are themselves difficult to define.

Lubold and Pon-Barry (2014) defines rapport in terms of a feeling of closeness, and many rapport-level questionnaires ask about a feeling of connectedness. Still other studies define rapport by the degree to which the partner "paid attention" to the subject, or the conversation was "enggrossing" and "worthwhile" (LaBahn, 1996). Regardless of definitions, it seems that researchers must repeatedly resort to abstract concepts in order to explain rapport.

These definitions point to the possible utility of connecting rapport to keystroke analysis. In Table 3.3 below, I highlight how keystroke timing could add a concrete measurement to some of the more abstract definitions of rapport.

The motivation behind why keystroke analysis could be a more accurate method to measure rapport comes from Chung and Pennebaker (2014, p. 5):

> A consistent finding is that many of the word categories used to reliably classify psychological states can be considered to be a part of language style as opposed to language content. That is, **how** people say things is often more revealing that **what** they are saying. [Emphasis added]

That being said, perhaps like many aspects of life, rapport can only be reduced to “I know it when I see it.” (Vogel, 2010). The motivation for my study, though, is to make rapport easier to see.

3.7.2 Measuring rapport

Early attempts to quantify rapport mostly focused on the relationship between psychotherapists and their patients, since rapport is critical to building a productive therapeutical relationship. Anderson and Anderson (1962) looked at quantifying rapport by looking at the proportion of matching definitions between a therapist and client, where higher rapport was correlated with a greater proportion of matching definitions, since both parties saw certain concepts in the same light. Relying on word-matching, though, is a crude estimate; by using linguistic and psychological categorization in my thesis, I hope to be able to capture more robust similarities.

In terms of the actual classifications of rapport, two primary methods exist: self-reports from interaction participants, and external observers who assign a rapport rating to an interaction. My thesis will rely on subjective self-reporting, which is a common way to measure rapport. As examples, Nomura and Kanda (2014), Hagad et al. (2011), and Bernieri et al. (1996) rely on self-reporting and surveys of public opinion to quantify rapport and found strong agreement between participants when self-reporting rapport in similar experimental situations. However, accurate impressions of rapport can also come from external evaluators, even when relying on very brief snippets of a conversation. For instance, Madaio et al. (2017) found high levels of inter-annotator agreement when viewing slices of conversations as small as 30 seconds.

3.7.3 The complexity of rapport

Seo et al. (2018) highlights the complex interaction of individual verbal behaviors that contribute to an increased feeling of good rapport. For example, asking an off-topic question, such as personal information, can increase rapport in certain settings, whereas in other settings or at the wrong time it can seem rude. As such, measuring straightforward semantic similarity between turns would not be a sufficient surrogate for rapport estimation. Similarly, some statements or questions are properly responded to with short responses, while others have more appropriate long responses. Therefore, in this case, measuring simple turn length, or turn length similarity, would also be insufficient.

The model devised by Tickle-Degnen and Rosenthal (1990) provides an attractive apparatus for experimentation, as they break rapport down into three dimensions: *attentiveness*, *positivity*, and *coordination* (not to be confused the term I introduced in Section 3.6). But they also found that each dimension does not exude equal influence on rapport throughout the course of a conversation. For example, early in an interaction positivity and attentiveness are most important for establishing good rapport; in the later stages of an interaction coordination and attentiveness are more influential on good rapport.

The findings from studies such as Tickle-Degnen and Rosenthal (1990) and Seo et al. (2018) point to the importance of studying social interaction at both a higher- and lower-level. A feeling such as rapport is multidimensional, and is made up of dimensions such as appropriate response type (dialogue acts), as well as sentiment expressed and word choice as it relates to what has been said previously (semantic similarity). Finally, as described below, rapport can also be communicated in spoken conversations through prosodic differences.

For these reasons, the multi-dimensional approach I will take, using production patterns and psychological categories seems most apt to capture rapport.

3.7.4 Rapport and speech prosody

In a study highly germane to my own thesis, Lubold and Pon-Barry (2014) studies how "prosodic coordination" correlates to rapport. (In this case, "coordination" is synonymous with "coordination," which I use throughout my thesis.) Interestingly, they find the strongest connection between rapport and prosody to be on a turn-by-turn basis, rather than across an entire conversation. In other words, high rapport comes when each pair of turns is similar, but overall patterns are not (necessarily) similar. Nonetheless, they still do find overall coordination to be informative as relating to prosody, as well.

This will be interesting to test in my thesis. I can look at the degree of correlation between typing patterns and rapport on a turn-by-turn basis, and also across an entire conversation. If my findings parallel those of Lubold and Pon-Barry (2014), it would suggest that rapport levels can be assessed continuously, at each turn, rather than only at the end of a conversation.

An additional distinction is made in Michalsky and Schoormann (2017), which studied not only overall pitch coordination in a conversation when speakers find their partners more socially attractive, but also at the individual level: how much does a single speaker try to match the pitch of their conversational partner if they find them to be likable and socially attractive? They observe that pitch coordination is affected by individual perceptions, and therefore it is not a mutual process, but rather is a signal of individual feelings towards an interlocutor.

This will also be interesting to investigate in my own studies. I can look at whether rapports levels are always similar and how matching rapport levels affect typing. Similarly, when a large mismatch exists in rapport levels, how does this asymmetry affect typing? Given the findings from Michalsky and Schoormann (2017), it may be that a speaker who finds their partner to be more socially attractive will put greater cognitive effort into matching their partner's style, which will translate to more cognitive lapses on matching words.

A succinct summarization of many of these results is provided by one of the conclusions in Michalsky and Schoormann (2017), which found that "prosodic [coordination] seems to be more than a mere automatic categorical adaptation to an interlocutor in social interaction but is instead dependent on social variables."

3.7.5 Rapport and cognition

Finally, it seems that rapport is ideal to study through keystroke analysis, given the sensitivity of this type of analysis. Barnett et al. (2018) and Barnett et al. (2020) found that when an examiner intentionally established either high or low rapport with a subject, even though they did have meaningful interactions with the subjects, the level of rapport affected

Definition (source)	Findings from keystrokes and prosody
"... an individual's experience of harmonious interaction or 'clicking' ..." (Tickle-Degnen and Rosenthal, 1990)	Pauses during typing, as well as increased mistakes, are associated with increased cognitive load or strain. In a "harmonious" interaction, one would expect fewer prolonged pauses and fewer mistakes, because the subject is more comfortable with expressing their thoughts.
"mutual understanding ..." (Saidla, 1990), "the perception of having established similarity with another person" (Nickels et al., 1983), "... the sense of closeness [between speakers] ..." (Lubold and Pon-Barry, 2014)	Mutual understanding is, among other signifiers, signaled by coordination at many levels, including speech prosody. By studying semantic and cognitive effort similarity in typing, I will also gain an understanding of which occurrences of similarity are indicative of a positive relationship, and hence could be the result of higher rapport.
"... the perception that a relationship has the right 'chemistry' and is enjoyable." (LaBahn, 1996)	Multiple studies have found that typing patterns are sensitive to a typist's emotions. Notions such as enjoyment seem to fit this category, and are therefore likely to be perceptible in typing patterns.
"engrossing... involving... worthwhile..." (Grahe and Bernieri, 2002)	Studies of the connection between keystrokes and emotions also study the intensity of emotions, not just the positivity/negativity of the emotions themselves. It seems that a conversation that is engrossing or involving will evoke more intense and less apathetic contributions.

Table 3.3

Definitions of rapport and possible quantification by keystroke patterns

performance on cognitive tasks such as the Stroop test and word association tests. In these investigations, it was found that high rapport improves results on cognitive assessments.

Findings such as those by Barnett and colleagues seem to underscore the importance of studies in my thesis. If keystroke analysis can provide accurate predictions of perceived rapport, and rapport helps improve cognitive functioning, then increasing rapport will not only create a more agreeable interaction, but also a more productive interaction.

3.7.6 Rapport and keystrokes(?)

Nonetheless, a fair question to ask is "Why is typing analysis good for rapport prediction?" I'd like to highlight some (abstract) definitions of rapport and demonstrate how typing analysis can address these abstractions.

In light of the evidence presented here, it seems that keystroke patterns should be sensitive to many aspects of rapport that are difficult to define concretely, but will be straightforward to meansure.

Chapter 4

Studies Overview

4.1 Studies Summary

Tables 4.1a, 4.1b and 4.1c provide a summary of all of the studies I am proposing. They present a very high-level overview, and may not make sense in isolation.

Study	**Method and Task**	**Data Collection Status**	**Completed Analysis**
(1) Dialogue Acts	• Recruit two participants to discuss movies • Each provides movie recommendations to the other, after getting to know each other • Each participant will provide recommendations for 5 minutes	• Pilot data available from similar Liebman studies • Interface currently being built to collect movie recommendation conversational data	• Pilot study using Liebman data (see Section 5.4) • Shows differences in pauses for forward- vs. backward-facing dialogue acts
(2) Sentiment and Similarity	Same as above	Same as above	None
(3) Rapport	Same as above	Same as above	• Minimal study using Liebman data. See Section 7.4 • Promising results showing consistency in rapport even in small-scale and unstructured conversation

Table 4.1a

Study	**Features**	**Estimated participant count (n)**	**Unit of Analysis**
(1) Dialogue Acts	• Pause before turn starts • Pause after first word • Revision count/length • Pause before revision (Collectively, "cognitive lapses") • Typing rate (keystrokes/second)	150 (Tiong and Lee, 2021; Monaco, 2021)	Dialogue acts are assigned to individual turns
(2) Sentiment and Similarity	• All features above • Inter- & intra-word pause duration • Keypress duration • Length of typing burst • Stylometric features (word length, utterance length, etc. • Pause lengths by semantic type, e.g. function vs content word"	150	The similarities and differences between two adjacent turns, or a dyad, will be measured.
(3) Rapport	• All features above • Extend feature engineering by using other properties of features above, e.g. SDs and quartiles • Extend features using n-gram combinations • Self-reported impressions of partner, from questionnaire, and combinations of these impression ratings	150	Participants rate rapport only for the entire conversation

Table 4.1b

Study	**Research Questions**	**Proposed Analysis**	**Hypotheses**
(1) Dialogue Acts	• Do typing patterns change depending on the conversational function of the utterance being typed? • What can we learn about cognition in dialogue if conversational function dictates pause and revision times?	• Typing Feature ∼ Dialogue Act • Compare the effect of dialogue acts on typing patterns • Use an ANOVA-style approach • Control for lexical variables that are known to affect language production	DAs that require more context to be taken into account (e.g. a backwards-facing question) will result in more cognitive lapses because of additional relevant information.
(2) Sentiment and Similarity	• Does the sentiment of an utterance change typing patterns? • Does the semantic similarity of an utterance to previous utterances change typing patterns? • Does sentiment *change* between a previous utterance and the current utterance affect typing patterns? • Do sentiment and similarity of a statement interact to dictate typing behavior changes?	• Typing Feature ∼ Similarity : Sentiment • Measure the effect of variables relating to sentiment and similarity on typing patterns, including interactions • Use a linear (mixed) modeling approach to study comparative strength of coefficients	Both sentiment and similarity will affect typing patterns, but the effects will interact. Typing patterns will be jointly modulated by how much the typist is changing the sentiment of a dialogue, as well as how familiar the words are to previously seen words.
(3) Rapport	• Can self-reported ratings of rapport be predicted from typing patterns? • Which typing features and psychological categories are most strongly connected to rapport ratings?	• Rapport ∼ Typing Features • Use a number of typing-derived features to predict rapport ratings • Use a neural-network or similar high-dimensional supervised ML model to create predictions	Typing will be able to make more accurate predictions of rapport than linguistic features alone, because typing patterns are connected to many cognitive and production properties and simultaneously can be very accurately measured.

Table 4.1c

4.2 Overall Methodology

Throughout the chapters on my studies, two different datasets will be referenced. To avoid confusion, below I will outline the different data being used.

4.2.1 Exploratory study data

My exploratory study data was originally collected for use in Liebman and Gergle (2016a,b), which investigated the use of social cues in CMC, and linguistic coordination in online chats, respectively. These studies used a customized instant messaging client called the Dialogue Experimentation Toolkit (DiET) (Healey and Mills, 2008; Healey et al., 2018), which also allowed experimenters to manipulate the messages by modifying the transmitted text before the partner saw it.

In the original experiments, 60 pairs of university students, who did not know each other, were seated in front of computers in separate rooms. In order to generate a more lively conversation, the participants were tasked with discussing a moral dilemma, such as learning that a friend's romantic partner was being unfaithful, as in (2). Participants were also told to take a certain stnace on the issue. Beyond these initial prompts, the conversations were freewheeling, and were either 5 or 15 minutes in length.

