Isak Enger

Language Variation Online

A Case Study of Two Internet Users' Writing on Four Online Platforms

Master's thesis in English

Supervisor: Annjo Klungervik Greenall and Stian Hårstad

June 2020

Master's thesis

NTNU

Norwegian University of
Science and Technology

Isak Enger

Language Variation Online

A Case Study of Two Internet Users' Writing on Four
Online Platforms

NTNU

Norwegian University of
Science and Technology

Abstract

This thesis uses data from two individual participants in order to investigate how English written online varies, and what potentially motivates such variation. It deploys a mixed method approach, where a quantitative set of linguistic features are identified and analyzed, and a qualitative set of interviews are performed in order to gather participants' reflections on their own writing. By first gathering authentic excerpts of text, from two unique online platforms per participant, and then locating and counting various salient features across all excerpts, the thesis establishes a general trend of variations made on each platform. This process established a wide set of differences in amount of several features between one participant's platforms, and fewer differences between the other participant's platforms. In order to get insight on the possible reasons or motivations behind such differences, or lack thereof, between platforms and participants, participants were each sent interview sheets containing a mix of general questions posed to both, and more specific ones unique to each participant. These were constructed on the basis of findings in their textual data. Interviews showed a general notion from both participants that the context for their writing was the most prevalent factor for their usage of variations. The participant whose data varied greatly across platforms attributed this to the differences between the purposes and audiences of the two platforms. The participant with less varying data across platforms states that he uses variation on both, but to a smaller degree on one platform, partially as a result of a smaller sense of community. The findings of the thesis are thus, primarily, that written English online varies through a set of features that point to a variety of meanings and communities. Participants deploy these features with varying degrees of awareness but are primarily motivated by the context the writing occurs in, their audience, and their mood. Furthermore, they are motivated by which of their identities they wish to express, and which affective and epistemic stances they wish to enact.

Sammendrag

Denne masteroppgaven bruker data fra to individuelle deltakere for å undersøke hvordan engelsk skrevet på nettet varieres, og hva som potensielt motiverer slik variasjon. Det brukes en blandet metodetilnærming, hvor et kvantitativt sett med språklige trekk identifiseres og analyseres, og et kvalitativt sett med intervjuer gjennomføres for å samle deltakernes refleksjoner rundt deres egen skriving. Gjennom å først samle autentiske tekstutdrag fra to ulike plattformer per deltaker, og så lokalisere og telle diverse fremtredende trekk fra alle utdragene, etablerer oppgaven en generell trend for variasjoner gjort på hver plattform. Denne prosessen etablerte et visst sett med forskjeller i mengde for flere trekk mellom den ene deltakerens plattformer, og færre forskjeller mellom den andre deltakerens plattformer. For å skaffe insikt i mulige grunner til eller motivasjoner for slike forskjeller, eller mangler derav, mellom plattformer og deltakere, ble hver deltaker sent et intervjuskjema som inneholdt en blanding av generelle spørsmål til begge, og mer spesifike spørsmål som var unike til hver deltakere. Disse ble konstruert basert på funn i data fra tekstutdragene deres. Intervjuene viste en generell forestilling fra begge deltakerne om at konteksten for skrivingen deres var den mest utbredte faktoren for deres bruk av variasjon. Deltakeren som hadde data som varierte mye på tvers av plattformene sine begrunnet dette i forskjeller mellom hver plattforms hensikt og publikum. Deltakeren som hadde data med mindre variasjon på tvers av plattformene sine fortalte at han bruker variasjon på begge, men i mindre grad på den ene plattformen, delvis som et resultat av en mindre følelse av fellesskap. Funnene til oppgaven er dermed, i hovedsak, at skriftlig engelsk på nettet varierer gjennom et sett med trekk som peker mot et mangfold av betydninger og fellesskap. Deltakerne bruker disse trekkene med varierende grad av bevissthet men er hovedsaklig motivert av konteksten skrivingen skjer i, deres publikum, og deres humør. Videre motiveres de av hvilken av deres identiteter de ønsker å uttrykke, og hvilke affektive og epistemiske holdninger de ønsker å vedta.

Acknowledgements

A five-year long path has ended, and I would like to extend my gratitude to those that were in any way involved, both near and far.

First, of course, I would like to thank both my supervisors, Annjo Klungervik Greenall and Stian Hårstad. Without their continuing interest in and support for this project and my ideas, it is safe to say there would simply be no thesis. Despite hardships of various kinds, they were always supportive, understanding, and helpful in the resolving of issues and towards progress of the paper. Thank you both.

I naturally also want to thank both my participants. Although you have received your fair share of gratitude throughout the process, I think you deserve a final thank you. Without your patience and willingness to help and accommodate throughout, this thesis would never have been completed.

I am grateful to all the people that I surround myself with, from both before and during my time at NTNU. In matters personal and academic, you've made the road easier. You should know who you are, and I thank you all equally.

Thank you to Even for proofreading and giving valuable feedback.

I also want to thank my girlfriend, Hanne, for her ongoing support. Thank you for listening to me airing my frustrations, answering my impossible questions, and letting me have an extra ice cream sandwich.
My parents, Britt and Rolf, who I know are, were, and always will be proud of me. My cousin, Joachim. Thank you.

Finally, a thank you to those whose music helped me get through it all.

Isak Enger
Trondheim, June 2020

Table of Contents

List of Figures

List of Tables

1 Introduction

In this paper, I seek to explore the idea of sociolinguistic variation within a selection of online spaces. Just as one's language is modified when speaking to a family member versus one's boss at work, it seems logical to assume that one's language is similarly modified when writing a private message to a friend versus a public post on Facebook. What, then, might motivate choosing one way of writing over another? What do the writers themselves think about their chosen forms, and do they even think about them at all?

These questions became the steppingstones for the development of a full research question: How does written English vary within a selection of online spaces, and what motivates this variation? In order to answer this question, the research analyzes a selection of authentic excerpts written by two users of English online. One of the participants provided text from the social news aggregate website Reddit and from a forum dedicated to music, while the other participant provided excerpts from her E-mails as well as chats on the Instant Messaging application WhatsApp. Additionally, the research interviews both participants in order to gain insight into what motivates certain ways of using language, that are identified by the analysis. Data from their excerpts, as well as data from their interviews, are used to discuss the research question in light of relevant sociolinguistic theory.

1.1 Structure of the Thesis

The paper will first explore and present relevant sociolinguistic theory generally in section 2.1. Section 2.2 discusses theory surrounding language as it unfolds online. Section 2.3 and 2.4 delve into identity and stance, as well as how identity is especially relevant for online language, which are key concepts within the thesis. Section 3.1 is dedicated to presenting methodological points that are relevant for the process of the research, in particular the manner in which data from text and interviews were acquired. Section 3.2 briefly discusses the process of finding willing participants for the research, and section 3.3 explores some ethical considerations that were taken into account throughout the research. Section 4 and 4.1 present the way textual data was processed and sorted, with sections 4.2 and 4.3 discussing findings from each participant and their platforms. Section 4.4 presents relevant points that were brought up during each participant's interviews. Finally, section 5 discusses salient findings from data in light of the theory, and section 6 provides a summary and brief suggestions for further research.

