# Chapter 2

# Related Work

This thesis brings together three areas of research that have thus far not been fully integrated: dialogue analysis, keystroke pattern analysis, and speech prosody. I begin by setting up prosody, especially *implicit* or *silent* prosody. I then introduce keystroke dynamics, providing a background as well as the diversity and depth of keystroke research. Following this, I present relevant aspects of sentiment analysis and rapport with a partner. Throughout this section, I highlight relevant prior studies that have preceded the studies in my thesis, and show how these studies can be expanded by the studies proposed in my thesis.

## 2.1 Explicit and Implicit Prosody

Prosody is the study of all the elements of language that contribute toward acoustic and rhythmic effects, but primarily those that occur above the individual sound or phoneme level, and instead are focused 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.

### 2.1.1 Implicit Prosody Hypothesis

Almost all studies of prosody 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, such as speaking rate, also impact the speed at which we comprehend and read language, e.g. we read at the same rate that the words would usually be spoken.

To measure silent reading speeds, researchers use high-precision procedures such as eye-tracking, which can measure exactly how long an eye is focused on a single word or sequence of words. Prior research has observed 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 (2014b).

### 2.1.2 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.

Previous studies of dialogue usually collect one of two types of data: spoken conversations, e.g. the Switchboard Corpus (Godfrey et al., 1992), where timing metrics are available but laborious to measure, and text-based messaging, e.g. an online chat such as (Liebman and Gergle, 2016a), where entire messages are analyzed, sometimes along with the time when the message was transmitted.

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. This also provides a significant opportunity for HCI and text-based CMC: It was traditionally assumed that when spoken speech is absent from a conversation, as in a text-based dialogue, that some source of information is lost, or at least significantly altered (Daft and Lengel, 1986). By investigating *implicit* prosody in text-based chats, I am able to capture aspects of sentiment and thought that were previously thought to be limited to *explicit* prosody. Finding evidence of prosody in text-based chat bolsters constraint-based theories of CMC, since it points to the notion that the same processes are involved in text-based CMC, but represented differently (Clark, 1996; Gergle, 2017).

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.

### 2.1.3 Spoken prosody and the current studies

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 their respective chapters, and so the list below is a brief introduction to the relationship between prosody and the studies in my thesis.

A more comprehensive list of parallels between spoken prosody and prosodic patterns in typing is produced in Appendix C. As my thesis does not look for direct analogs between speech and typing timing patterns, but rather is motivated by speech prosody and the IPH, these direct parallels are not immediately germane.

- 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 a user's opinion of their conversational partner affects their typing patterns. Studies such as Gravano et al. (2011) show how prosodic characteristics of speech affect social perceptions, while Li et al. (2017a) demonstrates that sentiment analysis can be significantly improved using prosodic information.

- 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.

As an example of a 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.

Ballier et al. (2019) provides an interesting preliminary look at the relationship between speech prosody and keystroke patterns, as well. Their findings are limited, and are primarily at the syllable-level, as opposed to higher-level analysis, such as sentence- or paragraph-level. Nonetheless, their results are still interesting and relevant to the present study, and provide a rationale for the distinction in my studies between *inter*-word pause timing and *intra*-word pause timing, because syllable-level pauses would primarily be evident only within a word.

Finally, it is well-established that many syntactic unit boundaries are well-marked by changes in spoken prosody (Vicsi and Szaszák, 2010), i.e. a noun phrase is demarcated by longer pauses or larger pitch changes at its edges. Plank (2016) looks at the granularity of syntactic information available in keystroke data. The authors ask whether pause times can be used to locate the boundaries of every prepositional phrase, or only the boundaries of every sentence. By adding pause time data to a feature-set in a bidirectional Long Short-Term Memory (bi-LSTM) model, the researchers are able to improve over baseline accuracy in part-of-speech tagging and chunk labeling. For this reason, the studies in my thesis also look at pause times at phrasal boundaries, which are very similar to syntactic unit boundaries. If these timing differences are important in dialogue typing, as well, then it points to the use of silent prosody in typing.