(2) You and your partner have a mutual friend who is engaged to be married in two months. You both just learned that your friend's fiance may have recently cheated on your friend. You and your partner are discussing what to do with this information. The goal of your discussion is to try to agree on what to do in this situation. For the purposes of this discussion, your opinion is that you need to tell your friend what you heard.
For the purposes of this discussion, your opinion is that you need to tell your friend what you heard.

The DiET toolkit recorded the timing of every keystroke press as well as the final text that was transmitted to the conversational partner. The partner only saw the final text; they were unaware of any revisions, additions, or deletions made before the message was sent, either in the course of a participant producing a message or subsequent experimental modifications.

4.3 Thesis data collection

While the pilot study's experimental setup was successful for the purposes of the prior studies, I feel it will fall short in my studies for two reasons:

1. The conversations were abstract and did not have any real buy-in from participants. While the conversations did raise ethical issues, and it is likely that participants could relate to the scenarios on some level, the scenarios did not necessarily reflect any lived experiences. Because keystroke-level analysis can be more fine-grained than lexical-level analysis, it is important that memory and opinion retrieval represent actual previous experiences or situations. Conjuring a hypothetical experience likely involves

different cognitive pathways or processes, which would be reflected in temporal differences in language production Aldridge and Fontaine (2019).

2. I want to avoid asking participants to champion a viewpoint they do not necessarily believe, as this also affects cognition. The reasoning is similar to that raised above in (1). Specifically, arguing a viewpoint that a participant does not agree with can be thought of as a form of deception. Multiple studies have shown that typing production is sensitive to deception, and can essentially be used as a lie detector in that it can pick up differences between how participants produce a true sentence versus a false sentence Banerjee et al. (2014); Monaro et al. (2018).

3. Typing about hypothetical emotions, or the emotions of others, e.g. "the person taking part in this conversation" is not the same as typing about one's own emotions. Seminal studies such as James Pennebaker's *The Secret Life of Pronouns* show that different pronouns imply different underlying psychological connections, showing that if a person not truly typing about themselves, then they are not employing the same mental processes (Pennebaker, 2011).

That being said, a brief analysis of keystroke data from Liebman and Gergle (2016a,b) *does* show that their more abstract tasks were able to evoke meaningfully consistent patterns in typing production. To clarify, if a task did not evoke any consistent cognitive processes that are used in language production, then we would expect the data to be random noise, since similar cognitive processes would not be utilized from utterance to utterance or between different participants. Rather, I do see slight evidence that typing patterns do reflect more genuine emotions or thought processes. These results, which are spelled out below, are promising though not conclusive. They point to the need for both more data points, as well as a more structured task that should yield more consistent results.

In order to avoid the issues raised above, but still keep the conversation lively where strong opinions might be expressed, my proposed task will focus on movie recommendations. In many ways, my own study will parallel the study performed in Pecune et al. (2019). However, whereas Pecune et al. (2019) tested human-computer agent interactions, my studies will test text-based human-human interactions.

4.3.1 Proposed experimental setup

Participants for my studies will be recruited via Amazon Mechanical Turk or a similar subject pool. The purpose of this is two-fold: 1) Participants are likely to have less in common, whereas in the pilot study all participants were undergraduates at Northwestern. By having less in common, a more lively conversation could be generated from more divergent backgrounds and viewpoints. 2) The models I would like to train require more data than that collected from the pilot study. An online setup will scale well and lead to easier data collection.

Using the JATOS toolkit (Lange et al., 2015), I can coordinate an experiment that requires two participants simultaneously. Once two participants have entered the experiment, they will see instructions as well as a chat window. This fairly standard chat window will allow a participant to see a full chat history, including what the other person transmitted, as well as the messages the participant themselves have transmitted as well. A scenario will be presented on the screen, which both participants will see, such as:

(3) [PARTICIPANT1] has had a long week at work, and would like to relax and watch a movie to unwind. [PARTICIPANT2], what movie or movies would you recommend and why?
Feel free to get to know each other, your tastes in movies, and discuss why you've recommended these movies. Do not hesitate to express opinions, for example about what you like or don't like about certain movies or movie genres, or certain actors and actresses.

After completing the first recommendation conversation, another prompt will appear, which will generate an additional conversation in the "opposite" direction, where the recommendee becomes the recommender. An example prompt is provided below. By also using a different prompt, the hope is that the second conversation will not simply mimic the structure of the previous conversation.

(4) [PARTICIPANT2] is bored, and would like to watch a really thought-provoking or stimulating movie. [PARTICIPANT1], what movie or movies would you recommend and why?

These prompts for generating interactions should get at the essential phenomena I would like to investigate, such as organizing a conversation and forming opinions of a partner.

1. Participants should use a diversity of dialogue acts, since clarifying questions will need to be asked, and forward-facing dialogue acts will be used to advance the recommendation conversations. This should provide a wide range of dialogue acts for Study 1.

2. From anecdotal experience, most people have strong opinions of movies and movie genres. Since Study 2 aims to not only look at lexical choice, but also sentiment, it is likely that a discussion of movies could evoke both strong agreement and strong displeasure.

3. While "rapport" is hard to define, it seems that these conversations will be effective at evoking or necessitating rapport. At the very least, the conversations should give each participant a sense of the rapport level established during the interaction. I will also get a better sense of rapport by asking in the post-test questionnaire "Will you watch any of the movies that [PARTICIPANT1] recommended during your conversation?" or "Do you think you would enjoy watching a movie together with [PARTICIPANT1]?"

4.3.2 Post-test questionnaire

I will model my questions using a number of well-tested questions that get at subjective measures of rapport, trust and likeability (Liebman and Gergle, 2016b, inter alia). Liebman and Gergle (2016b) also used Principal Component Analysis (PCA) to determine which question or questions were most informative. Their tests found a single factor existed, and so they averaged the ratings from multiple questions. I can perform the same analysis and possibly reduce the dimensionality of my data while better understanding how certain perceptions are related to each other and also related to rapport.

I will also utilize questionnaires that were designed for assessing human-computer interactions, but can be applied readily to human-human interactions. For example, Pecune et al. (2019) asked users to rate the degree to which they felt that the computer agent was interested in what they were saying. I can easily modify this question to apply to humans, e.g. "Did you feel that your partner was paying attention to your own opinions?". The questions in Pecune et al. (2019) were based on those in Zhao et al. (2018), which asked users to rate whether they felt "in sync" with the computer agent. I can again modify this to a human-human interaction by asking if the participant felt they were in agreement with or working well with their partner.

4.3.3 Participant information

Given the novelty of parts of my proposed experiments, I would like to avoid any unnecessary confounds. All participants will need to be native English speakers, to avoid issues where word retrieval is more difficult due to lack of language familiarity. Participants will also need to have completed high school, in order to allow participants to converse at a similar level of complexity. Finally, all participants will need a high rating on Amazon Mechanical Turk to better ensure that they are engaged in the task.

The data for all 3 studies will be from the same experimental collection, but will be split and analyzed in different ways. Study 3 requires enough data to train and test a neural network. In private communication with Vinnie Monaco, an experienced researcher in keystroke dynamics, he proposed a minimum need for 100,000 keystrokes for a small RNN. The Liebman experiments used in the pilot studies generated about 98,000 keystrokes from 38 pairs, although the 38 pairs represent a subset of the original data collection. Assuming we also need a test set and validation set, about 150 pairs in 5-10 minute conversations should be adequate.

4.4 Overall Features

Table 4.2 provides an overview of all of the features to be used in each study in my thesis, as well as features that have been used in my pilot studies.

Feature	**Study 1 Pilot**	**Study 1 Thesis**	**Study 2 Pilot**	**Study 2 Thesis**	**Study 3 Pilot**	**Study 3 Thesis**
Pause before turn starts	✓	✓	✓	✓	✓	✓
Pause after first word	✓	✓		✓		✓
Revision count		✓	✓	✓	✓	✓
Revision length		✓		✓		✓
Pause before revision		✓		✓		✓
Typing rate (keystrokes/second)		✓		✓		✓
Inter-word pause				✓		✓
Intra-word pause				✓		✓
Keypress duration				✓		✓
Length of typing burst				✓		✓
Average word length				✓		✓
Average utterance length (in words)				✓		✓
Average utterance length (in keystrokes)				✓		✓
Pauses before function words				✓		✓
Pauses before content words				✓		✓
Expanded statistical features of all above						✓
Pause duration between n-grams (combinations of 1+ adjacent keystrokes)						✓
Keypress duration of n-grams						✓
Expanded features divided by parts-of-speech						✓
Expanded features divided by semantic type						✓

Table 4.2
Overview of keystroke features used in all studies

Chapter 5

Study 1: Effects of Dialogue Act function on typing patterns

5.1 Introduction

Since (spoken) prosodic cues can be used to improve the accuracy of dialogue act identification (Bothe et al., 2018b), I would like to investigate how the typing analogues of spoken prosody are also unique to types of dialogue acts. It seems that if I can identify silent prosodic similarities, then dialogue act prediction in typed text can use these to improve accuracy in the same way that spoken prosodic cues can improve predictions of spoken dialogue acts. In addition, timing differences also provide clues about the cognitive state of the speaker, and so I can better understand how different types of dialogue acts require more or less cognitive effort.

5.2 Research questions

- Can I use the pause times surrounding the first word of an utterance to predict the underlying function of that utterance, which could ultimately lead to more appropriate responses to that utterance?

- Do different types of dialogue acts require more or less cognitive effort to produce? In other words, should mistakes or lengthened pauses in only certain dialogue acts be treated as signs of cognitive stress, and not be weighted equally in terms of judging a typist's cognitive state?

5.3 Hypotheses

This study will investigate three key moments in composing a message, as illustrated in Figure 5.2, as well as overall typing speed and accuracy throughout an utterance. At t_0, the speaker retrieves from memory the message they plan to transmit. At this point, no text has been produced. The next point, t_1, represents a point during message production. A message has been partially or fully composed, and the speaker checks (or does not check) the status

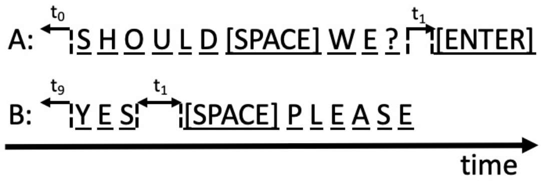


Figure 5.1
A toy dialogue illustrates the pause intervals that are used for this study. All solid lines with a letter above represent the time that key was depressed. t_0 represents the time during which a message is retrieved or planned, before production begins. t_1 represents timing between words, when the state of the conversation can be checked.. t_2 represents the pause before the message was transmitted.

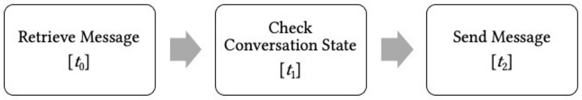


Figure 5.2
An abstraction of the cognitive processes underlying the production of the utterances in Figure 5.1.

of the conversation. At t_2, the speaker is satisfied with the message they have composed, and transmits the message for message. Depending on the dialogic function of a message, I hypothesize that different dialogue acts will apply cognitive strain at different points in time, which will be reflected in differing typing patterns. Understanding these differences will allow researchers to better understand the underlying function of the utterance being produced.

My first hypothesis investigates t_0 in Figure 5.2. Different dialogic functions require different amounts of retrieval before being produced. A backwards-functioning dialogue act requires reference to a previous utterance, in order to ask a clarifying question or make a clarifying remark. On the other hand, a forward-functioning dialogue act does not make reference to prior material in a conversation. Since studies such as Logan and Crump (2011) demonstrated that a greater amount of retrieval causes longer initial pauses before typing, I propose that initial pauses before message composition will be connected to the dialogic function of the message.

H1: Pauses before backward-functioning dialogue acts will be shorter as compared to forward-functioning acts.

Once message composition begins, a speaker may or may not need to check the current state of a conversation (t_1). For example, as Dhillon (2008) points out, if a speaker produces a (forward functioning) floor-grabber, then the speaker will need to take account of the state of the conversation to see if they have wrested control, or grabbed the floor. This state-check results in a pause after message production has begun. Taking this into account, I propose that pauses in the midst of message composition are also affected by the dialogic function of the message.

H2: Pauses after the initiation of a forward-functioning dialogue act will be longer than pauses after a backward-functioning utterance is initiated.

I will then look at how dialogic function affects the timing surrounding the transmission of a message (t_2). Since a backward-functioning dialogue act is connected to previous context, the material it's connected to accumulates as a conversation progresses. Some evidence for this comes from Bothe et al. (2017) and Bothe et al. (2018b), which found that only incorporating local context, i.e. the previous two utterances, deteriorated the accuracy of their dialogue act labeling model. If a backward-functioning dialogue act requires consideration more than local context, e.g. the entire previous conversation, then taking this into account will improve the accuracy of predicting a dialogue act. On the other hand, a forward-functioning dialogue act does not take account of the previous material, as it's function is to advance the conversation to a new state or new topic.