2 Theoretical Background

2.1 Sociolinguistics and Variation

Sociolinguistics is the main framework in this thesis because it concerns how language carries social significance. More specifically, one of its key concerns regards how variation in language carries such social significance (Chambers, 2003, pp. 1-2). Eckert (2000) views variation as a social practice by seeing speakers not as representatives, but as constituents of larger social categories, in the sense that they are not only members of such categories, but also the building blocks in their construction. Speakers both construct and respond to the social meaning carried by variation, such that variation is used not only to show an adherence to a group, but also as a way of creating, maintaining, and adding meaning to this adherence (pp. 3-4). Thus, variation not only spawns from a specific context, but is also used to affect it by virtue of its property as carrier of social significance.

Variation is brought on by several factors, which influence a speaker such that they might modify their language to fit a certain situation. Some of the more prominent factors lie in the social context, and there are a number of social factors that contribute to how one's language takes shape within a given context. According to Chambers (2003), social differences between two users of a language are a good example of this, as the number of such differences stand in direct correlation to the formality of the language. Age, for instance, is a readily apparent and common social difference, and therefore has well-documented effects on how language is modified towards a more or less formal tone. A simple illustrative example could be to imagine two women, whose only relation is that they are neighbors, meeting at a grocery store. The imagined conversation is likely to differ if we imagine them both to be in their 20's, compared to if one is her 20's and the other in her 60's. The latter situation would undoubtedly have a more formal air around it, and with it comes changes to language such as slower, more deliberate, and clearer speech (pp. 4-5). This is an intentionally simple example, and real-life equivalents are likely to have far more factors affecting the language at different levels. Nonetheless, it illustrates how language usage varies in response to the situation.

Variation itself takes many forms and, according to Chambers (2003), is perhaps easy to ignore by the average user of a language. When one says *I'm walking*, it is functionally the same as saying *I'm walkin'*. They both convey the same basic semantic meaning. The same is the case for saying *car* vs. *automobile* (p. 13); the two are synonymous, yet they are clearly phonologically and orthographically different. In the first example this difference is shown via reduction of the nasal velar /ŋ/ into an alveolar /n/, and the compression of the segment -ng into -n' in writing to reflect this change in sound. In the second, the two are orthographically and phonetically completely different words. Chambers (2003) further states that variation takes lexical, syntactic, phonological, and morphological form and is observed and produced all the time without most people making specific notice of it (p. 13). When we put focus on the variation itself, by for instance imagining the persons that said *car* and *automobile* respectively, we might imagine two different people with various backgrounds and lives that affect their specific

way of referring to the same unit. Alternatively, one could imagine the same person within a specific context that leads him or her to choosing one form over the other. While variation might go unnoticed in daily lives, most people have an idea that we tend to use different words in different situations and with different people, which the layperson may be able to bring to consciousness if prompted to.

According to Schilling (2013), "stylistic variation" is the key resource that users of a language employ in order to shape and reshape their personal identities, interactions, memberships, social orders, and attitudes.[1] Thus, variation works as a tool for users to enact a stance towards the ongoing conversation, and to project attributes such as 'learnedness' or 'precision'.[2] Furthermore, these attributes and stances can be associated with specific individuals, character types, or larger social groups, tying the local together with the global, as usage of a form shapes social identities and meanings. Simultaneously, the established social meanings shape what an individual can do with that language (pp. 327-8). For the above case of *car* vs. *automobile*, the user might produce the latter form in an attempt to project an attribute of 'learnedness', however depending on the context he may in one instance be successful and in another inadvertently project one of 'pretentiousness'.

2.1.1 Third Wave Variation

Within the study of sociolinguistic variation, Eckert (2012) identifies three distinct waves. Without going into too much depth, the first two waves viewed variation as an incidental result of social space, where speakers are static and their identities are equated with category affiliation. The first wave was characterized by a search for the connection between variation and predefined categories such as class, age, gender, and ethnicity, while the second was concerned with discovering rather than presupposing local and socially meaningful categories (p. 94). Relating the first two waves with the example above of *car* versus *automobile*, the variation between them would be tied to specific social categories, whether they are discovered or presupposed. According to Eckert (2012), the third wave instead sees variation as an essential feature of language rather than a result of a social space. Variation constitutes a social semiotic system that can express an ever-changing range of social concerns within a community (p. 94). While the three waves are distinct, they are not in direct opposition, and are instead considered research practices that influence each other (Quist, 2009, pp. 117-8).

2.1.2 Style

Instead of seeing variables as reflections of social categories, a third wave approach sees variations as elements in the construction of a style (Quist, 2009, p. 118), social meanings and categories (Hårstad, 2010, p. 14). A style involves how the same user of language will use this language differently on different occasions, rather than the ways in which language is used when different users talk differently from each other (Bell, 1997, p. 240). Eckert (2004) defines style as a *practice* rather than a *thing*. It is the visible manifestation of social meaning and the activity in which people create it. Neither style nor social meaning are static, and people combine existing resources as a way of

---

[1] Schilling uses this term to label the study of variation in individual speakers' speech, but the term is equally relevant for written language. The term is abbreviated to 'variation' throughout the paper.

[2] Stance is discussed further in section 2.4.

constructing brand new meanings, as well as twisting and tweaking pre-existing ones (p. 2). The focus within third-wave approaches is therefore more on the construction of identity and personae rather than predefined categories of language use, with the major point being how variables gain social meaning (Quist, 2009, p. 118), and how such variables are created and become meaningful resources to be utilized in the construction of styles (Hårstad, 2010, p. 14).

2.1.3 Indexicality

Style is thus a way of deploying language in order to create social meaning and personae. Bell (1997) explains that style is done in relation to other people and derives its meaning from associations between linguistic features and particular social groups (p. 243). The way this occurs is explained by Eckert (2012), in which she highlights the importance of Silverstein's (2003) indexical order. Briefly put, this order begins with a portion of a population becoming salient, and a distinguishing feature of their language attracting attention. Once this feature is recognized, it can be extracted from linguistic settings and used to index membership in that population subset from which it spawned. It can be used in ideological ways to show belonging to, or characteristics or stances associated with, that subset. Those outside of it may use the feature in order to invoke stereotypes associated to the subset. Repeated indexing leads to the conventionalization of the new sign, which opens it up for further indexical moves. All manner of linguistic features can come to index a variety of meanings. The process may seem linear but can progress in multiple directions at multiple points in time. An indexical meaning is therefore part of a larger network of linked meanings, where the context allows for the retrieval of a specific region of meaning (Eckert, 2012, p. 94). Thus, a single variable does not index a single meaning. As mentioned in the above *car* vs. *automobile* example, choosing *automobile* can in one context express a style by indexing a meaning that is formal, for instance in an effort to show belonging or adherence to a norm if the speaker is in a formal situation in which they wish to inhabit a formal role. In another context, such as among close friends, the same speaker can index a style of irony and playful mockery of formality by using the same variable.

2.1.4 Audience Design

Style is also considered by Bell (1997) to be modulated primarily as a response to the speaker's audience, rather than other factors, which he based on an experience of the same newsreader for two New Zealand radio stations differing his pronunciation during newscasts. The only variable between the two newscasts was the audience, and Bell (1997) refers to this as "audience design". He further explains that it occurs as a result of a change in the speaker's audience, and generally manifests itself via the speaker shifting their style such that it is more similar to that of their audience. Audience design applies to all levels of a language user's repertoire, from lower-level factors such as phonetics to larger-level ones, such as the shift to a wholly different language (pp. 244-5). Relating to an above example, the choice of *car* over *automobile* is then primarily affected by who is at the receiving end of either word. Bell (1997) further explains that one can style shift toward a third party, called a "referee", by using linguistic features that are associated with identification with that group (p. 248). This is in line with the concept of indexicality as discussed above, and a user of language can thus use the formal *automobile* to an informal audience as an intentional style shift in order to associate with a formal referee. Bell (1997) also explains that the topic at hand can cause a shift that echoes one that occurs in response to the addressee associated with

the specific topic. This means, for example, that one's style when talking about work echoes how one would talk to their boss (pp. 246-7).