## 2.2 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 computer-mediated 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, and 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 (Locklear et al., 2014; Vizer and Sears, 2017).

### 2.2.1 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). Prior studies, on the other hand, have

estimated that transcribing an hour of speech data requires a trained researcher or professional anywhere from 4-10 hours (Bazillon et al., 2008; Novotney and Callison-Burch, 2010).

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 often negligible (Bridges et al., 2020; Pinet et al., 2017).

Another unique advantage of keystroke analysis is that the keylogs 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. For example, Lindgren et al. (2019) points out that a revision that is immediate versus a revision after a pause indicates different underlying cognitive processes.[1]

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

[1]While I do not delineate revision types in my analysis, but rather group together all revisions, the data is structured in such a way that this is trivial to extract, and will be a topic of future studies.

the reader, or changes their perception of the reader, analogous to audience design principles (Clark and Murphy, 1982; Horton and Gerrig, 2016).

### 2.2.2 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, Figure 2.1 reproduces a figure from Conijn (2020), which provides a succinct illustration of fundamental available features, which can be combined and expanded 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]

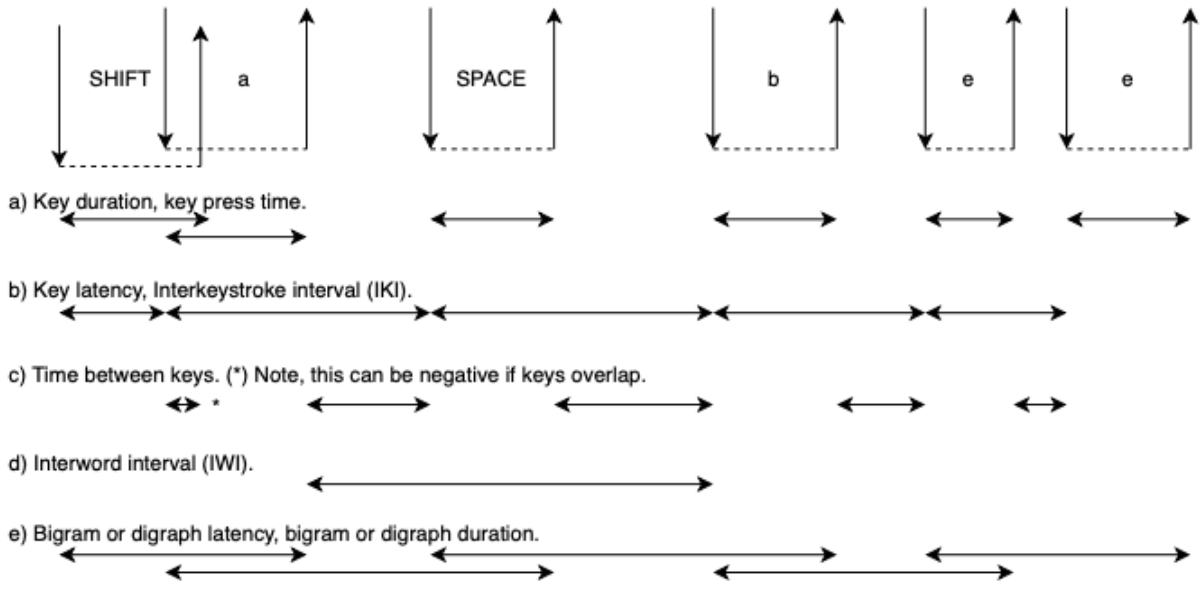


**Figure 2.1**
A set of features available from extracted keystroke timing. Reproduced from Conijn (2020)

[2]A more intuitive analog in speech to keypress duration would be phoneme or letter duration. However studies such as Lee et al. (2015a) 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.