Chukharev-Hudilainen (2014) notes that longer pauses seem to represent more complex linguistic processing. As an example, the overall composition of an utterance is more complex than letter-by-letter orthography. From this, it seems that the composition of a backwards-functioning utterance would become more complex as a conversation goes on, because the composition takes account of more and more context, and so the final composition decision becomes more difficult to make.

H3: As a conversation proceeds, more cognitive lapses will occur in backward-functioning dialogue acts. Conversely, the frequency of cognitive lapses in forward-facing dialogue acts will stay more constant throughout a conversation.

H3a: As a conversation proceeds, pauses within backward-functioning dialogue acts will becomes longer. Conversely, pauses within forward-facing dialogue acts will stay more constant throughout a conversation.

H3b: As a conversation proceeds, revisions within backward-functioning dialogue acts will becomes more frequent, and be preceded by longer pauses. Conversely, within forward-facing dialogue acts, revision count and pauses before revisions will stay more constant throughout a conversation.

5.4 Exploratory study

The results reported below examine the pauses before and after the first word of an utterance, respectively. In Figure 5.1 these pauses refer to t_0 and t_1.

For my first exploratory study, the aim was to test which factors most strongly influenced the length of a pause before a participant begins their next utterance. In line with studies such as Alves et al. (2007), if a pause is indicative of additional cognitive load, I was curious if different dialogue acts require different amounts of cognitive effort to process and produce.

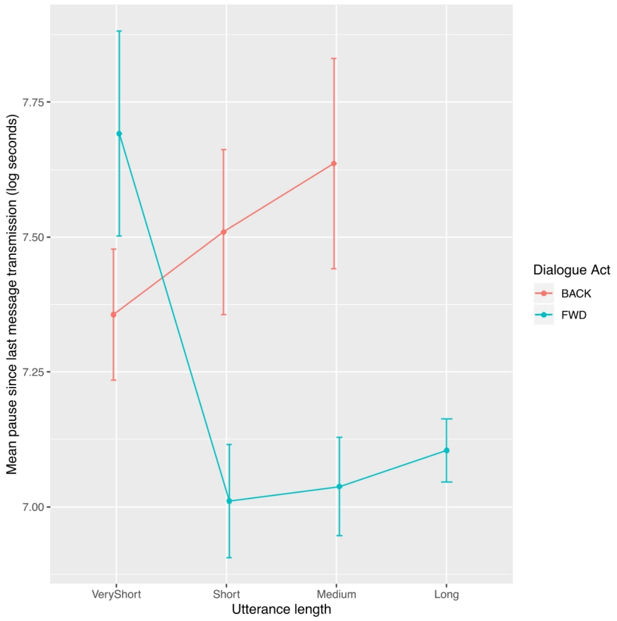


Figure 5.3
The interaction between dialogue act and utterance length. The error bars represent the standard error ($\frac{\sigma}{\sqrt{n}}$)

5.4.1 Pauses before the first word of an utterance

To test the differences in pause lengths, and what accounted for them, I ran a factorial ANOVA to predict the log pause time before an utterance. The predictors of interest were word length, utterance length, dialogue act, and most recent sender. Interactions between all terms were also included. With the exception of utterance length, all main effects were significant at the $p < 0.05$ level. In addition, the interaction between utterance length and dialogue act was also significant.

- Word length - The pause before long words was lengthier than pause before medium and short words.
- Dialogue acts - The pause before backward-functioning DAs was lengthier than the pause before forward-functioning DAs.
- Interaction between utterance length and dialogue act - Pauses before backward-functioning DAs get longer as the utterance length increases, but that pauses before forward-functioning DAs display the opposite effect, at least from very-short to short utterances. In longer forward-functioning utterances, the effects are not significant. This is illustrated in Figure 5.3.

This interaction is interesting, and will require further investigation. It could be the case that the previous utterance primed the speaker for certain words, and so those words were already more highly activated. In addition, I will also look at whether a word or phrase that occurs even further back in the conversation affected this production pattern as well.

5.4.2 Pauses after the first word of an utterance

After observing that the dialogue act does affect the initial pause before an utterance, I wanted to look for a parallel effect *after* the initial word. However, different opening words introduce a high amount of variability. In other words, at the end of different words I would expect the typist to be in a different cognitive state; thus any variability would be difficult to connect independently with the dialogue act function. By implementing the controls below, I can be more certain that production variability is due to dialogue act differences rather than word length differences.

This part of the study is based on Dhillon (2008), which looked at a set of "floor grabber" words that can serve dual purposes in speech, and measured the pause following only very specific words. In text-based communication, though, the concept of a "floor grabber" is not nearly as overt. While a speaker may look at a status in a chat window that illustrates their interlocutor "is typing," it is difficult for a speaker to reliably judge whether or not they have gained control of the conversation. Thus, a narrow set of words would probably provide very parse data, with little information.

In lieu of this consideration, I chose instead to limit the exploratory study to affirmative words, or words that are synonymous with "yes." These words are similar in length, and all have the same set of ambiguous pragmatic functions. Similar to the conversation in (1), an affirmative word can serve different dialogue functions. I selected all utterances that began with one of the following words: *yes, yeh, yup, yep, yeah, yea, yah, ya.* This filter resulted in 81 utterances for consideration.

To test sources of variability I set up a factorial ANOVA similar to that in the first half of this study. The model was set up to predict the pause length after the last letter of the words. The predictors of interest were dialogue act function, whether the next keystroke was a SPACE or ENTER, word length, utterance length, whether or not this utterance overlapped with the previous utterance, and finally which word was used to initiate the utterance. I also looked at interactions between predictors.

Results of the ANOVA found significant predictors in dialogue act function, next keystroke content, and whether or not this utterance overlapped the previous utterance. It also showed a significant interaction between the first word used and whether the utterance overlapped the previous utterance. Non-significant predictors included word length, utterance length, the first word used, and the remaining interactions.

- Dialogue act - The pause following the first word of a forward-functioning DA was shorter than the pause following the first word of a backward-functioning DA.

- Subsequent keystroke - Pauses that preceded SPACE (indicating more words to come) were shorter than pauses preceded ENTER.

5.4.3 What I learned

The exploratory results above point to the influence of dialogue function on the process of language production. While it is well-established that factors such as syntactic planning and lexical selection play a role in determining pause duration (Butterworth, 1980), the results of this exploratory study point to additional factors in language production. Similar to Dhillon

(2008), the exploratory study shows that dialogue act function can also play a significant role in affecting pauses in typing as well as speech.

Both studies showed that the DA played a significant role in determining the pause preceding the first word of an utterance, and following the first word of an utterance, respectively. For forward-functioning DAs, the pause before the first word was shorter when compared to the first word of similar backward-functioning DAs. Perhaps this points to the urgency of interrupting the direction of a conversation, where a speaker will more quickly speak up when they want to gain control and also change direction.

A similar trend is seen for the pause following the first word of an utterance. For forward-functioning DAs, the pause after the first word was also shorter when compared to the first word of similar backward-functioning DAs. This finding diverges from those of Dhillon (2008), which found a longer pause after a floor grabber, perhaps because the speaker needed to take account of whether they had gained control of the conversation.

The preliminary finding above seem to show that at least in respect to pauses following certain words, the pause duration does not directly parallel similarly located pauses found in spoken language production. However, to reiterate what has been previously written, I do not expect every phenomena in spoken prosody to be mirrored in implicit written prosody. But it is interesting to note where parallel relationships, orthogonal relationships, and ambiguous relationships exist.

This points to the need for more extensive investigations of the effects of dialogue act on different typing phenomena. The interaction results, especially, suggest that there are nuances that impact the processes involved in producing differently functioning dialogue acts.

It seems likely that incorporating keystroke timing patterns into predictions of dialogue act labels will be helpful. However, straightforward pause timings do not seem to be the most accurate predictors, in and of themselves.

Nonetheless, this exploratory study still shows consistent differences for production of different dialogue acts. Thus, my conclusion is still similar to that reached by Dhillon (2008, p. 5353): “The current study adds to this body of work by demonstrating that pauses also represent delays incurred in pragmatic selection.”

5.5 Expansion in thesis

The findings of the exploratory study are interesting on their own, but also call for further clarification. For example, it is possible that pause times are also related to the pace at which a conversational partner is responding. If one speaker is replying slowly, does this cause the other to do the same? Or are responsive typing rates unaffected by the speed at which a partner is replying?

Most importantly, as laid out in the overall feature list (Table 4.2), my thesis will also test the connection between DA and more sophisticated keystroke production features. The exploratory study only investigated pauses surrounding the first word of an utterance. However, this overlooks a number of other important locations and features in an utterance. My thesis will add the following:

- All inter-word pauses, to test if prolonged pauses take place at other locations

- Revision counts and pauses preceding revisions, to test how often mistakes and/or corrections are made during different dialogue acts.

- Overall typing rate, to test if different dialogue acts at different locations in a conversation affect the overall typing process.

In addition, this study only reported results from ANOVA tests. Further statistical testing and modeling is clearly necessary. For example, while an ANOVA points to the sources of variance, we do not know to what extent each factor influences the pause times. Running linear regression models will shed light on this. Moreover, these findings need to be clarified through hierarchical models, where different participants or participant dyads are considered separately.

Finally, my thesis will test the ubiquity of these findings, by also testing them in a different conversational setting. Whereas the feasibility study used a less structured and goal-motivated dialogue, my thesis will use conversations from a structured game that necessitates more engagement between partners and evokes more genuine reactions.

5.6 Contributions

Being able to accurately identify different types of dialogue acts is important for keeping the quality of a conversation high. For example, if a speaker is using a lot of backward-facing clarifications, comments, or questions, this could imply that the speaker is confused. On the other hand, if a speaker uses a lot of forward-facing dialogue acts, this could imply that a speaker understands, and is ready to move to the next topic.

The results of the exploratory study provide evidence that while text-based conversations necessarily lack spoken prosody, some of the information provided through spoken prosody is still transmitted in alternative fashions. This would challenge "bandwidth"-oriented theories (Gergle, 2017, p. 193 and *inter alia*) that posit that text-based communication lacks the bandwidth, and thus does not contain, information that is privileged in or exclusive to spoken communication. While this information, using current technology, is not immediately apparent to a message recipient, my studies suggest that the information does exist, and perhaps needs to be made more overt to the receiver of a message.

These findings also have interesting cognitive architecture implications. A recent theory of keystroke dynamics purports that the typing process is made up of two hierarchical loops: an outer loop is affected by lexical retrieval, and this then cascades to an inner loop that is affected by actual letter-by-letter word execution (Logan and Crump, 2011; Yamaguchi et al., 2013). Because my exploratory study found effects of dialogue act function both before and after a word, it seems that dialogue function affects both the inner and outer loop of this cognitive model, whether independently or flowing from one loop into another.

These findings could also contribute to improvements to a chatbot, where an algorithmically-determined response is generated. For chatbots, understanding elements of a conversation such as common ground can be very difficult (Traum, 1994; Thorne, 2017). However, understanding the dialogue function could possibly serve as a proxy for common ground, e.g. if a speaker expresses agreement, it implies that the facts agreed upon are part of the established common ground.

Finally, and as is echoed throughout my thesis, these findings could improve a tool that assists people with better comprehending online conversations. The cause of this difficulty could be cognitive or emotional impairments, or even lack of familiarity with the language. One could imagine a chat communication platform that visualizes the state of the conversation, and assists a user in understanding whether their conversational partner is agreeing, or requires additional information.

5.7 Progress

Similar to my other studies, I am working with a research assistant to build an apparatus for data collection. Upon receiving IRB approval, I will begin testing the apparatus on a pool of subjects, and then begin the data collection process.

Chapter 6

Study 2: Typing patterns are most accurately modeled by the interaction of semantic similarity and sentiment

6.1 Introduction

One of the most important differences between this study and Study 1 is the unit of analysis. Whereas Study 1 considered each utterance individually, this study will look at pairs of adjacent utterances that were produced by different speakers, or dyads. This study looks at these more granular elements of dialogue to ask if semantic similarity to a previous utterance, sentiment similarity to a previous utterance, and the sentiment of the message itself affect typing patterns. This is an important investigation because if sentiment interacts with similarity, it could mean that straightforward sentiment analysis or straightforward similarity measurement are not adequate for uncovering latent feelings or thoughts during a text-based conversation.