2.2 Language Online

The Internet is unquestionably vast, and all types of language are used on it. Plenty of unique words, terms, abbreviations and morphologies have been introduced online, and Androutsopoulos (2006) presents "Webslang" and "Chat speak" as some of the labels proposed by laypeople in order to categorize language that is unique to the Internet (p. 419). Androutsopoulos (2006) points to an attempt made at naming the 'new' language, in which Crystal (2001) coined the term "netspeak". This was defined as a language that displayed features considered to be unique to the Internet, that arise as a result of the medium's properties as electronic, interactive, and global. Netspeak is considered a language variety rather than a wholly unique language, with its own set of sub-varieties. Thus, the "language of the internet" has, for example, a subcategory of "language of E-mails". In this sub-variety, features that are unique to E-mail, such as headers, greetings, quotations, along with more local points of variety, such as spelling variation, would count as the functionally distinct elements that serve to classify the language as "the language of E-mails". Similar distinct elements were found in "the language of chats", "the language of forums", and so on (Androutsopoulos, 2006, pp. 419-20).

Later approaches would problematize some of the suggestions made by Crystal (2001). Barton & Lee (2013) criticize such deterministic views of language online, since terms like "netspeak" consider the primary cause for new forms of language to be the technology itself and pay little to no attention to the contribution of contextual and social factors (p. 5). Androutsopoulos (2006) points out one such failing of "netspeak", which is to what degree it is viable to suppose "the language of E-mails" even exists, as its described criteria of common linguistic features are far outweighed by the many and diverse settings and purposes in which E-mails are composed (p. 420). Consider the following: I write two E-mails, where one is addressed to the head of the Department of Language and Literature at NTNU, and the other to a pen pal living in Chile that I have known for several years. Both would be E-mails, but the inclusion/exclusion of greetings or responsive quotations would undoubtedly vary between the two, as would the syntax, level of formality, attention to spelling etc., enough for the disparity between the two to go well beyond the local points of variety that Crystal (2001) described. Androutsopoulos (2006) further explains that the diversity of purposes and settings outweighing the common linguistic features is the case for other proposed 'languages' as well, such as "the language of chat". Finally, such an approach as that made by Crystal (2001), which Androutsopoulos (2006) dubs "internet linguistics", obscures the interplay between contextual, social, and technological factors in constructing language patterns online, nor does it adequately highlight how linguistic variability helps shape identities and interactions between people on the Internet (Androutsopoulos, 2006, pp. 420-1).

2.2.1 CMC

Given the flaws of the term "netspeak", it will not hold for the purposes of this paper. Instead, the more nuanced term "computer-mediated communication" (CMC) will be used as it is a similar but more matured term that, among other things, captures the importance of context and the user in online language and is open to more types of language on the Internet.

According to Barton & Lee (2013), CMC research draws upon existing linguistic concepts in an effort to understand online language. CMC is considered a variety of language characterized by features such as acronyms/initialisms, word reductions, emoticons, unconventional punctuation etc. (p. 5). Another important point regarding CMC research brought up by Barton & Lee (2013), that is important to the focus of this paper, is that you cannot separate written genres from their contexts or the users that wrote them. CMC language is both shaped by various social factors, as well as situated in specific contexts of use (pp. 5-6). In other words, Crystal's (2001) conceptualization of, for example, "the language of E-mail" is too limiting within CMC research, as it does not adequately take into account the manner in which the language is shaped by the specific social factors, as well as how it is deeply rooted in specific contexts of use. Despite the issues discussed in regards to concepts like "the language of E-mails", Barton & Lee (2013) point out that there are regular similarities and differences across modes of CMC (p. 6).[3] This is definitely the case, as most people would agree that the vast majority generally tend to, for example, use more formal language when composing an E-mail compared to when using instant messaging (IM). Barton & Lee (2013) expand on this by highlighting the fact that a CMC user might not use a mode's features in every context. They might also reappropriate the features of a mode into a different context, according to the needs of the current purpose for writing (p. 6). Thus, a user might write an E-mail with few or none of the features typical to E-mails, or they might appropriate features that are typical of E-mails into a non-E-mail context in order to achieve some effect. Barton & Lee explain that, in light of these new directions of research on CMC, the study of social variation in CMC language would emerge (2013, p. 6), and it is from this field this paper draws its usage of the term CMC.

2.3 Identity

A highly relevant aspect that motivates variation within CMC is identity, from which this paper's understanding is based on the definition made by Bucholtz & Hall (2009): "identity is the social positioning of self and other" (p. 18). Such an understanding is intentionally broad, and they further specify it by explaining that identity is not localized within any singular level of analysis, be that vowel quality, code choice, or ideological structure, but rather that it operates on several levels simultaneously. They privilege the interactional level, since it is on this plane that resources gain their social meaning (p. 19). Put simply, there can be no identity in language outside of an interaction, since the variations and tweaking of language that give rise to an understanding of an identity require not just someone that produces them, but also someone that receives and interprets them within a context. However, Bucholtz & Hall (2009) clarify that a language

---

[3] Barton & Lee's (2013) term "mode" is understood as analogous to platforms or wider contexts of CMC text production.

user's own sense of self is important in the construction of an identity, but that this is not where one's identity is housed exclusively. They further argue that identity will emerge from specific conditions of linguistic interactions, but they do not discount the possibility that resources used in identity construction in a specific interaction may derive from previously developed resources created in past interactions, with ideology being a key example (pp. 19-20).

Bucholtz & Hall (2009) make a further distinction in terms of levels of identity. On the upper level, there are macro identity categories, that include age, gender, and social class.[4] On a lower level, we find local and ethnographically specific cultural positions. Finally, on the lowest level, we find micro-level categories, which include those participant roles and stances which we temporarily enact according to the specific interaction. Here, identity emerges through fleeting roles or orientations that are specific to the interaction at hand. Although the micro-level is referred to as a lower level, it is an equally important category as those pertaining to sociological and ethnographic identities on the upper levels, and is an important contributor to how one forms subjectivity and intersubjectivity within a discourse (pp. 20-21).

2.3.1 Identity in CMC

Online, we have unique opportunities for modifying or playing with our identities. Barton & Lee (2013) explain that our identities, whether expressed through the language we use, the clothes we wear or the media we consume, are fluid. Depending on the social context, we may choose to obscure or flaunt one or more aspects of our identities. Some, such as age or gender may be harder to manipulate, while others such as our hobbies or friendships, can be easier to control and may change over time irrespective of our own choices or wishes, as a result of different contexts of interaction, or based on to whom we are asserting our identities. Therefore, it makes more sense to speak of plural identities rather than one singular identity (p. 68). Within this understanding of identities, their role within CMC is further complicated. Users can in many cases choose to obscure aspects of their identities, such as their age or gender. They can also further enhance the aspects they are particularly fond of, proud of, or more interested in, by having greater choice regarding where, when, and with whom they engage in social correspondence. Someone who is particularly fond of a band may delve deep into a forum dedicated to that very band, immerse themselves in discussion of their music, performances, members etc., all without revealing a single other aspect of their identity, be it age, gender, or other fields of interest.