Using the two dimensions of keystroke features (latency and duration), features of prior work can be organized into methodological categories. Table 2.1 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.

| Category | Examples in keystrokes | Analogues in speech |
|---|---|---|
| Pause timings or latencies | Interkeystroke intervals (IKI) between or within words, (e.g. Medimorec and Risko, 2017), or initial pause time, (e.g. Allen 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). |
| Revision 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 (e.g. 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 2.1**
Categories of keystroke features, along with possible parallels in speech production.

As mentioned above, one advantage of features such as those outlined in Table 2.1 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 (Brizan et al., 2015; Goodkind et al., 2017; 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. (Abadi and Hazan, 2020; Kołakowska, 2015, 2018).

### 2.2.3 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; Lee et al., 2015a; López-Carral et al., 2019).

Emotion can be classified either discretely or continuously, and keystrokes seem sensitive to both (Epp et al., 2011). In a discrete classification, emotions are categorical, e.g. happy, sad, neutral (Cowen et al., 2019). In a continuous classification, emotions are evaluated on dimensional spectrums, e.g. *valence*, or the degree of negativity or positivity, and *arousal*, the intensity of the evoked emotion (e.g. Lee et al., 2014). Lee et al. (2014) and Lee et al. (2015a) used the same general experimental framework but presented emotional stimuli visually and aurally, 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.

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. (2015a), 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.

My studies use takeaways from both research lines: sentiment and rapport are tested as both categorical variables as well as continuous spectrums that can be evaluated using a regression-type analysis. The studies cited herein seem to point to the value of both approaches.

Another important takeaway from these studies is that typing patterns are useful both as dependent and independent variables. My studies also extend this line of research. For example, Study 1 uses a binary classification task to measure how well a collection of keystroke variables can predict the correct dialogue act. On the other hand, within Study 2 I investigate how well the sentiment of an utterance along with the participant's overall opinion of their partner can predict the timing of specific keystroke patterns.

### 2.2.4 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 (Logan and Crump, 2011; Yamaguchi et al., 2013).

Vizer and Sears (2017) created a *continuous* classification system to measure cognitive demand. Unlike many classification studies that output a single discrete classification at the end of a training

instance, a continuous classification system is constantly updating and changing its predictions. In a conversational situation such as a game or a troubleshooting call, 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. This is why Study 3 also examines the effectiveness of using subsets of typing data to predict a typist's mindset. For example, Study 3 predicts overall rapport given only the keystrokes from the first half of the conversation. While this does not constitute a *continuous* measure of rapport, it does provide a foundation for further subsetting.

Effects of cognitive changes are relevant to my thesis because providing recommendations requires a different amount of cognition from receiving recommendations. Moreover, as seen in studies such as Branigan et al. (2011), different perceptions of an interlocutor require different amounts of effort to formulate an appropriate response. If these differences carry over to dialogues, then noticeable differences should exist when a dialogue act changes in Study 1 or a participant's role changes in Study 3.

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."

These studies provide motivation for the subdivisions of timing features used in my studies. For example, rather than solely considering the average overall dwell time for a user, my studies also look at features such as the average dwell time before or within content words.

### 2.2.5 Keystrokes in chats

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.

Buker and Vinciarelli (2021) used an experimental setup very similar to my own, and investigated similar questions. However, whereas Studies 2 and 3 in my thesis investigate a participant's opinion of their partner, Buker and Vinciarelli (2021) investigates how well keystroke dynamics can predict different personality traits. 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).

Borj and Bours (2019) used keystroke analysis to identify liars in a chat. The central finding was that being deceitful required more deliberate effort and less natural thoughts, and this different mode of thinking was evident in different typing patterns. This is relevant to Study 1 in my thesis, where a statement-dialogue-act might only report facts, whereas an opinion-dialogue-act might require personal imagination, though not necessarily deceit.