Prior studies have shown two consistent facets of similarity or coordination: 1) Nearly all conversations coordinate on multiple levels: the phonetic properties of sounds, turn lengths, and word choice all becoming more similar throughout a conversation. 2) However, not all coordination is a sign of a positive interaction or a positive impression of a partner. For example, Branigan et al. (2011) shows that if a speaker thinks their partner is less intelligent, then the speaker will become more similar in word choice with the partner's word choices. Relatedly, Scissors et al. (2009) shows that matching on certain topics, such as the the past (rather than the present or future) is also not a sign of a positive interaction. (But see Section 3.6 for a more complete rundown of findings.)

As another precedent for this study, Bothe et al. (2017) points out that changes in sentiment often signal a change in situation. Further, their study points out that these changes can be an "early warning" of not-yet-observed cognitive or situational changes. Similarly, Zhou et al. (2018, *inter alia*) points out that sensitivity to emotion in dialogue improves user satisfaction, creates a more positive perception of the dialogue, and leads to fewer breakdowns in dialogue.

Relating specifically to typewritten language production, though, it is instructive to look at a set of previous studies: Lee et al. (2014, 2015); López-Carral et al. (2019). These studies investigated how visually and aurally evoked emotions affected keystroke output. They found consistent effects both on how long a key was held down as well as the latency between keys. They found that as a typist became more excited (very happy and very sad), their keypresses became shorter or quicker. The pauses between keystrokes, however, were affected by the positivity or negativity of the sentiment: as sentiment became more positive, latency between keystrokes was reduced. As a crude summary, when people are unenthusiastic and sad, their typing is slowest, whereas when they are happy and excited, their typing is more rapid.

Being sensitive to changes in the emotional state of a conversation is also important for creating more naturally flowing conversations. Sordoni et al. (2015) points out that sensitivity to "context" is important in this case, both in terms of the lexical content that has been previously produced, as well as the contextual sentiment of what has already been produced.

Finally, predicting sentiment and similarity simultaneously is not without precedent. Wang and Manning (2012) was able to do just this, using variants of a Naive Bayes (NB) and a Support Vector Machine (SVM) classification algorithm. Mao et al. (2014) performed a similar analysis of Wikipedia discussion forums at a less granular time scale, and while they found utility in considering both similarity and sentiment of comments, their methodology largely kept similarity separate from sentiment analysis. Wang et al. (2020) presents an interesting approach as well, where sentiment analysis is improved by using a topic-aware model. They employ this in a synchronous setting, by analyzing the text of customer service call transcripts.

If this study shows that typing patterns can predict phenomena such as a change in sentiment or change in a situation, then this adds support to the notion that tracking typing patterns adds useful information to tracking dialogue state or whether a conversation is proceeding successfully or not.

6.2 Research question

This study will use the same data collected for Study 1, where two partners discuss movie recommendations.

- If an utterance is more semantically similar to the previous utterance, will this affect the typing speed at which the utterance is produced?
- If an utterance uses very different sentiment compared to the previous utterance's sentiment, will this affect how the utterance is typed?
- Are typing patterns sensitive to an interaction between sentiment and similarity, which could provide a more nuanced picture of dialogue state?

6.3 Hypotheses

Although prior studies of both sentiment and similarity are scarce, we can make certain assumptions based on prior studies of either sentiment or similarity. As Bothe et al. (2017)

points out, changes in sentiment can be taken as signs that a situation has changed, which evokes heightened vigilance or emotional changes in a listener. Relatedly, studies such as Lee et al. (2014) and Lee et al. (2015) found that different levels of emotional response can be detected through typing behavior.

While no prior studies exist that examine the effect of textual similarity in text-based CMC, studies such as Villani et al. (2006) and Killourhy and Maxion (2012) compared typing consistency when typing free text (responding to a prompt) versus copying a fixed text. From this comparison, they do find significant more differences that depend on whether a user is copying existing text versus producing original content. Although these studies are not perfect analogs to producing typewritten language that is similar to previously produced text, they do show that the originality of a text influences typing patterns.

From the side of speech prosody, studies such as Pell (2001) and Pell et al. (2011) show that the interaction of emotion and other linguistic properties affects prosodic properties of speech such as word duration.

Combining the findings of these studies, it seems that typing is affected both by sentiment as well as by similarity to existing text. As such, I propose the following:

H1: Typing will most accurately be predicted by accounting for both sentiment change as well as text similarity.

> **H1a:** Keypress duration, a keystroke analog to voice intensity, will most accurately be predicted by accounting for both sentiment change as well as text similarity.
>
> **H1b:** Interword pauses, a keystroke analog to word duration, will most accurately be predicted by accounting for both sentiment change as well as text similarity.

As mentioned by Bothe et al. (2017), a *change* in sentiment could signal a change in a situation, which would affect how fluidly and confidently a speaker would respond. This would also make sense under the *collaborative* model of conversation, where dialogue is a joint activity, in contrast to the *autonomous* model of a speaker acting independently (Clark, 1996; Phillips, 2007). Because of this, I would make two predictions:

H2: The effect of sentiment *change* between utterances will be greater than the effect of the sentiment of the utterance itself.

> **H2a:** Pauses throughout an utterance will be longer when the utterance's sentiment is more different than the preceding utterance's sentiment.
>
> **H2b:** Revision counts in utterances will be higher when the utterance's sentiment is more different than the preceding utterance's sentiment.

6.4 Exploratory study

While the studies mentioned above point to both sentiment and similarity affecting language production, none of them come to that conclusion directly. Therefore I wanted to run an exploratory study to see if promising signs exist that this is a worthwhile inquiry.

For measuring similarity, I used a very basic 'R' package that relies on a pre-trained word term-document frequency count to produce a 300-dimensional vector for over 100,000 words (Günther et al., 2015). I then used a bag-of-words approach, where word order is not taken into consideration, to measure the cosine similarity between an utterance and the previous utterance.

For measuring sentiment, I used 'sentimentr' (Rinker, 2016) which measures the sentiment of an entire sentence rather than taking the mean sentiment of each word. As an example of the utility of this approach, an intensifier plus an adjective, e.g. "very bad", would not be considered a positive/neutral word plus a negative word; rather, the sentiment of "bad" would be multiplied because of the intensifier "very."

After measuring both similarity and sentiment, I filtered out any utterances that did not have a sentiment score and a similarity score, due to words being unrecognizable or outside the vocabulary of the systems (OOV). This left me with only 256 utterances from 49 participants.

6.4.1 Effects of sentiment and similarity on initial pauses

Similar to the exploratory study for Study 1, I measured the initial pause before an utterance is produced (t_0 in Figure 5.1). I used an ANOVA that predicted pause times from similarity scores, sentiment scores, dialogue act functions (forward vs backward), and whether the utterance was a continuation of the speaker's own utterance or a response to their partner's utterance. All interactions were also measured.

An interesting interaction was found between sentiment, similarity, and dialogue function ($p = 0.046$). This could point to the notion that sentiment and similarity matter more when a speaker is only replying to their partner.

I also ran a similar ANOVA in which sentiment change was used in place sentiment *per se*. In contrast to **H2**, I did not find an effect of change, only of the sentiment of the utterance itself.

6.4.2 Effects of sentiment and similarity on revision counts

I also ran an ANOVA comparing how well revision counts are predicted by similarity scores, sentiment scores, dialogue act functions (forward vs backward), and whether the utterance was a continuation of the speaker's own utterance or a response to their partner's utterance. All interactions were also measured.

The only significant predictor of revision could was similarity score. Sentiment change and utterance sentiment were not significant predictors. These findings make sense in light of findings from Killourhy and Maxion (2012), where the originality of a text affects typing patterns. However, the findings would run counter to my own hypotheses.

6.4.3 What I learned

More than anything, I believe this exploratory study pointed to how underpowered the current data is. After filtering, I was left with only 256 utterances from 49 participants. My thesis data collection methods should generate a larger sample size.

It also pointed to the limitations of using only very basic models, in that the models were difficult to refine, since I did not know which directions to refine them in.

At the same time, I did find promising signs of the effect of sentiment and similarity in a text-based CMC environment. This is important because it signals that the findings of previous literature that studied typing in isolation could still be applicable to my studies of typing in dialogue.

6.5 Expansion in thesis

The effects of sentiment and similarity on typing should extend far beyond the straightforward tests in my exploratory study. While the same basic structure (Typing Feature ∼ Similarity * Sentiment) will be employed, the studies in my thesis will bear little resemblance to the studies in my exploratory study.

I will expand the typing features that are predicted, as I would expect other areas of typing production to show sensitivity to sentiment and similarity. A full list of additional features is listed in Table 4.2.

Medimorec and Risko (2017, etc.) highlight the importance of all inter-keystroke intervals, whether pauses surrounding a word or pauses within a word. While pause lengths provide critical information about a speaker's state of mind, the location of these pauses relative to words or the entire utterance are also important and have different underlying influences.

As pointed out in studies such as Lee et al. (2015), keypress duration (from KeyDown to KeyUp) is also intimately tied to emotion. Not only will I measure keypress duration for individual keys, I will also look at the cumulative duration of keypress bigraphs (KeyPress_n + KeyPress_{n+1}), which are employed in many studies especially around personal biometric verification (Monrose and Rubin, 1997b; Raul et al., 2020).

As mentioned in Study 1, a straightforward count of revisions is hardly representative of the overall nature of revision. Studies such as Lindgren et al. (2019) and Zhang et al. (2019) point to the importance of also considering the length of revisions, the pauses before revisions, and the amount of time spent in revisions. These further dimensions of revision behavior should provide better insight into how sentiment and similarity affect the surety of a speaker, reducing the amount of revision necessary.

I will also look at how sentiment and similarity affect typing bursts (Baaijen et al., 2012; Van Waes et al., 2012). To measure the nature of bursts, I will look at sequence lengths between significant pauses, as well as sequence lengths between significant revisions.

All keystroke features mentioned in the exploratory study for Study 2 will be reused, along with additional features as laid out in Table 4.2. As an overview of these additions, they will fall into four main categories

In addition, I will use much more subtle statistical tests. There is little reason to expect all features to have a normal distribution. Further, not all features will have monotonic linear effects, and my statistical tests should be sensitive to this.

6.6 Contributions

My first contribution will be in adding more research to the understudied phenomenon of sentiment and similarity interaction (Lukyamuzi et al., 2020). If my ultimate goal is to make latent motivations more apparent to a text-based interlocutor, then I will first need to know what those latent motivations actually consist of.

Many industries are using sentiment analysis to evaluate everything from dissemination of false news to customer satisfaction (e.g. Medhat et al., 2014; Wang et al., 2020). By using typing patterns to measure cognitive load, and thus its connection to both sentiment and similarity, I hope to improve traditional use of sentiment analysis, using keystroke analysis to demonstrate the insufficiency of sentiment analysis or semantic similarity alone. This may allow for more accurate measurement of user sentiment.

These more granular elements will also be important for the subsequent studies in my thesis, though those studies are not contingent upon the results of this study. Rather, similarity and sentiment are two elements that modulate feelings of rapport, so the findings of this study could play a part in understanding overall rapport. Perhaps even since rapport is difficult to define, similarity and sentiment measures could be an adequate and observable proxy for rapport.

6.7 Progress

The progress for Study 2 is the same as Study 1. A pilot study has been run, and my research assistant and I are in the process of developing an apparatus for data collection.

Chapter 7

Study 3: Typing patterns are affected by feelings of rapport with a conversational partner

7.1 Introduction

My previous studies demonstrated (1) how Dialogue Act function affects typing patterns and (2) how semantic similarity and sentiment towards a partner affect typing.

Given that certain components of dialogue affect typing patterns, the present study expands the scope of inquiry by examining the connection between typing patterns and rapport, where rapport is a notoriously ill-defined and multi-dimensional phenomenon. It is important to note, however, that while the results of Studies 1 and 2 could be instructive in the feature engineering for this study, strong results from Studies 1 and 2 are not a necessity for designing this study's features nor supporting its results.

At the end of this section, I provide two preliminary results of pilot studies that seem to hint that typing data does have some connection to the more abstract terms in definitions of rapport (see Table 3.3). If a connection does exist, then the results of this study in my thesis could help a conversational partner who, even if unable to define rapport, still make an accurate prediction of the feeling of rapport from a conversational partner. In a situation where high rapport is necessary, the speaker could also make necessary adjustments to raise the level of rapport, and gain a sense of whether their changes are effective.

7.2 Research questions

- Can self-reported rapport ratings be accurately predicted based on differences in typing patterns?

7.3 Hypotheses

While rapport is a pervasive phenomena, in that it exists throughout a conversation, it is also a dynamic phenomena and can change for different reasons at different points (Tickle-Degnen and Rosenthal, 1990). In a recent study, Lubold and Pon-Barry (2014) investigated the time-course of coordination and its connection to rapport. They measured rapport using both external annotators as well as self-reported scores and found that these different sources were most in agreement in the second half of a conversation. In addition, they found that the most informative type of coordination was immediate proximity (see Section 3.6 for an overview of different types of coordination).