Barton & Lee (2013) further highlight that CMC can be a somewhat restrictive mode of communication, particularly in text-based forms. Here, users are limited in their physical contextual cues, and must rely solely on the written word and their language to share their identities (p. 68). One might assume this would severely disrupt the conveying of identities online, however Barton & Lee (2013) explain that previous studies show how online participants use non-standard spelling to signal their belonging to a specific subculture (p. 68). Thus, when faced with limitations in physical contextual cues, we make strategic uses of language to construct and perform our identities through written linguistic means.

---

[4] As is discussed further in 2.3.1, these are the categories a user is able to intentionally manipulate in CMC.

2.4 Stance

As mentioned, the notion of identity is an important facet of social context, with which variation interacts. Jaffe (2009) explains that one of the ways in which people express their identity is through stance (p.10), which Du Bois (2007) defines as "… a public act by a social actor, achieved dialogically through overt communicative means, of simultaneously evaluating objects, positioning subjects (self and others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field" (p. 163). Jaffe (2009) also explains that stances are ways of indexing specific and shared cultural structures of feelings and norms and can be used to draw social boundaries that are core to social differentiation and categorization. Individuals may use stances to lay claim to particular identities and statuses, and additionally evaluate another individual's claims to identity and status (p. 7).

Barton & Lee (2013) elaborate that stance can be divided into affective stance and epistemic stance, where the former refers to the speaker or writer's feelings, while the latter signals knowledge and belief towards a statement (pp. 86-7). However, the boundaries between the two may not always be clear, especially in situations where people talk about themselves. Stance-taking is interactive and intersubjective, meaning it is often directed to a particular audience, which in turn has to interpret it (Barton & Lee, 2013, pp. 86-7). Stance may also be taken by choosing specific writing styles, such as an "expert voice" when discussing something they are knowledgeable about (p. 91). Bucholtz & Hall (2009) explain that, in the concept of indexicality, linguistic forms index interactional stances that come to be associated with certain social categories (p. 22). Stances are thus a part the indexical process of association between linguistic features and various identity categories.

3 Method

Method is a term for how one collects, processes, and analyzes data. The chosen method decides what information it is possible to gather, and it is therefore important to consider what data is needed to answer the question posed by the thesis (Akselberg & Mæhlum, 2014, pp. 74-5).

3.1 Choice of Method

Since the thesis of this research seeks to explore how variation in written English takes place online, and what potential factors drive such variation, it is clear that the method requires at least one immediate form of data: text from online spaces. This forms the backbone of the paper, with supplementary interview data to aid in analysis, and provide context and insight from the authors of respective text excerpts. To provide this data, two willing participants have sent text they have written on two platforms each within the past 5 years.[5] Each participant has come to an agreement with me in order to establish which two platforms their text is sent from. Agreeing to participate involved reading and agreeing with the requirements presented in a letter of information and consent (appendix 1). They were also sent an interview sheet via E-mail, that included both general questions that are the same for each participant (appendix 2), as well as questions based specifically on findings from the data each of them provided.[6]

3.1.1 Collecting the Textual Data

Due to a number of restrictions for gathering data, as well as ethical considerations discussed in section 3.3, the research came to be oriented around larger volumes of data from a small number of participants. Thus, the research is case-study oriented, where portions of excerpts of text are provided by two participants. Further considerations of time and scope, as well as considerations regarding workload for participants and their skills required for the task, led to an aim of a total of 6000-8000 words, divided into 3000-4000 per participant, and 1500-2000 per online platform as the minimum requirement, not including supplementary information such as contextual notes.[7] [8]

Through a process of selecting several potential candidates as well as refining the details of participation in the project, two willing participants were acquired that were informed of and agreed to the requirements for participating in the study. One participant is a male L2 speaker of English, referred to as "Richard", and the other is a female L1

---

[5] The term 'platforms' is used to label any website or application on which text is routinely produced or read by those that participate in it. For the most part, it refers to the respective platforms from which participants have sent their data.

[6] In order to preserve their anonymity, neither full questions nor answers regarding specific findings are presented in this paper.

[7] Contextual notes included information that made the interaction, purpose, or topics for posts or messages understandable to the researcher, while still preserving the anonymity of third parties and obscuring their sensitive information.

[8] To ensure a more equal amount of text from each platform, a minimum word-count was set rather than a message-count, since the amount of text in messages/posts varies across platforms.

speaker of English, referred to as "Maria".[9] In order to make the most of the limited number of participants, the aim was for the cases to be as diverse as possible, both in terms of personal life and internet habits. To that degree, recruitment was somewhat successful, since the two participants are quite different from each other in both aspects.[10]

Through her work, which she characterizes as "computer-mediated discourse analysis", Herring organized a set of analytic priorities for research within the field:

- technological variables such as synchronicity, size of message buffer, anonymous messaging, persistence of transcript, channels of communication (e.g., text, audio, video), automatic filtering;
- situational variables such as participation structure (e.g., public/private, number of participants), demographics, setting, purpose, topic, tone, norms of participation, linguistic code; and
- linguistic variables (or discourse features) such as structure (e.g., typography, spelling, word choice, sentence structure), meaning (i.e., of symbols, words, utterances, exchanges), interaction (e.g., turn taking, topic development, back-channels, repairs), and social function (e.g., identity markers, humor and play, face management, conflict).

(Herring, 1996, 2001, 2004, in Thurlow & Mroczek, 2011, p. xx)

Such a list can form a great analytical tool for the research at hand to allow for pinpointing of variables that may be carriers of social meaning or used in identity construction. However, it is not without flaw, and Androutsopoulos (2010) notes that researchers should strive to include a willingness to challenge assumptions about the distinctiveness of new media language as well as a need to shift away from a focus on the medium to a focus on the user within research, and that attention should be placed on the situated practices of the communicators (Thurlow & Mroczek, 2011, pp. xx-xxi). Bearing Herring and Androutsopoulos in mind, the research at hand benefits from both viewpoints. Herring's explicit list is a great tool for knowing what to look for in the data and is used in the construction of the tables used for analysis. Androutsopoulos' insights serve as a constant reminder that these points do not form a clear enough picture, and that the context and situations are inevitable and highly relevant variables to take into account. Thus, data on the participants' own thoughts and attitudes to their own text will be as relevant as the texts themselves.

The primary focus of analysis of text excerpts is on variations in linguistic form across distinct platforms. For instance, one participant may use more formal grammar on one platform when compared to another. Widerberg (2011) explains that textual data such as this benefits from being frozen in time, and cannot be affected by the researcher (pp. 16-17). By identifying certain salient linguistic forms, and comparing their appearance, or lack thereof, in a different linguistic setting, the research aims to gain insight on these variations through participants' views and attitudes towards them. Text excerpt data is

---

[9] Richard and Maria are put in quotes as they are pseudonyms. Several other details regarding both participants, such as Maria's country of residence, are omitted from discussion to preserve their anonymity. Despite their omission, they are still taken into account for the purposes of analysis but will not be presented or discussed.