What ties all of these studies, as well as my studies, together is the notion that typing patterns reflect innate features of a typist. While this type of investigation is very relevant to HCI, my thesis instead focuses on the interactive aspect, and looks at how keystrokes can elucidate dimensions of an interaction, rather than a person in an interaction.

Each of the studies above *could have*, in theory, been conducted in isolation, since they only study each typist on their own. As a significant distinction, my thesis advances the stance taken by Clark (1996) that language is a game "designed for two" and can best be understood when looking

at a dyad, or call-and-response, between speakers. The studies in my thesis use typing patterns to better understand the nature of a relationship, and how these patterns reflect the way a speaker feels towards their partner.

### 2.2.6 Ethical issues with keystroke collection

This section should conclude with a discussion of the ethical issues surrounding keystroke analysis. Keystrokes are a "biometric" or personal identifier, like a fingerprint or iris scan (Banerjee and Woodard, 2012; Epp et al., 2011; Locklear et al., 2014; Monrose and Rubin, 1997b). 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. Goodkind et al., 2017; Tsimperidis and Arampatzis, 2020), the information contained in our keystroke patterns should be protected.

All of the experiments in my thesis obtained IRB approval, and all participant identities are anonymized during and after the experiment (see Section 4.4.1). I did wait to notify participants only *after* the experiment that their keystrokes were logged, so that they were not self-conscious about their typing. This notification, though, included 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.

## 2.3 Dialogue

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. 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." While my thesis does not directly address this problem, it does take it into account in the controls and limitations of the studies.

In contrast to interactions, linguistics has traditionally 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.

Another interesting way to view this distinction between dialogue and monologue 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 my thesis this is reflected by the fact that features are engineered not only from a message in isolation, but also how a message relates to preceding and following messages.

As mentioned in Section 2.2, 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 Borj and Bours (2019); Bukeer et al. (2019); Roffo et al. (2014) 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, rather than each partner in isolation.

### 2.3.1 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. My thesis uses a semi-naturalistic setup, where the dialogue is spontaneous, but the prompts and assigned roles are constant between experiments.

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*. Many types of adjacency pairs exist, such as *question-answer*, *greeting-greeting*, and *inform-acknowledge* (Stivers, 2012). Because of this, conversation has been studied at the group level (more than one person), as opposed to studying individuals and their mental processes in

conversation. Again, keystroke-level analysis of conversational text will allow us to bridge the gap between individuals and dyadic constructs: Language production is a reflection of internal cognitive processes, while the text of a conversation also reflects group-level interaction. Both of these are trivial to capture and measure using the data collection methodology in my thesis.

A 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.

Studies 2 and 3 further utilize the turn-taking nature of a dyadic construct, and specifically a spontaneous dialogue. Edelsky (1981) makes an important distinction in turn-taking between an *exclusive floor* and a *cooperative floor*. An example of an exclusive construct is a professor delivering a lecture, where they hold the floor until they entertain a question. On the other hand, cooperative floors exist only by the cooperative nature of a conversation, where one interlocutor waits for the other, but is free to interrupt at any time. For this reason, Study 2 only looks at turns that are not interrupted, so as to avoid any confounds introduced when the cooperative nature of a dialogue is violated. Study 3 looks at the ratio of interrupted and uninterrupted turns, to measure whether experimental partners are exhibiting the same level of cooperativeness.

My thesis will add more data to the language production processes that underlie turn-taking. Since CMC does not contain explicit non-verbal cues, and yet users engaged in CMC still report positive conversational experiences that are marked by few prolonged pauses, and few instances of one participant trying to type simultaneously to another participant typing a message, then signals

aside from intonation must also exist in conversation.[3] If cues such as intonation or gesture are not available, then my thesis may shed light on what cues are available, which help promote a positive conversational experience.

### 2.3.2 Sentiment analysis in dialogue

Most sentiment identification in dialogue has been performed on audio data (Shon et al., 2021; Yeh et al., 2019). 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 utilize both timing information and textual information.