H1: The most accurate predictions of rapport will come from using only the second half of a conversation, and only considering parallels in immediately adjacent utterance pairs.

Given the complexity of rapport, I hypothesize generally that keystroke dynamics is an ideal modality for detecting and predicting levels of rapport. Keystroke data reflects both latent tendencies that a partner cannot observe, as well as the explicit language produced and received. Both of these sources are likely valuable sources of information about feelings of rapport.

The value of these dual sources can be seen from various studies that predict rapport from different angles: Acosta and Ward (2011) used the sentiments expressed in a dyad of utterances to predict rapport, while Olson and Parkhurst (2013) showed that fine-grained behavior such as reaction times was connected to levels of rapport.

One study that points to this dual nature was able to raise rapport levels by creating a voice-based computer agent that matched both prosody and sentiment matching in its responses (Li et al., 2017). They attribute these gains to the perception that the robot is perceived as matching the mental state of the user.

In order to identify matching mental states, I will rely on the Linguistic Inquiry and Word Count (LIWC) system to classify words by their emotional underpinnings, cognitive approach, and even their linguistic parts-of-speech (Niederhoffer and Pennebaker, 2002; Chung and Pennebaker, 2014). These classifications get at *how* a speaker is expressing something, e.g. a pronoun versus proper name, rather than *what* they are expressing, e.g. a person. The *how* is often more revealing than the *what* (Niederhoffer and Pennebaker, 2002).

This study will go one step further than current LIWC-based studies have gone, though, and measure the amount of cognitive effort that goes into producing word classes. By investigating this, I can look deeper than whether or not interlocutors are agreeing on content by matching word choices, but look at the effort that goes into agreeing or disagreeing. To measure cognitive effort, I will use pause times before words as well as revision properties (Stromqvist, 2007).

H2: Self-reported rapport levels are best predicted when matching linguistic styles are produced with low cognitive effort (shorter pauses and fewer revisions).

Finally, studies of cognition and keystrokes have shown that the length of a typing burst is due partially to cognitive load (Galbraith and Baaijen, 2019). Further, pauses between typing bursts are due to planning, replanning, and revision. By looking at many definitions of rapport, it seems evident that rapport is associated with feeling comfortable and at-ease in a conversation. These feelings then should be borne out by typing patterns.

H3: Conversations with high rapport ratings will exhibit longer typing bursts, with more consistent pauses between bursts.

7.4 Exploratory study

To see if there were any promising signs that keystroke patterns provided reliable predictors of rapport ratings, I ran a cursory analysis of two keystroke patterns, also using data from Liebman and Gergle (2016a,b). These illustrate some encouraging trends, despite that the conversational tasks are less structured than the studies I am proposing for my thesis. In my thesis, I will also run a thorough statistical analyses, rather than relying on visual impressions of charts.

This exploratory study used ratings that subjects provided in a post-test questionnaire, after completing the conversations in the Liebman experiments. These ratings were based on 6 questions that used a 7-point bipolar scale proposed by LaFrance (1979). As an example of one rating to judge rapport, a subject would rate to what degree they felt out-of-step (0) to in-step (7) with their partner. The rapport ratings below took the mean rating of all 6 questions.

In the second half of this study I looked at the amount of revision in an utterance, quantified by how often a subject used the Backspace or Delete key when producing an utterance.

7.4.1 Distribution of rapport

I first looked at the overall distribution of rapport ratings that a subject assigned to their partner. From Figure 7.1, there seems to be two interesting trends: 1) The distribution of ratings for 5-minute conversations is less normal than the corresponding distribution for 15-minute conversations. This seems to suggest that as speakers talk for longer periods of time, their impressions of each other become more nuanced. Whereas a 5-minute conversation leaves a more noisy impression, a longer conversation leaves a more expected distribution of some high ratings, some low ratings, and mostly medium-level ratings. Still, my takeaway is that these types of conversations can lead to accurate impressions of rapport.

7.4.2 Revisions and rapport

I then measured the number of deletions by rapport rating, where the count of deletions is a proxy for the extent of revisions in a conversation. A conversation that is high in rapport should result in fewer deletions: If interlocutors are comfortable with each other, then they will not need to be as careful about their wording, and will feel less stressed and cognitively burdened. Given the results of previous studies of typographical errors and cognition, e.g. Lindgren et al. (2019), which shows that revisions can be due to the need to change or

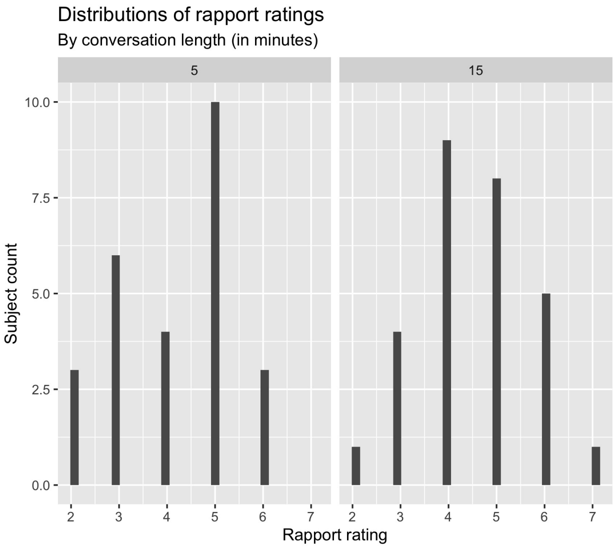


Figure 7.1

elaborate ideas, I would expect fewer mistakes (and hence fewer deletions) when rapport is higher. On the other hand, when rapport is low and a speaker does not feel at ease with their partner, I would expect that speaker to make more mistakes and also be less certain of the words and ideas they are producing.

From the chart above, a number of interesting trends seem evident. The first comes from the top and bottom rows. Each row divided the individual turns by whether a speaker was continuing an utterance they previously produced ("Same Speaker" or a continuation of a multi-utterance turn) or responding to their partner ("New Speaker"). Looking at the bottom row first, we see that rapport ratings seem to affect the number of deletions. Though the trend is not perfect, it seems that as rapport becomes higher, the typists make fewer revisions. This is what we would expect: as subjects become more comfortable with their partners, they do not need to make as many revisions to what they have produced.

However, the trend of fewer revisions as rapport grows only applies to the initial utterance when responding to a partner (a New Speaker). As seems evident from the top row, once a speaker has produced the initial utterance of their response, they seem to level off and make a similar number of revisions regardless of rapport level.

To confirm this, an ANOVA of only the New Speaker shows that rapport level does have a marginally significant effect on the number of revisions ($p = 0.046$).

This could help inform my feature engineering for the corresponding study in my thesis, in that the initial utterance of a turn should be more highly weighted than subsequent utterances. It will be discussed in the following section.

The second noticeable difference is between columns of data, which are divided by conversation length. Although this pilot study comes from a smaller data set, when focusing on the bottom row (responses rather than continuations), it does seem that longer conversations provide a more reliable or monotonic decrease in revision count as rapport becomes higher. On the other hand, the data for shorter conversations seems noisier and less consistent, and does not reliably change in one direction or another.

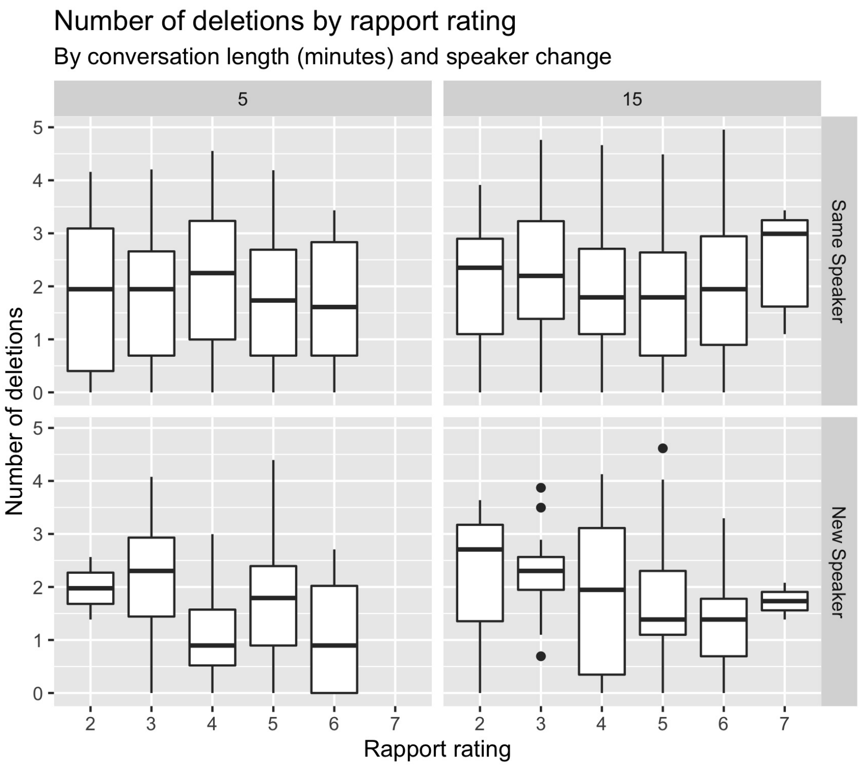


Figure 7.2
Number of revisions given rapport ratings, broken down by two dimensions: rows represent whether a speaker was responding to a partner or continuing their own utterance; columns represent the length of the conversation.

7.4.3 What I learned

From this pilot study, I learned two important pieces of information. First, even in the more unstructured conversations used in the Liebman experiments, where participants might also have false motives, meaningful differences in rapport ratings still exist. I hope that by adding more structure as well as more data points (more subjects) these trends will become more clear.

It was also interesting to see how rapport primarily affected the amount of revision in an initial utterance, rather than in an utterance from the same speaker that continues the same turn. Although this will be further investigated in my thesis, it could point to the idea that rapport only affects initial retrieval or planning. However once an utterance has been retrieved from memory and has started to be produced, rapport no longer affects the subsequent execution of the words.

One important factor missing from this study is consideration of utterance length. As an utterance becomes longer, I should naturally expect more opportunities for mistakes and revisions. In my thesis, I will take this into account, whether by normalizing revision count by utterance length or only comparing similar length utterances.

Further, I did not consider the length of each individual revision, but rather grouped all revisions together. As pointed out in studies such as Lindgren et al. (2019), the reasons for correcting one mistyped character are very different from the reasons for deleting and retyping an entire sentence.

Given this, Figure 7.2 could be composed of a collection of single character deletes or it could be a collection of entire sentence revisions. My thesis will control for this, and only compare similar revision types.

In summary, it does seem that an exploratory analysis of pilot data shows that this is a promising direction. My full analysis (Study 3), though, will use more rigorous testing with a larger sample size. Nonetheless, it does seem that typing patterns have a somewhat reliable connection to feelings of rapport.

7.5 Expansion in thesis

questionairre - since questions come from robot intx, means that findings could also be useful for robot HCI, not just F2F

By far the most significant change in my thesis is that this study will use typing patterns as predictors, rather than dependent outcomes. In this way, the utility of Study 3 will be that it would be possible to use non-obtrusively collected keystrokes to make predictions of rapport, a phenomenon that has repeatedly proven to be difficult to define and measure

To create the typing pattern predictors, I will rely on many of the same features as in the previous studies, such as pauses, revisions, and typing rate. However, all of these features will be enhanced by breaking down each feature by its LIWC category, and looking at differences between partners.

As an example, one predictor will be `differences-revision-count-pronoun`. This feature would be built through the following routine:

1. Find all instances of pronouns being typed by User1 and User2, regardless of whether the pronoun is in the final text.

2. Of all pronouns, count how many were revised, deleted, etc. for each user.

3. Create the feature value by comparing the difference between User1's proportion and User2's proportion.

If speakers are putting the same effort into pronoun choice, then this may be a strong predictor of rapport between the two speakers.

Study 3 will use a machine learning model that can handle a high number of dimensions. Since a single datapoint will generate a high number of features, a neural network is likely well-suited for this task. I will split my data into 80% training/10% validation/10% testing in order to test network architectures. I think a large part of this determination will come from correlation between features, which I can use a validation set to measure.

7.6 Contributions

As a practical application of these findings, good rapport is essential to successful interactions. One area where this is especially important is in healthcare, where trust and rapport between a care provider and a patient is critical Harrigan et al. (1985). By using the results of this study, a provider could be alerted when rapport is low, and thus the provider should adjust their own approach to interaction.

Similarly, good rapport implies engagement. When selling a product or giving detailed instructions over a computer, it is important that both sides stay engaged, since the conversational partners cannot see one another.