[10] Some of these aspects are briefly elaborated on in sections 4.2 and 4.3.

provided in the form of screenshots, rather than copied and pasted text. While this method is slightly more demanding on participants, it was considered necessary in order to capture elements such as layout, structure, and use of emoticons, such that they too could be included in analysis. Elements such as emoticons are also part of the data set, since, according to McCulloch (2019), these can act as replacements for gestures or emblems that are normally lost in written language (pp. 156-59). No specific directions were given as to what should or should not be included in a screenshot. The only requirement was the anonymizing of sensitive third party information, the process for which was outlined in the letter of information and consent (appendix 1). They were also informed that any uncertainties in this regard could be voiced to me directly, to avoid sending any unusable or irrelevant excerpts.

Lüders (2015) states that empirical examples must be reconstructed, in an effort to preserve the privacy of participants by having their original content unavailable through search, while still maintaining the original meaning of the message (p. 82). For this reason, presentations of textual data consist of individual features, and longer sequences of text from participants' excerpts are not included. This is also the case for interview data, as questions and answers may include identifying information. Lüders (2015) stresses that her cases were particularly vulnerable, necessitating a fuller anonymization process (p. 94). The participants of this study, however, are not as vulnerable, and have expressed their contentedness with the way in which their data is processed.

3.1.2 Mixed Method Approach

Barton & Lee's (2013) approach to understanding language online is one where they combine the study of practices with analysis of text (p. 11). They stress the value of connecting the two, since without closely looking at texts, one would miss the actual linguistic products that are created online, and without studying people's lives as well as practices and beliefs about their own online text production, the dynamic nature of online language would be lost (p. 167). Likewise, the linguistic and the social are inseparable, and the study of both is required, such that a combination of quantitative, qualitative, and mixed methodological approaches may be employed in sociolinguistic research (Mallinson, 2018, p. 3). Simply looking at the text excerpts themselves is therefore inefficient (and likely more challenging) without some notion regarding their production and the individual practices surrounding them. Thus, the inclusion of some contextual information for the excerpts, as well as various information regarding the participants themselves are also necessary parts of the data set. This research is then analogous to a study done by Barton & Lee (2013) in which they studied how Hong Kong youth deployed multilingualism and multimodal resources when communicating via IM (p. 168).

Such a mixed method approach contrasts with what Horvath (2010) describes as a typical sociolinguistic quantitative/qualitative divide, wherein the former studies language change over time, and the latter is concerned with the meaning and use of linguistic variability (p. 10). The approach employed in this research is a direct result of its purpose, which seeks to both identify and to some degree quantify certain characteristics and explore what causes them, from the view of the producers of the text themselves. As a result of a mixed method approach, data can cover a wide span of relevant areas, and there is great opportunity for identifying features and gaining insight on a speaker's attitudes towards it.

3.1.3 Interview Method

Analysis of text is, as mentioned, not the only part of the data set, but is supplemented by the inclusion of data from interviews with participants. Widerberg (2011) explains that the inclusion of other points of view from other methods is a way of illuminating the project. She further states that an interview consists of the researcher using conversational approaches in order to elicit information, stories and understandings surrounding a topic (pp. 16-17). The qualitative approach concerns not only the circumstances of the person but focuses on dimensions of experience regarding these (Dalen, 2004, p. 17). This is the case for the interviews in this research, with the notable difference being that interviews are not vocal or face-to-face, but instead conducted via E-mail. This was chosen in order to save time and resources, as the organizational logistics and transcription of interviews would be an additional time-consuming process for a small-scale project such as this. This includes a risk of potentially limiting the data. Some participants may, for instance, lack the will or ability to express themselves accurately through text, and some might find the written medium to be cumbersome, whereas they may have been more prone to deep elaborations had they been able to speak face-to-face with the researcher.

The interview questions were constructed on the basis of findings found in their data as discussed in section 4. Questions were structured by presenting a salient feature and asking the participant to explain what, if anything, motivates the choice of this feature.

3.2   Choice of Participants

In order to achieve its research goal, the project needed to find willing participants. This was done using a combination of opportunity sampling and volunteer sampling. Opportunity sampling relies on the researcher's past knowledge or experience in order to find suitable participants (Brady, 2011, p. 2). Participants were identified based on my pre-existing knowledge of them fulfilling criteria for being proficient and frequent users of English online. However, participation in the project involved some degree of sensitive information, which Jupp (2011) identifies as a criterion, along with participants agreeing to participate, for volunteer sampling (p. 2).

3.3   Ethical Considerations

Kvale & Brinkmann (2009) highlight four areas that are typically discussed as ethical considerations for researchers: *informed consent, confidentiality, consequences, and the role of the researcher.* They further highlight that such considerations are not meant as problems to be solved, but rather that researchers operate within a field where it is important to be open to inevitable dilemmas, ambivalences, and conflicts that arise within such considerations (pp. 86-7), several of which arose during the process of this research. Kvale & Brinkmann (2009) also explain that considerations should not be applied mechanically, since their relevance is determined by situational factors. (p. 88).

While these considerations cover research somewhat generally, there is also a set of considerations specifically tied to internet research. The Norwegian National Research Ethics Committees (NESH) detail these, while adding that general considerations, such as the ones already mentioned, also hold for internet research. They further stress the importance, to all research that is socio-scientific and concerned with the humanities, of

securing the dignity and integrity of participants. In ethically evaluating one's internet research, NESH highlight four primary factors: *the publicness of the utterance, the sensitivity of the information, the vulnerability of the affected,* and *the interactions and consequences of the research* (NESH, 2019, pp. 5-6).

3.3.1 Informed Consent

Kvale & Brinkmann (2009) define informed consent as participants in the project being informed about the research's overarching goal and main design points. Additionally, participants are to be informed about potential risks or benefits as a result of participating, so that their participation is completely voluntary and that they have a right to withdraw from the research. They should also be informed about procedures involved in the project, and how their data is handled. There is also a question of how much and when information should be given to participants. Care must be taken not to inform too much, so that the data might be affected, and informing too little, so that participants might be partially unaware of what they are partaking in (pp. 88-9). This point was taken into account, and details such as the precise manner of analysis or a preview of interview questions were omitted to avoid interference in both the textual and interview data. NESH (2019) highlight that informed consent is not solely a matter of mandates by law, but also of an ethical responsibility of the researcher, even when it is not particularly sensitive or personal. They also stress that, when collecting data from online platforms, it is not sufficient to point to general consent given when signing up for online forums such as Facebook, and in such cases it is necessary to reacquire consent while clarifying the purpose of the research (p. 13-14). All these aspects have been considered for the present project, in accordance with the requirements set by The Norwegian Centre for Research Data (NSD). The letter of information (appendix 1) that was sent to participants as part of the collection of their consent covers all these points and was constructed based on NSD's model text for such letters. The letter itself was approved by NSD as being in accordance with European law

3.3.2 Confidentiality

Confidentiality, as explained by Kvale & Brinkmann (2009), means that private data that might identify participants is not revealed (p. 90). In order to achieve this, there is often a process of anonymization which can, according to NESH (2019), be particularly challenging when doing internet research as a result of technological prerequisites and the publicness of the utterance. Since information is in most cases stored online, there is the possibility of participants being identified through searching for their utterances. This means that the researcher cannot make the same guarantees for confidentiality when doing internet research, and they should take care to inform participants of this when collecting their consent (p. 16). These considerations are also made for this project, and the identity of those involved is not revealed through the project. Only some personal information is gathered, such as gender, general age range, and information regarding their social lives and backgrounds. As a result of the paper not presenting longer sequences of text, the searchability of participants' utterances is low, as is the risk of them being identified in this manner.