A key difference between sentiment analysis in monologic text and sentiment analysis in dialogue is that monologues 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).

### 2.3.3 Computer-mediated communication in dialogue

As mentioned in the Introduction, it is important to constrain our discussion of dialogue to a specific subset: text-based computer-mediated communication (CMC). Although I will continue to use the term CMC, the term sometimes is too restricting. A recent review of literature titled its chapter on CMC "Discourse processing in *technology*-mediated environments" (Gergle, 2017). 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 AOL Instant Messaging. Technology-mediated

[3] However, see Riordan (2011) for a discussion of non-verbal cues used in CMC, such as emoticons and font changes.

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.

Below I provide a brief overview of relevant theories of CMC. While my thesis does not aim to provide a novel theoretical contribution to CMC, the findings of my studies bolster current evidence for certain theories. Further, the importance of theoretical inquiry should be underscored, because in a rapidly-changing world of technology: "We can't keep up with new innovations, so we need theory and models that can." (Scott, 2009, p. 754). While all of the theories below do not hold face-to-face communication as a "gold standard" (Gergle, 2017), they do all help to explain what strategies we use to replace the verbal and nonverbal cues used in face-to-face conversation, which are lost in text-based communication. (Riordan, 2011).

#### 2.3.3.1 Theories of CMC

Social Presence Theory posits that the cues present in face-to-face communication are filtered out, and that a lack of cues makes communication feel less intimate and involved (Short et al., 1976). As Walther (2011) points out, though, this theory can be challenged by the fact that, on various occasions, people intentionally choose to use alternative forms of communication, even when face-to-face communication is available.

Other approaches such as Media Richness Theory (Daft and Lengel, 1986) and Social Information Processing (SIP Walther, 1992, 2018) posit that different mediums make different channels of information available to its users and that users adapt to the affordances of different channels. Further, users are willing to adjust the time-course of information transmission to accommodate the medium, since relationship development takes longer in a computer-mediated environment versus a face-to-face encounter (Walther and Parks, 2002). As opposed to a bandwidth-based theory, where a necessary hierarchy is erected with certain mediums lacking channels available in other mediums, the channels available in a given medium are adaptable.

The affordance-based theory of Clark and Brennan (1991) looks at eight different constraints, each unique to the specific modality or technology being used to communicate. Further, Clark and

Brennan (1991) also frames grounding within a cost-benefit analysis, in that each technology or modality imposes different constraints or costs on communication, which in turn impose different costs on grounding. As an example, if a conversation is extremely important or higher-stakes, such as a space shuttle launch, an interlocutor will be willing to pay a higher cost or expend more effort in order to establish common ground. This may take the form of a repeated phrase, or a more verbose statement that provides every detail and additional word.

To relate affordances to text-based chats (and keystroke analysis), we can look at the affordances of Reviewability and Reviseability. The former affordance speaks to the ability to review one's own utterances, or a partner's utterances, and is manifested by the ability to look at the history of a chat in most platforms. The latter affordance concerns the ability to change or edit an utterance, both before and after it is transmitted. As (Chafe and Tannen, 1987) points out, the amount of effort a speaker needs to put into being understood is drastically different when a message can be planned, reviewed and revised. Since spoken language is not effortlessly planned or revised, the result of these affordances is that speakers are held less accountable than writers, since an addressee expects written language to be more accurate and articulate (Horton, 2017).

One of the main goals of this thesis 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. This, then could add support to a social information-based theory. For example, if in Study 3 participants consistently report high rapport, it could point to the idea that participants take advantage of differences in timing patterns to adapt their communication to a text-based medium, ad thereby provide evidence for a central tenet of SIP: "...communicators are just as motivated to reduce interpersonal uncertainty, form impressions, and develop affinity in on-line settings as they are in other settings." (Walther and Parks, 2002, 535).