Further, a number of studies have found connections between high teacher rapport and improved student outcomes Estepp (2015); Zhao et al. (2016); Aryadoust (2017). Results of this study could improve measurement of rapport, and therefore provide feedback for teachers to achieve greater student success.

Because my post-test questionnaire will be based on questions from robot-human experiments such as Pecune et al. (2019), my results could also be informative beyond human-human interaction. Although rapport with a human partner is not identical to rapport with a computer-agent partner, if my results show promise for using typing patterns to predict rapport, then they could be adapted by designers of chatbots to tune chatbot responses based on typing patterns.

From a theoretical perspective, the insights from this study will be beneficial to fields such as social psychology in its study of rapport. Although the goal of this study is not to more accurately define rapport, the study will provide more fine-grained accounts of behavior connected to different levels of rapport. This may enable researchers to conduct more detailed experiments to test rapport, which in turn may help better understand and define the phenomenon of rapport.

7.7 Progress

The progress for Study 3 is the same as Studies 1 and 2. A pilot study has been run, and my research assistant and I are in the process of developing an apparatus for data collection.

Chapter 8

Contributions and Publications

8.1 Overall Contributions

The data collected for my thesis, as well as a sanitized version of the Liebman, will add one or two new datasets of keystrokes in dialogues. The only existing dataset is in Italian, and is therefore not accessible to a large portion of researchers.

I would also like to open-source the API being created for keystroke collection of dialogues. An API like this doesn't exist yet. However the creator of jsPsych, a large library for running psychology experiments online, communicated to me that many researchers recently have asked for this.

8.1.1 Possible Publications

Dissertation study	**Contributions**	**Possible venues**
(1) Keystrokes and Dialogue Acts	Keystroke production is affected by the dialogic function of an utterance	ACL, CSCW, CHI
(2) Sentiment and similarity influence typing	Keystroke production is affected by both the sentiment of an utterance and its similarity to previous utterances	CHI
(3) Rapport's influence on typing	Partner rapport can be predicted by keystroke information	CHI, CSCW

Table 8.1

Chapter 9

Timeline

Please see https://docs.google.com/spreadsheets/d/1AFCI-8NvF2BweStUBzcqiroPw1hJyICtjFNhMj0v-oo/edit?usp=sharing

Chapter 10

Concerns and Next Steps

Next steps

- I began working with my RA around July 7. We will hopefully get a prototype set up in 2-3 weeks, and begin testing it.
- While my RA is coding up the API, I will work on my IRB for my proposed studies.

Concerns

- I am concerned about whether my experiment will generate meaningful results.
- I need to brush up on the sophisticated linear modeling and machine learning techniques I would like to implement.
- I need to become more familiar with Javascript, so that I can maintain the API once my RA has finished his coding job.

Chapter 11

Bibliography

Abadi, E. and Hazan, I. (2020). Improved Feature Engineering for Free-Text Keystroke Dynamics. In Markantonakis, K. and Petrocchi, M., editors, *Security and Trust Management*, Lecture Notes in Computer Science, pages 93–105, Cham. Springer International Publishing.

Abramson, A. S. and Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63:75–86.

Acosta, J. C. and Ward, N. G. (2011). Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication*, 53(9):1137–1148.

Aldridge, M. and Fontaine, L. (2019). Using Keystroke Logging to Capture the Impact of Cognitive Complexity and Typing Fluency on Written Language Production. In *Observing Writing*, pages 285–305. Brill.

Allen, L. K., Jacovina, M. E., Dascalu, M., Roscoe, R. D., Kent, K. M., Likens, A. D., and McNamara, D. S. (2016). {ENTER} ing the time series {SPACE}: Uncovering the writing process through keystroke analyses. *International Educational Data Mining Society.*

Alves, R. A., Castro, S. L., de Sousa, L., and Strömqvist, S. (2007). Influence of Typing Skill on Pause–Execution Cycles in Written Composition. *Writing and Cognition*, pages 55–65.

Anderson, R. P. and Anderson, G. V. (1962). Development of an Instrument for Measuring Rapport. *The Personnel and Guidance Journal*, 41(1):18–24.

Aryadoust, V. (2017). Understanding the Role of Likeability in the Peer Assessments of University Students' Oral Presentation Skills: A Latent Variable Approach. *Language Assessment Quarterly*, 14:398–419.

Ashby, J. and Clifton Jr., C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, 96(3):B89–B100.

Baaijen, V. M., Galbraith, D., and de Glopper, K. (2012). Keystroke Analysis: Reflections on Procedures and Measures. *Written Communication*, 29(3):246–277.

Ballier, N., Pacquetet, E., and Arnold, T. (2019). Investigating Keylogs as Time-Stamped Graphemics. In *Graphemics in the 21st Century*, pages 353–365, Brest. Fluxus Editions.

Banerjee, R., Feng, S., Kang, J. S., and Choi, Y. (2014). Keystroke Patterns as Prosody in Digital Writings: A Case Study with Deceptive Reviews and Essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1469–1473, Doha, Qatar. Association for Computational Linguistics.

Banerjee, S. P. and Woodard, D. (2012). Biometric Authentication and Identification Using Keystroke Dynamics: A Survey. *Journal of Pattern Recognition Research*, 7(1):116–139.

Barnett, M., Sawyer, J., and Moore, J. (2020). An experimental investigation of the impact of rapport on Stroop test performance. *Applied Neuropsychology: Adult*, pages 1–5.

Barnett, M. D., Parsons, T. D., Reynolds, B. L., and Bedford, L. A. (2018). Impact of rapport on neuropsychological test performance. *Applied Neuropsychology: Adult*, 25(3):258–265.

Bazillon, T., Esteve, Y., and Luzzati, D. (2008). Manual vs assisted transcription of prepared and spontaneous speech. In *LREC*, page 5.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.

Bernieri, F. J., Gillis, J. S., Davis, J. M., and Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology*, 71(1):110.

Bhagat, S., Carvey, H., and Shriberg, E. (2003). Automatically generated prosodic cues to lexically ambiguous dialog acts in multiparty meetings. In *Proceedings International Congress of Phonetic Sciences*, pages 2961–2964.

Blacfkmer, E. R. and Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39(3):173–194.

Borj, P. R. and Bours, P. (2019). Detecting Liars in Chats using Keystroke Dynamics. In *Proceedings of the 2019 3rd International Conference on Biometric Engineering and Applications - ICBEA 2019*, pages 1–6, Stockholm, Sweden. ACM Press.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, 44(2):123–147.

Bothe, C., Magg, S., Weber, C., and Wermter, S. (2017). Dialogue-Based Neural Learning to Estimate the Sentiment of a Next Upcoming Utterance. In Lintas, A., Rovetta, S., Verschure, P. F., and Villa, A. E., editors, *Artificial Neural Networks and Machine Learning – ICANN 2017*, volume 10614, pages 477–485, Cham. Springer International Publishing.

Bothe, C., Magg, S., Weber, C., and Wermter, S. (2018a). Discourse-Wizard: Discovering Deep Discourse Structure in your Conversation with RNNs. *arXiv:1806.11420 [cs]*.

Bothe, C., Weber, C., Magg, S., and Wermter, S. (2018b). A Context-based Approach for Dialogue Act Recognition using Simple Recurrent Neural Networks. *arXiv:1805.06280 [cs]*.

Brandt, D. (2014). *The Rise of Writing: Redefining Mass Literacy*. Cambridge University Press.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., and Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, 121(1):41–57.

Breen, M. (2014). Empirical Investigations of the Role of Implicit Prosody in Sentence Processing. *Language and Linguistics Compass*, 8(2):37–50.

Brennan, S. E., Galati, A., and Kuhlen, A. K. (2010). Two Minds, One Dialog. In *Psychology of Learning and Motivation*, volume 53, pages 301–344. Elsevier.

Bridges, D., Pitiot, A., MacAskill, M. R., and Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8:e9414.

Brizan, D. G., Goodkind, A., Koch, P., Balagani, K., Phoha, V. V., and Rosenberg, A. (2015). Utilizing linguistically enhanced keystroke dynamics to predict typist cognition and demographics. *International Journal of Human-Computer Studies*, 82:57–68.

Bukeer, A., Roffo, G., and Vinciarelli, A. (2019). Type Like a Man! Inferring Gender From Keystroke Dynamics in Live-Chats. *AFFECTIVE COMPUTING AND SENTIMENT ANALYSIS*, page 9.

Bunt, H. (2005). A framework for dialogue act specification. *Proceedings of SIGSEM WG on Representation of Multimodal Semantic Information*.

Bunt, H. and Romary, L. (2009). Towards Multimodal Content Representation. *arXiv:0909.4280 [cs]*.

Butterworth, B. (1980). Evidence from pauses in speech. *Language production*, 1:155–176.

Chukharev-Hudilainen, E. (2014). Pauses in spontaneous written communication: A keystroke logging study. *Journal of Writing Research*, 6(1):61–84.

Chung, C. K. and Pennebaker, J. W. (2014). Using Computerized Text Analysis to Track Social Processes. In Holtgraves, T. M., editor, *The Oxford Handbook of Language and Social Psychology*. Oxford University Press.

Clark, H. and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1):73–111.

Clark, H. H. (1996). *Using Language*. Cambridge university press.

Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. In Resnick, L. B., Levine, J. M., and Teasley, S. D., editors, *Perspectives on Socially Shared Cognition.*, pages 127–149. American Psychological Association, Washington.

Clark, H. H. and Murphy, G. L. (1982). Audience design in meaning and reference. In *Advances in Psychology*, volume 9, pages 287–299. Elsevier.

Clark, H. H. and Schaefer, E. F. (1989). Contributing to Discourse. *Cognitive Science*, 13(2):259–294.

Conijn, R. (2020). *The Keys to Writing: A Writing Analytics Approach to Studying Writing Processes Using Keystroke Logging.* PhD thesis, Tilburg University.

Conijn, R., Roeser, J., and Van Zaanen, M. (2019). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing*, 32(9):2353–2374.

Conroy, N. K., Rubin, V. L., and Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4.

Coover, J. E. (1923). A method of teaching typewriting based on a psychological analysis of expert typing. In *National Education Association Addresses and Proceedings*, volume 61, pages 561–567.

Core, M. G. and Allen, J. F. (1997). Coding dialogs with the DAMSL annotation scheme. In *In Proc. Working Notes AAAI Fall Symp. Commun. Action in Humans.*

Dahlmann, I. and Adolphs, S. (2007). Pauses as an indicator of psycholinguistically valid multi-word expressions (MWEs)? In *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions - MWE '07*, pages 49–56, Prague, Czech Republic. Association for Computational Linguistics.

Dammalapati, S., Rajkumar, R., and Agarwal, S. (2021). Effects of duration, locality, and surprisal in speech disfluency prediction in english spontaneous speech. *Proceedings of the Society for Computation in Linguistics*, 4(1):91–101.

Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18(1):7–24.

Dhakal, V., Feit, A. M., Kristensson, P. O., and Oulasvirta, A. (2018). Observations on Typing from 136 Million Keystrokes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–12, Montreal QC, Canada. ACM Press.

Dhillon, R. (2008). Using pause durations to discriminate between lexically ambiguous words and dialog acts in spontaneous speeech. *The Journal of the Acoustical Society of America*, 123(5):3425–3425.

Epp, C., Lippold, M., and Mandryk, R. L. (2011). Identifying emotional states using keystroke dynamics. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, page 715, Vancouver, BC, Canada. ACM Press.

Estepp, C. M. (2015). Teacher Immediacy and Professor/Student Rapport as Predictors of Motivation and Engagement. *NACTA Journal*, page 9.

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2):133–145.

Ferreira, F. and Bailey, K. G. (2004). Disfluencies and human language comprehension. *Trends in Cognitive Sciences*, 8(5):231–237.

Fodor, J. D. (2002a). Prosodic Disambiguation In Silent Reading. *North East Linguistics Society (NELS)*, 32:21.

Fodor, J. D. (2002b). Psycholinguistics Cannot Escape Prosody. In *International Conference in Speech Prosody*, page 7.

Forsyth, E. N. (2007). *Improving Automated Lexical and Discourse Analysis of Online Chat Dialog*. PhD thesis, Naval Postgraduate School, Monterey, CA.

Fox Tree, J. E. (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.

Galbraith, D. and Baaijen, V. M. (2019). Aligning keystrokes with cognitive processes in writing. In *Observing Writing*, pages 306–325. Brill.

Garrod, S. (1999). The challenge of dialogue for theories of language processing. *Language processing*, pages 389–415.

Garrod, S. and Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, 1(2):292–304.

Gergle, D. (2017). Discourse Processing in Technology-Mediated Environments. In *The Routledge Handbook of Discourse Processes*. Routledge.

Gill, A. J., Gergle, D., French, R. M., and Oberlander, J. (2008). Emotion rating from short blog texts. In *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, page 1121, Florence, Italy. ACM Press.