There is, however, a small possibility that someone within shared social circles of me and the participants could identify them through some of the information presented here. This risk is hard to eliminate but very unlikely, and is therefore considered a non-issue. Another result of feedback from NSD was the storage location for the data, which went from being stored on a personal computer to being stored within NTNU's secure cloud-storage services.

Another highly relevant point of ethical consideration to this project was that of third parties. Since the project asks for authentic excerpts with sufficient context so as to properly analyze them, the inclusion of text from third parties in interactions became a factor. NESH (2019) highlight that the four primary factors mentioned in section 3.3 are particularly relevant for information originating from a third party. Data from third parties could be sensitive and should be processed accordingly to ensure their privacy (p. 13). In order to avoid the inclusion of sensitive third party data, several steps were taken to ensure that only non-sensitive third party information was included in the data. These included 1) expunging third party's username and sensitive information (such as addresses or private matters) found within their text, 2) in some cases completely excluding their text, replacing it with a restructured contextual note, 3) encouraging participants to avoid the selection of excerpts that include particularly sensitive information 4) having these steps be performed by participants, ensuring that third party data was already anonymized before the researcher could begin analyzing it. In order to accommodate for an increase in amount of work for participants, the total word-count was reduced per platform from about 3000 to 1500. A guide for sufficient anonymization (blocking out names and text by, for instance, drawing a black line over it with various computer software) was also included in the letter of information (appendix 1), and participants accepted these terms.

3.3.3 Consequences

Kvale & Brinkmann (2009) explain that consequences within qualitative research entails keeping in mind both the potential damages and benefits that a participant might experience as a result of their participation in the research. The ethical principle of beneficence states that the risk of injury to a participant should be as low as possible. The sum of potential benefits should outweigh the risk of damage, and the researcher should reflect over potential consequences not just to participants, but to the group(s) that participants come to represent through the research. Researchers should also be aware that the openness and intimacy of qualitative research might be alluring to participants, such that they might reveal information they later could come to regret sharing (p. 91). There is no explicit or direct benefit to participating in the research at hand; they were not promised any material reward or compensation, and the risk of them being identified is very low.

3.3.4 Role of the Researcher

As pointed out, the researcher themselves is the most important tool in the analysis of text excerpts. The researcher, however, is not a neutral entity. Dalen (2004) explains that preconceptions consist of the opinions and conceptions we have in advance of the phenomena being studied. Such preconceptions are inevitable, and the point is not to ignore or repress them, but to use them to open up to better understandings of a participants' experiences and statements (pp. 18-19). Dalen's statements pertain

specifically to interviews, and her points are vital for analysis of this type of data within this project. They are equally important for the other set of data for analysis, since I, as researcher, carry my own preconceptions towards text excerpts, for instance based on which platform they are taken from. Barton & Lee (2013) expand on this, by making note of the researcher's position as researcher being enabled by their role as insiders or active users of a specific online arena. The multiple angles of participation and research on a website leads to a variety of researcher roles, where boundaries between them can be fuzzy and overlapping, which has, in their experience, been of great benefit. One such benefit is in the form of being insiders and active participants in the website, since they then come pre-equipped with knowledge of, for instance, IM-specific language features or specific emoticons (p. 176). These points are analogous to my own experiences as a user of a variety of online platforms, allowing for a beneficial angle of approach to my analysis of excerpts through knowledge of certain conventions.

3.3.5 Publicness of Utterances

NESH's general rule for researchers, that they should collect consent and inform participants when processing their sensitive data, is made more complicated when observation occurs in public spheres, such as on the street or at plazas, or if the person of study is a public person. Despite information being openly available, it does not necessarily mean that it is public. Just as one cannot gather data from intimate, private conversations at a public café, nor should one gather such data from an openly available internet forum without properly informing and collecting consent from those affected (NESH, 2019, p. 9). Lüders (2015) adds that online content is not "up for grabs" without researchers bearing privacy in mind simply because it is public, and that clearly private groups or accounts must have their use consented to and anonymized (p. 81). While this research does collect information from public internet arenas, care is taken so that the content for analysis is from consenting participants, and that third parties are fully anonymized and their content is of a non-sensitive nature.

4 Analysis

Data was analyzed through the identification and quantification of various salient features.[11] These are presented below, and the following sections present the findings from each participant's platforms. Features were sorted into tables and used to evaluate the level of formality in an excerpt. Tables 1, 2, 3, and 4 are abbreviated forms of the full tables that show the total tally of features, as well as the percentage each feature makes up out of the total word count for that platform. The full tables are included as appendices 3, 4, 5, and 6.

Features are non-exclusive, meaning that an emoticon is counted both as an emoticon and a non-standard. By presenting amounts of features, the paper establishes somewhat general quantities of features for each platform and for each participant, therefore allowing for the identification of trends, and insight into which features are more/less frequent on which platforms.

There were numerous interesting findings in both the data and interviews. In the interest of keeping the size of the paper within bounds, only a selection of salient findings will be presented in this section and later discussed in section 5.

---

[11] 'Features' refers to any salient linguistic element that has been or can be varied. For instance, *u'll* and *you'll* count as features since they are both reductions of *you will*, where the former is non-standard.

4 Analysis

Data was analyzed through the identification and quantification of various salient features.[11] These are presented below, and the following sections present the findings from each participant's platforms. Features were sorted into tables and used to evaluate the level of formality in an excerpt. Tables 1, 2, 3, and 4 are abbreviated forms of the full tables that show the total tally of features, as well as the percentage each feature makes up out of the total word count for that platform. The full tables are included as appendices 3, 4, 5, and 6.

Features are non-exclusive, meaning that an emoticon is counted both as an emoticon and a non-standard. By presenting amounts of features, the paper establishes somewhat general quantities of features for each platform and for each participant, therefore allowing for the identification of trends, and insight into which features are more/less frequent on which platforms.

There were numerous interesting findings in both the data and interviews. In the interest of keeping the size of the paper within bounds, only a selection of salient findings will be presented in this section and later discussed in section 5.

---

[11] 'Features' refers to any salient linguistic element that has been or can be varied. For instance, *u'll* and *you'll* count as features since they are both reductions of *you will*, where the former is non-standard.

4.1 Overall Platform Data

The features identified and used in analysis are as follows:

- Emoticons (abbr. Emot): Use of emoticons.[12] Emojis are also counted as this feature.[13]

- Contractions (abbr. Cont): Grammatically standard contractions, such as *she'll* instead of *she will*.

- Errors (abbr. Err): Errors that are presumed to be unintentional, in the form of misspellings or mistakes in grammar. Example: *after the meetingvv*.

- Expletives (abbr. Expl): Swear words or otherwise unsavory language. Example: *fuck*.

- Non-Standard Punctuation (abbr. Npunct): Cases where punctuation deviates from standard usage, most commonly through repetitions such as *??* or *....*

- Non-Standards (abbr. Nstand): Non-standard words or forms of words, that are still recognizable as regular words. Example: *gurl* and *yyyyoooouuuuu*.

- Word Reductions (abbr. Wred): Compressing one or more words into smaller units. Separate from contractions in that they are non-grammatical or uncommon reductions. Example: *alts* and *boome*r.

- Non-Standard Grammaticality (abbr. Ngram): Cases where words or sentences are grammatically non-standard/incorrect, without sacrificing the readability of the word(s). Most commonly through lack of apostrophes in words like *shes* and *Im*.

- Acronyms (abbr. ACR): Acronyms and initialisms such as *OP* or *PM*.