Godfrey, J., Holliman, E., and McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 517–520.

Goodkind, A., Brizan, D. G., and Rosenberg, A. (2017). Utilizing overt and latent linguistic structure to improve keystroke-based authentication. *Image and Vision Computing*, 58:230–238.

Goodkind, A. and Rosenberg, A. (2015). Muddying The Multiword Expression Waters: How Cognitive Demand Affects Multiword Expression Production. In *Proceedings of the 11th Workshop on Multiword Expressions*, pages 87–95, Denver, Colorado. Association for Computational Linguistics.

Grahe, J. E. and Bernieri, F. J. (2002). Self-awareness of judgment policies of rapport. *Personality and Social Psychology Bulletin*, 28(10):1407–1418.

Grau, S., Sanchis, E., Castro, M. J., and Vilar, D. (2005). Dialogue act classification using a Bayesian approach. *9th Conference on Speech and Computers*, page 5.

Gravano, A. and Hirschberg, J. (2009). Turn-Yielding Cues in Task-Oriented Dialogue. In *Proceedings of the SIGDIAL 2009 Conference*, pages 253–261, London, UK. Association for Computational Linguistics.

Gravano, A. and Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech & Language*, 25(3):601–634.

Gravano, A., Levitan, R., Willson, L., Beňuš, Š., Hirschberg, J., and Nenkova, A. (2011). Acoustic and prosodic correlates of social behavior. In *Twelfth Annual Conference of the International Speech Communication Association*.

Gregory, M., Johnson, M., and Charniak, E. (2004). Sentence-internal prosody does not help parsing the way punctuation does. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, pages 81–88.

Günther, F., Dudschig, C., and Kaup, B. (2015). Lsafun-an r package for computations based on latent semantic analysis. *Behavior research methods*, 47(4):930–944.

Hagad, J. L., Legaspi, R., Numao, M., and Suarez, M. (2011). Predicting Levels of Rapport in Dyadic Interactions through Automatic Detection of Posture and Posture Congruence. In *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, pages 613–616, Boston, MA, USA. IEEE.

Harrigan, J. A., Oxman, T. E., and Rosenthal, R. (1985). Rapport expressed through nonverbal behavior. *Journal of Nonverbal Behavior*, 9(2):95–110.

Healey, P. and Mills, G. (2008). A dialogue experimentation toolkit. *Language, Conition and Culture*.

Healey, P. G. T., Mills, G. J., Eshghi, A., and Howes, C. (2018). Running Repairs: Coordinating Meaning in Dialogue. *Topics in Cognitive Science*, 10(2):367–388.

Heath, M. (2021). No need to yell: A prosodic analysis of writing in all caps. *University of Pennsylvania Working Papers in Linguistics*, 27(1):10.

Heldner, M. and Edlund, J. (2010a). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568.

Heldner, M. and Edlund, J. (2010b). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568.

Horton, W. S. (2017). Theories and Approaches to the Study of Conversation and Interactive Discourse. *The Routledge Handbook of Discourse Processes.*

Horton, W. S. and Gerrig, R. J. (2016). Revisiting the Memory-Based Processing Approach to Common Ground. *Topics in Cognitive Science*, 8(4):780–795.

Housen, A. and Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4):461–473.

Ivanovic, E. (2005). Dialogue act tagging for instant messaging chat sessions. In *Proceedings of the ACL Student Research Workshop on - ACL '05*, page 79, Ann Arbor, Michigan. Association for Computational Linguistics.

Jurafsky, D. (1997). Switchboard Discourse Language Modeling Project (Final Report).

Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., Shriberg, E., Stolcke, A., Taylor, P., and Van Ess-Dykema, C. (1997). Automatic detection of discourse structure for speech recognition and understanding. In *1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings*, pages 88–95, Santa Barbara, CA, USA. IEEE.

Kalman, Y. M. and Gergle, D. (2009). Repeats as Cues in CMC Letter and Punctuation Mark Repeats as Cues in Computer-Mediated Communication. In *5th Annual Meeting of the National Communication Association*, Chicago, IL.

Kalman, Y. M., Scissors, L. E., Gill, A. J., and Gergle, D. (2013). Online chronemics convey social information. *Computers in Human Behavior*, 29(3):1260–1269.

Kasher, A. (1972). Sentences and Utterances Reconsidered. *Foundations of Language*, 8(3):313–345.

Keizer, S., op den Akker, R., and Nijholt, A. (2002). Dialogue Act Recognition with Bayesian Networks for Dutch Dialogues. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages 88–94, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Khanpour, H., Guntakandla, N., and Nielsen, R. (2016). Dialogue Act Classification in Domain-Independent Conversations Using a Deep Recurrent Neural Network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2012–2021, Osaka, Japan. The COLING 2016 Organizing Committee.

Killourhy, K. S. and Maxion, R. A. (2012). Free vs. transcribed text for keystroke-dynamics evaluations. In *Proceedings of the 2012 Workshop on Learning from Authoritative Security Experiment Results - LASER '12*, pages 1–8, Arlington, Virginia. ACM Press.

Kołakowska, A. (2013). A review of emotion recognition methods based on keystroke dynamics and mouse movements. In *2013 6th International Conference on Human System Interactions (HSI)*, pages 548–555.

Kołakowska, A. (2015). Recognizing emotions on the basis of keystroke dynamics. In *2015 8th International Conference on Human System Interaction (HSI)*, pages 291–297.

Kołakowska, A. (2018). Usefulness of Keystroke Dynamics Features in User Authentication and Emotion Recognition. In Hippe, Z. S., Kulikowski, J. L., and Mroczek, T., editors, *Human-Computer Systems Interaction: Backgrounds and Applications 4*, Advances in Intelligent Systems and Computing, pages 42–52. Springer International Publishing, Cham.

Krauss, R. M. and Fussell, S. R. (1996). Social psychological models of interpersonal communication. In *Social Psychology: Handbook of Basic Principles*. Guilford Publications.

LaBahn, D. W. (1996). Advertiser Perceptions of Fair Compensation, Confidentiality and Rapport: The Influence of Advertising Agency Cooperativeness and Diligence. *Journal of Advertising Research*, 36.

LaFrance, M. (1979). Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Social Psychology Quarterly*, pages 66–70.

Lange, K., Kühn, S., and Filevich, E. (2015). ”Just Another Tool for Online Studies” (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLOS ONE*, 10(6):e0130834.

Laskowski, K. and Shriberg, E. (2010). Comparing the contributions of context and prosody in text-independent dialog act recognition. In *In Proc. ICASSP*.

Lee, P.-M., Tsui, W.-H., and Hsiao, T.-C. (2014). The influence of emotion on keyboard typing: An experimental study using visual stimuli. *BioMedical Engineering OnLine*, 13(1):81.

Lee, P.-M., Tsui, W.-H., and Hsiao, T.-C. (2015). The Influence of Emotion on Keyboard Typing: An Experimental Study Using Auditory Stimuli. *PLOS ONE*, 10(6):e0129056.

Leinonen, J., Ihantola, P., and Hellas, A. (2017). Preventing Keystroke Based Identification in Open Data Sets. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, pages 101–109, Cambridge Massachusetts USA. ACM.

Levinson, S. C. and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, 6:731.

Levitan, R., Gravano, A., Willson, L., Beňuš, Š., Hirschberg, J., and Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 11–19.

Levitan, S. I., An, G., Ma, M., Levitan, R., Rosenberg, A., and Hirschberg, J. (2016). Combining Acoustic-Prosodic, Lexical, and Phonotactic Features for Automatic Deception Detection. In *INTERSPEECH*.

Li, G., Borj, P. R., Bergeron, L., and Bours, P. (2019). Exploring Keystroke Dynamics and Stylometry Features for Gender Prediction on Chat Data. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1049–1054, Opatija, Croatia. IEEE.

Li, Y., Ishi, C. T., Ward, N., Inoue, K., Nakamura, S., Takanashi, K., and Kawahara, T. (2017). Emotion recognition by combining prosody and sentiment analysis for expressing reactive emotion by humanoid robot. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1356–1359, Kuala Lumpur. IEEE.

Liebman, N. and Gergle, D. (2016a). Capturing Turn-by-Turn Lexical Similarity in Text-Based Communication. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, pages 552–558, San Francisco, California, USA. ACM Press.

Liebman, N. and Gergle, D. (2016b). It's (Not) Simply a Matter of Time: The Relationship Between CMC Cues and Interpersonal Affinity. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, pages 569–580, San Francisco, California, USA. ACM Press.

Lindgren, E., Westum, A., Outakoski, H., and Sullivan, K. P. (2019). Revising at the leading edge: Shaping ideas or clearing up noise. In *Observing Writing*, pages 346–365. Brill.

Locklear, H., Govindarajan, S., Sitová, Z., Goodkind, A., Brizan, D. G., Rosenberg, A., Phoha, V. V., Gasti, P., and Balagani, K. S. (2014). Continuous authentication with cognition-centric text production and revision features. In *IEEE International Joint Conference on Biometrics*, pages 1–8.

Logan, G. D. and Crump, M. J. (2011). Hierarchical Control of Cognitive Processes. In *Psychology of Learning and Motivation*, volume 54, pages 1–27. Elsevier.

López-Carral, H., Santos-Pata, D., Zucca, R., and Verschure, P. F. (2019). How you type is what you type: Keystroke dynamics correlate with affective content. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–5.

Lovric, N. (2003). *Implicit Prosody in Silent Reading: Relative Clause Attachment in Croatian.* PhD thesis, City University of New York (CUNY Graduate Center).

Lubold, N. and Pon-Barry, H. (2014). Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues. In *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pages 5–12, Istanbul Turkey. ACM.

Lukyamuzi, A., Ngubiri, J., and Okori, W. (2020). Polarity and Similarity Measures Towards Classifying an Article on Food Insecurity. *International Journal of Technology and Management*, 5(2):1–10.

Madaio, M., Lasko, R., Ogan, A., and Cassell, J. (2017). Using Temporal Association Rule Mining to Predict Dyadic Rapport in Peer Tutoring. *International Educational Data Mining Society.*

Mao, W., Xiao, L., and Mercer, R. (2014). The Use of Text Similarity and Sentiment Analysis to Examine Rationales in the Large-Scale Online Deliberations. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 147–153, Baltimore, Maryland. Association for Computational Linguistics.

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4):1093–1113.

Medimorec, S. and Risko, E. F. (2017). Pauses in written composition: On the importance of where writers pause. *Reading and Writing*, 30(6):1267–1285.

Michalsky, J. and Schoormann, H. (2017). Pitch Convergence as an Effect of Perceived Attractiveness and Likability. In *Interspeech*, pages 2253–2256.

Microsoft (2021). The next great disruption is hybrid work: Are we ready? https://www.microsoft.com/en-us/worklab/work-trend-index/hybrid-work.

Mijic, I., Sarlija, M., and Petrinovic, D. (2017). Classification of cognitive load using voice features: A preliminary investigation. In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 000345–000350, Debrecen. IEEE.

Monaco, J. V. (2021). personal communication.

Monaco, J. V. and Tappert, C. C. (2017). Obfuscating Keystroke Time Intervals to Avoid Identification and Impersonation. *arXiv:1609.07612 [cs].*

Monaro, M., Galante, C., Spolaor, R., Li, Q. Q., Gamberini, L., Conti, M., and Sartori, G. (2018). Covert lie detection using keyboard dynamics. *Scientific Reports*, 8(1):1976.

Monrose, F. and Rubin, A. (1997a). Authentication via keystroke dynamics. In *Proceedings of the 4th ACM Conference on Computer and Communications Security - CCS '97*, pages 48–56, Zurich, Switzerland. ACM Press.

Monrose, F. and Rubin, A. (1997b). Authentication via keystroke dynamics. In *Proceedings of the 4th ACM Conference on Computer and Communications Security - CCS '97*, pages 48–56, Zurich, Switzerland. ACM Press.

Morales, A., Acien, A., Fierrez, J., Monaco, J. V., Tolosana, R., Vera-Rodriguez, R., and Ortega-Garcia, J. (2020). Keystroke Biometrics in Response to Fake News Propagation in a Global Pandemic. *arXiv:2005.07688 [cs].*

Muir, K., Joinson, A., Cotterill, R., and Dewdney, N. (2017). Linguistic Style Accommodation Shapes Impression Formation and Rapport in Computer-Mediated Communication. *Journal of Language and Social Psychology*, 36(5):525–548.

Mushin, I., Stirling, L., Fletcher, J., and Wales, R. (2003). Discourse Structure, Grounding, and Prosody in Task-Oriented Dialogue. *Discourse Processes*, 35(1):1–31.