- Full Capitalizations (abbr. CAP): Whole words that are fully capitalized, and that are not normally capitalized fully (such as acronyms). Example: *you SHOULD* or *why the FUCK*.

- Formality (abbr. Formal): A general ranking of the formality of the text contained within the image, separated into High (H), Middle/Mixed (M), and Low (L), in tables 1-4 and appendices 3-6. These rankings were based on the above features, where, for instance, a high number of expletives and low amount of grammatical contractions were indicators for a low level of formality.

---

[12] While emoticons generally refer to representations of facial expressions, they, and particularly emoji, also occasionally represent various objects, animals, or activities, all of which were counted for analysis. Most emoticons were representations of facial expressions.

[13] Emojis refer to the recently popular style of emoticons that come included in several mobile phones and are part of recent versions of Unicode, which is "the universal character encoding, maintained by the Unicode Consortium. This encoding standard provides the basis for processing, storage and interchange of text data in any language in all modern software an information technology protocols." (Unicode.org, 2019).

Features were identified on the basis Barton & Lee's (2013) description of CMC, which they consider a variety of language that is categorized by features such as emoticons, acronyms, unconventional punctuation etc. (p. 5), as well as the points highlighted in Herring's list of analytical priorities, including typography, spelling, word choice, symbols, words, utterances, (Herring, 1996, 2001, 2004, in Thurlow & Mroczek, 2011, p. xx), as presented in section 3.1.1.

Features were counted as tokens rather than types. A type is something abstract and unique, whereas a token is a concrete particular (Wetzel, 2018). Thus, if an excerpt contained the acronym *lol* (a singular type) three times in the text, they were counted as three tokens for the total tally within the tables. Counting of tokens was done manually, where each excerpt was read and reread multiple times, searching for tokens of a singular feature each time. Information that was censored counted as a single word. Contractions of several words, as well as numbers, emoticons, or repetitions of punctuation marks counted as a single word. The number of tokens of each feature within an excerpt was used to evaluate a degree of deviation from a standard. For instance, a high number of emoticons was a high level of deviation from a standard. These were combined with the researcher's intuition and evaluation to assign a level of formality per excerpt, where more deviations were indicative of a low level of formality. The levels of formality per excerpt were then used to judge the overall formality of the total data within a given platform.

This approach to evaluating formality is not without flaw. For instance, comparing the formality of WhatsApp versus E-mails can be seen as an unfair prospect, since the vast majority of E-mails will be more formal than the vast majority of WhatsApp messages, or formal in different ways. One could, for instance, write formally both in E-mails and on WhatsApp without one's language being identical on both platforms. One possible approach to this issue is by comparing data from participants to data present on the platform in general. For instance, one could compare Maria's WhatsApp data towards a variety of other users' WhatsApp data and compare their level of formality. In this paper, however, I chose to limit myself to the data being collected and analyzed. This is in large part due to the added time and resources required to collect similar data for comparison, but also because the research is case-oriented, meaning it is mostly interested in the cases themselves. Therefore, the level of formality is judged in comparison to the general level of formality in all the other excerpts from all platforms and both participants, leading to most of the E-mails to be considered highly formal, and most of the WhatsApp excerpts being considered informal.

Another issue regarding formality is how it is evaluated via the various categories. It is not the case that there is a direct link between, say, lower usage of acronyms to a high level of formality independent of platform. For instance, when writing on WhatsApp, it might seem unnatural and therefore less standard to not use the acronym *LOL* rather than fully write it out in a more formal way as *laughing out loud*, whereas writing it as an acronym in an E-mail would in most cases make the E-mail less formal. Likewise, some acronyms are sufficiently common, such as *ASAP,* so that they do not greatly affect formality in, for instance, an E-mail. These considerations are taken into account when analyzing the data. Much like how it would be wrong to say "wearing a suit is the correct attire for every occasion", so too is it wrong to say that "being formal is the correct use of language for every occasion". There are several cases where acronyms are standards within a context, however, to avoid a situation where I misinterpret or misunderstand a

usage, I am counting all acronyms as non-standards, regardless of the context they are in or my own familiarity with them.

In this paper, contractions are seen as features of formality, since they are directly contrasted with non-contracted word reductions, such as *hes* instead of *he's*. One could certainly argue that usage of contractions are also features indicating lower formality, when compared with, for instance, writing academic papers, where *he is* would be a more formal and appropriate form. However, as mentioned, evaluations of formality are done within the four platforms discussed in this paper, and it is clear through data from excerpts and interviews that using contracted forms are features that indicate a higher level of formality.

Another point worth mentioning is that the curation of excerpts was kept to a minimum, only asking for a mix of shorter and longer excerpts of text. This was done so that participants were able to represent themselves, for better or for worse. This likely affects the data, through the random omission of some highly interesting excerpts. Alternatively, it might lead to the data more evenly representing the participants, since they were not requested to send data they evaluated as good or bad examples.

4.2   Maria

Maria is an American woman in her early-mid 20's. She has resided in various states across the United States, and has briefly lived in some European countries, one of which she currently resides in. She has gone through various schooling on account of her frequent moving but has completed a university Programme in the Humanities. She states that she enjoys the online communities she engages in through playing video games, and that she has always found it easier to socialize through a screen rather than in person.

4.2.1 E-mails

Data from Maria consisted of 18 images of E-mails she had herself sent. The total word count for the E-mails was 1631, excluding words in the title of the E-mail and names in the signature. Since Maria lived in Europe during the data-gathering period, she uses a few non-English words and phrases in her E-mails. These were included in the word-count, since they form meaningful markers of identity and language usage, despite not being English.

Sorting of E-mail data into tables revealed a very high level of formality. Only two emojis were used, both within the same E-mail (a heart and a bee). The amount of contractions was the highest out of all the sets of data, as 4.78% of words in the E-mails were contractions, such as *I've* and *you'd*. Furthermore, there were very few errors, mostly consisting of very minor concord mistakes that are easy to make. Maria used no expletives in any of her E-mails, along with no fully capitalized words and no non-grammaticalities. Her non-standards and word reductions were very low, with her Npunct at 0.67% and Nstand at 1.34%. Finally, she used a few acronyms, but none that affected the level of formality, such as *ASAP* and *pdf*, and her Wreds were low and consisted entirely of her writing *ya'll*.

Thus, it is clear that the level of accuracy, grammatical correctness, and avoidance of non-standards were some of the major markers of Maria's language in E-mails.

Table 1: Total tally of features within Maria's E-mail excerpts.

E-Mails	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	CAP	Formal
Total features	2	78	3	0	11	22	9	0	4	0	H
% of total word count	0.12	4.78	0.18	0	0.67	1.34	0.55	0	0.24	0	

4.2.2 WhatsApp

Data from Maria also consisted of 30 images of conversations on the messaging app WhatsApp, which is a typical messaging app, allowing for long or short messages, emoticons, pictures, voice clips etc. The total word-count for the WhatsApp messages was 1585, and only the messages sent by the participant herself were counted.