Nickels, W. G., Everett, R. F., and Klein, R. (1983). Rapport building for salespeople: A neuro-linguistic approach. *Journal of Personal Selling & Sales Management*, 3(2):1–7.

Niederhoffer, K. G. and Pennebaker, J. W. (2002). Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology*, 21(4):337–360.

Nomura, T. and Kanda, T. (2014). Differences of expectation of rapport with robots dependent on situations. In *Proceedings of the Second International Conference on Human-Agent Interaction*, pages 383–389, Tsukuba Japan. ACM.

Nottbusch, G., Weingarten, R., and Sahel, S. (2007). From written word to written sentence production. *Studies in Writing*, pages 30–53.

Novotney, S. and Callison-Burch, C. (2010). Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-Expert Transcription. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 207–215, Los Angeles, California. Association for Computational Linguistics.

Olson, K. and Parkhurst, B. (2013). Collecting Paradata for Measurement Error Evaluations. In Kreuter, F., editor, *Improving Surveys with Paradata*, pages 43–72. John Wiley & Sons, Inc., Hoboken, New Jersey.

Pecune, F., Murali, S., Tsai, V., Matsuyama, Y., and Cassell, J. (2019). A Model of Social Explanations for a Conversational Movie Recommendation System. In *Proceedings of the 7th International Conference on Human-Agent Interaction*, pages 135–143, Kyoto Japan. ACM.

Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *The Journal of the Acoustical Society of America*, 109(4):1668–1680.

Pell, M. D., Jaywant, A., Monetta, L., and Kotz, S. A. (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition & Emotion*, 25(5):834–853.

Pennebaker, J. (2011). *The Secret Life of Pronouns: What Our Words Say about Us*. Bloomsbury Press.

Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54(1):547–577.

Pew Research Center (2019). Americans favor mobile devices over desktops and laptops for getting news. https://www.pewresearch.org/fact-tank/2019/11/19/americans-favor-mobile-devices-over-desktops-and-laptops-for-getting-news/.

Pew Research Center (2020). How Coronavirus Has Changed the Way Americans Work. https://www.pewresearch.org/social-trends/2020/12/09/how-the-coronavirus-outbreak-has-and-hasnt-changed-the-way-americans-work/.

Phillips, B. (2007). *A Comparison of Autonomous and Collaborative Models in Computer-Mediated Communication*. PhD thesis, University of Victoria.

Pickering, M. J. and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and brain sciences*, 36(4):329–347.

Pierrehumbert, J. and Hirschberg, J. B. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. *Intentions in Communication*.

Pinet, S., Zielinski, C., Mathôt, S., Dufau, S., Alario, F.-X., and Longcamp, M. (2017). Measuring sequences of keystrokes with jsPsych: Reliability of response times and interkeystroke intervals. *Behavior research methods*, 49(3):1163–1176.

Pisarevskaya, D. (2017). Deception Detection in News Reports in the Russian Language: Lexics and Discourse. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing Meets Journalism*, pages 74–79, Copenhagen, Denmark. Association for Computational Linguistics.

Plank, B. (2016). Keystroke dynamics as signal for shallow syntactic parsing. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 609–619, Osaka, Japan. The COLING 2016 Organizing Committee.

Priva Cohen, U. (2010). Constructing Typing-Time Corpora: A New Way to Answer Old Questions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, page 7.

Raul, N., Shankarmani, R., and Joshi, P. (2020). A Comprehensive Review of Keystroke Dynamics-Based Authentication Mechanism. In Khanna, A., Gupta, D., Bhattacharyya, S., Snasel, V., Platos, J., and Hassanien, A. E., editors, *International Conference on Innovative Computing and Communications*, Advances in Intelligent Systems and Computing, pages 149–162, Singapore. Springer.

Rinker, T. W. (2016). Sentimentr: Calculate text polarity sentiment. *University at Buffalo/SUNY, Buffalo, New York. version 0.5*, 3.

Roffo, G., Giorgetta, C., Ferrario, R., Riviera, W., and Cristani, M. (2014). Statistical Analysis of Personality and Identity in Chats Using a Keylogging Platform. In *Proceedings of the 16th International Conference on Multimodal Interaction - ICMI '14*, pages 224–231, Istanbul, Turkey. ACM Press.

Rozumowski, A. V., Kotowski, W., and Klaas, M. (2020). Resistance to Customer-driven Business Model Innovations: An Explorative Customer Experience Study on Voice Assistant Services of a Swiss Tourism Destination. *ATHENS JOURNAL OF TOURISM*, 7(4):191–208.

Rumelhart, D. E. and Norman, D. A. (1982). Simulating a Skilled Typist: A Study of Skilled Cognitive-Motor Performance. *Cognitive Science*, 6(1):1–36.

Saevanee, H., Clarke, N. L., and Furnell, S. M. (2012). Multi-modal Behavioural Biometric Authentication for Mobile Devices. In Gritzalis, D., Furnell, S., and Theoharidou, M., editors, *Information Security and Privacy Research*, volume 376, pages 465–474. Springer Berlin Heidelberg, Berlin, Heidelberg.

Saidla, D. D. (1990). Roommates' cognitive development, interpersonal understanding, and relationship rapport. *Journal of College Student Development*.

Schilperoord, J. (2002). On the cognitive status of pauses in discourse production. In *Contemporary Tools and Techniques for Studying Writing*, pages 61–87. Springer.

Schmader, C. and Horton, W. S. (2019). Conceptual Effects of Audience Design in Human–Computer and Human–Human Dialogue. *Discourse Processes*, 56(2):170–190.

Scissors, L. E., Gill, A. J., Geraghty, K., and Gergle, D. (2009). In CMC we trust: The role of similarity. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, page 527, Boston, MA, USA. ACM Press.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. *The handbook of phonological theory*, 1:550–569.

Seo, S. H., Griffin, K., Young, J. E., Bunt, A., Prentice, S., and Loureiro-Rodríguez, V. (2018). Investigating People's Rapport Building and Hindering Behaviors When Working with a Collaborative Robot. *International Journal of Social Robotics*, 10(1):147–161.

Serafin, R. and Di Eugenio, B. (2004). FLSA: Extending Latent Semantic Analysis with Features for Dialogue Act Classification. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 692–699, Barcelona, Spain.

Shon, S., Brusco, P., Pan, J., Han, K. J., and Watanabe, S. (2021). Leveraging Pre-trained Language Model for Speech Sentiment Analysis. *arXiv:2106.06598 [cs, eess]*.

Shriberg, E., Favre, B., Fung, J., Hakkani-Tur, D., and Cuendet, S. (2009). Prosodic Similarities of Dialog Act Boundaries Across Speaking Styles. *Linguistic Patterns in Spontaneous Speech*, page 27.

Shriberg, E., Stolcke, A., Hakkani-Tür, D., and Tür, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32(1):127–154.

Shriberg, E., Stolcke, A., Jurafsky, D., Coccaro, N., Meteer, M., Bates, R., Taylor, P., Ries, K., Martin, R., and van Ess-Dykema, C. (1998). Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech? *Language and Speech*, 41(3-4):443–492.

Sisman, B., Yamagishi, J., King, S., and Li, H. (2021). An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:132–157.

Snow, D. (1994). Phrase-Final Syllable Lengthening and Intonation in Early Child Speech. *Journal of Speech, Language, and Hearing Research*, 37(4):831–840.

Sordoni, A., Galley, M., Auli, M., Brockett, C., Ji, Y., Mitchell, M., Nie, J.-Y., Gao, J., and Dolan, B. (2015). A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 196–205, Denver, Colorado. Association for Computational Linguistics.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. V., and Meteer, M. (2000). Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics*, 26(3):339–373.

Stolcke, A. and Shriberg, E. (1996). Statistical language modeling for speech disfluencies. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, volume 1, pages 405–408 vol. 1.

Stolcke, A., Shriberg, E., Bates, R., Coccaro, N., Jurafsky, D., Martin, R., Meteer, M., Ries, K., Taylor, P., and Ess-Dykema, C. V. (1998). Dialog Act Modeling for Conversational Speech. In *AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, pages 98–105.

Stromqvist, S. (2007). Influence of Typing Skill on Pause–Execution Cycles in Written Composition. In *Studies in Writing*, volume 20, pages 55–65. Emerald Group Publishing, Bingley.

Svec, J. G. and Granqvist, S. (2010). Guidelines for Selecting Microphones for Human Voice Production Research. *American Journal of Speech-Language Pathology*, 19(4):356–368.

Swerts, M. and Geluykens, R. (1994). Prosody as a Marker of Information Flow in Spoken Discourse. *Language and Speech*, 37(1):21–43.

Thorne, C. (2017). Chatbots for troubleshooting: A survey. *Language and Linguistics Compass*, 11(10).

Tickle-Degnen, L. and Rosenthal, R. (1990). The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, 1(4):285–293.

Tiong, L. C. O. and Lee, H. J. (2021). E-cheating Prevention Measures: Detection of Cheating at Online Examinations Using Deep Learning Approach – A Case Study. *arXiv:2101.09841 [cs]*.

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversation.* PhD thesis, University of Rochester.

Trott, S., Reed, S., Ferreira, V., and Bergen, B. (2019). Prosodic cues signal the intent of potential indirect requests. In *Proceedings of CogSci 2019*, page 7.

Tsimperidis, I. and Arampatzis, A. (2020). The Keyboard Knows About You: Revealing User Characteristics via Keystroke Dynamics. *International Journal of Technoethics*, 11:34–51.

Twenge, J. M. and Farley, E. (2021). Not all screen time is created equal: Associations with mental health vary by activity and gender. *Social Psychiatry and Psychiatric Epidemiology*, 56(2):207–217.

Van Waes, L. and Leijten, M. (2015). Fluency in writing: A multidimensional perspective on writing fluency applied to l1 and l2. *Computers and Composition*, 38:79–95.

Van Waes, L., Leijten, M., Wengelin, A., and Lindgren, E. (2012). Logging tools to study digital writing processes. *Past, present, and future contributions of cognitive writing research to cognitive psychology*, pages 507–533.

Villani, M., Tappert, C., Ngo, G., Simone, J., Fort, H., and Cha, S.-H. (2006). Keystroke Biometric Recognition Studies on Long-Text Input under Ideal and Application-Oriented Conditions. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 39–39.

Vizer, L. M. and Sears, A. (2017). Efficacy of personalized models in discriminating high cognitive demand conditions using text-based interactions. *International Journal of Human-Computer Studies*, 104:80–96.

Vogel, P. (2010). SCOTUS: From Pornography's 'I Know It When I See It'to Social Media's 'I Don't Get It'. *E-Commerce times.*

Walther, J. B. (2011). Theories of Computer- Mediated Communication and Interpersonal Relations. In *The Handbook of Interpersonal Communication*, volume 4, pages 443–479. Sage.

Walther, J. B. and Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. *Handbook of interpersonal communication*, 3:529–563.

Wang, J., Wang, J., Sun, C., Li, S., Liu, X., Si, L., Zhang, M., and Zhou, G. (2020). Sentiment Classification in Customer Service Dialogue with Topic-Aware Multi-Task Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9177–9184.

Wang, S. and Manning, C. (2012). Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 90–94, Jeju Island, Korea. Association for Computational Linguistics.

Yamaguchi, M., Crump, M. J. C., and Logan, G. D. (2013). Speed–accuracy trade-off in skilled typewriting: Decomposing the contributions of hierarchical control loops. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3):678–699.

Yeh, S.-L., Lin, Y.-S., and Lee, C.-C. (2019). An Interaction-aware Attention Network for Speech Emotion Recognition in Spoken Dialogs. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6685–6689, Brighton, United Kingdom. IEEE.

Yu, G. (2010). Lexical Diversity in Writing and Speaking Task Performances. *Applied Linguistics*, 31(2):236–259.

Yuasa, M., Saito, K., and Mukawa, N. (2006). Emoticons convey emotions without cognition of faces: an fmri study. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1565–1570.

Zhang, M., Zhu, M., Deane, P., and Guo, H. (2019). Identifying and Comparing Writing Process Patterns Using Keystroke Logs. In *Quantitative Psychology*, volume 265, pages 367–381. Springer International Publishing, Cham.

Zhao, R., Romero, O. J., and Rudnicky, A. (2018). SOGO: A Social Intelligent Negotiation Dialogue System. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 239–246, Sydney NSW Australia. ACM.

Zhao, R., Sinha, T., Black, A. W., and Cassell, J. (2016). Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavior. In *International conference on intelligent virtual agents*, pages 218–233. Springer.

Zhou, H., Huang, M., Zhang, T., Zhu, X., and Liu, B. (2018). Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Zhou, L., Shi, Y., and Zhang, D. (2008). A Statistical Language Modeling Approach to Online Deception Detection. *IEEE Transactions on Knowledge and Data Engineering*, 20(8):1077–1081.