Sorting of Maria's data from WhatsApp revealed a direct opposite to her data in E-mails. Whereas her E-mails had the highest percentage of contractions from any of the platforms from both participants, her WhatsApp data consistently scored the highest in every category except contractions. Thus, her WhatsApp messages were considered to be overall low in formality, however, there were a selection of excerpts or individual messages wherein her formality rose to a middle level. Use of emoticons was the most frequent in Maria's WhatsApp messages, consisting mainly of emojis depicting various facial expressions. However, she also used regular emoticons such as *:)* and *:(,* sometimes depicted as *:))*. Her contraction use was at 3.22%, which is the second lowest out of all four sets of data. She has the highest amount of errors at 0.31%, mostly consisting of smaller ones such as writing *fuckinh* instead of *fucking*. She has the highest number of expletives, mostly consisting of swears such as *fuck* and *shit*, and the highest number of acronyms, most of which were common ones such as *omg* (oh my god) or *btw* (by the way). She also uses a lot of non-standards in grammar and punctuation. These come in the form of omissions of apostrophes in words like *shes* and frequent repetition of punctuations such as *????* or *.......* She also makes frequent use of fully capitalized words, the highest out of all four platforms by far, at a total of 3.86% of all her words. Finally, she uses a lot of word reductions wherein she omits the final letter in words such as *yea* or *goin*. It is clear that her language on WhatsApp is quite different from her E-mails, as it is marked by her being less grammatically accurate and making frequent use of non-standards of various kinds.

Table 2: Total tally of features within Maria's WhatsApp excerpts.

WhatsApp	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	CAP	Formal
Total features	21	51	5	18	34	94	60	36	12	61	L
% of total word count	1.32	3.22	0.31	1.13	2.15	5.94	3.79	2.27	0.75	3.86	

4.3 Richard

Richard is a Norwegian Man in his late 20's. He has completed normal Norwegian schooling through Upper Secondary (Videregående) and has taken some university courses. He is currently enrolled in a Master's Programme at a Norwegian university. He considers music his one true passion, both listening to various bands and creating his own. He also enjoys reading and writing, the latter of which he describes as being motivated by knowing someone will read what he writes and that it can offer help or entertain someone. Additionally, he notes he has spent "an absurd amount of time gaming", but he considers it a pastime rather than a "genuine interest".

4.3.1 Reddit

The first half of the data from Richard consisted of 10 images of posts he had made on the popular social news aggregate website Reddit. Reddit is composed of several smaller communities devoted to discussions of various topics, all of which vary in number of active members. Members of Reddit post content which can be rated up or down, where a high/low rating increases/decreases a post's visibility to users. Users can also comment on posts or reply to other comments, and comments are rated in the same way as posts. All of Richard's images are of comments on other people's posts, and the ratings for posts or comments were not included in the data. The total word count for the Reddit posts was 2103, and only text written by the participant himself was counted.

Going through Richard's data, it became clear his overall formality was at a high level, though a few excerpts were at a middle level. He used only two emoticons, one an emoji depicting a wide grin, while the other was the traditional *:(*. On Reddit, he used the lowest amount of contractions at 2.94%, which was the lowest out of all four platforms, though he used 0 non-grammatical contracted forms such as *shes*. He made few errors, mostly consisting of minor concord mistakes, and very few cases of fully capitalized words. He only has two instances of non-standard punctuation, one being *...* and the other being a seeming usage of a *:(* in place of a period, as the word right after it has its first letter capitalized. He uses few expletives, mostly consisting of a few instances of *shit* or *fuck*, and very few word reductions such as *boomer* and *alts*. Finally, his use of acronyms is fairly low, mostly consisting of more widely used ones, such as *IMO* (in my opinion). He also uses a few domain-specific acronyms such as *LFR* (looking for raid) or *OP* (overpowered) in a discussion about the videogame *World of Warcraft*.
It is thus clear that Richard's language on Reddit is marked by a certain degree of formality, though with some room for usage of non-standards and expletives.

Table 3: Total tally of features within Richard's Reddit excerpts.

Reddit	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	CAP	Formal
Total features	2	62	4	8	2	30	3	0	15	4	H
% of total word count	0.09	2.94	0.19	0.38	0.09	1.42	0.14	0	0.71	0.19	

4.3.2 Musical Forum

Data from Richard also consisted of 12 images from a forum dedicated to discussing a popular rock band, as well as bands from a similar era or style. While the main focus of the forum is discussion of the band and surrounding aspects, it also has separate space for general discussion, separated into smaller categories, such as discussion tailored specifically for those that are musicians themselves. The total word count was 2031.

Richard's data from the Musical Forum showed a high degree of formality. His data was consistently low in most categories, with the exception being his contractions, at 3.29, which is the second highest out of all participants and platforms. He used only three emoticons and he had the lowest amount of errors at 0.14%, no non-standard grammaticality, and only one instance of a fully capitalized word, namely *HATE*. He had the second highest amount of contractions, and a few expletives in the form of *shit* and *fuck*. He used a fair amount of word reductions, though these mostly consisted of shortenings of various band-names being discussed. His usage of acronyms included common ones like *PC* (politically correct) and *PM* (personal message), while others are acronyms for band-names, band-members, songs, albums, or members of the forum. These points make it clear that Richard's language on the forum is marked by a high level of formality. His language is accurate and informative, and his few deviations from these are through non-standards that are either fairly common or part of an internal discourse.

Table 4: Total tally of features within Richard's Forum excerpts.

Forum	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	CAP	Formal
Total features	3	67	3	5	2	5	11	0	10	1	H
% of total word count	0.14	3.29	0.14	0.24	0.09	0.24	0.54	0	0.49	0.04	

4.4   Interview data

The following will focus on a selection of statements made in reply to interview questions that are of most immediate and clear relevance to the research topic at hand, as well as previously presented theory. Interview data and brief references to theoretical points that pertain to participants' statements will be presented together. This is done in order to meld the empirical findings with the theoretical points made in section 2, as well as acknowledge that my findings in the interviews are colored by the theoretical viewpoints discussed earlier in the paper.

4.4.1 Interview Findings from Maria

One of the immediately apparent insights from Maria's interview was her explanation that her language is affected by contextual factors regarding the conversation, which, to her, primarily relies on the person(s) at the other end as well as the subject matter of the conversation. When asked to describe, using her own words, her written language online, she answers: "On some platforms, carefully. On others, comfortably. It also depends on who I'm talking to." Another point pertains to her usage of emoticons, where she states that they are directly reflective of the recipient's usage. If they use them, Maria will too, and if they don't, neither will she. Her data reflects this, where the only instance of emoji usage in an E-mail is explained as being in response to the recipient's previous E-mail's inclusion of them. Likewise, most of her WhatsApp messages show usage of emoticons when the recipient can also be seen using them. This is in line with Bell's (1997) theory on audience design, wherein the audience to the language production is the primary factor that leads to style shifting and variation such that it is more similar to that of the audience (p. 244). By using emojis in an E-mail, Maria exemplifies the critique made by Androutsopoulos towards Crystal's (2001) "netspeak" term (Androutsopoulos, 2006, p. 420), as presented in section 2.2., Maria can be seen to 'break the rules' of the "language of E-mail" by including emoji, which are typically not part of this common set of features.

Maria states that the person she is talking to is a key factor that affects her level of formality. She elaborates that she is more frequently informal when writing to those she considers to be close friends, and formal when writing to more socially distant recipients. Additionally, she explains that the level of formality is also modulated by the subject matter: "If I'm conversing with someone that I typically speak informally to about a serious topic, I tend to write more formally." This also carries over to platforms themselves, as she elaborates that she tends to write informal, quick responses on WhatsApp, while E-mails tend to be more elaborate and formal. She also claims that she considers E-mails to mimic a letter-writing format, and that she is more "serious-minded" when writing them. On this point of formality, she shares some insight on her usage of *ya'll* which is a frequent form used by her regardless of platform. She states that she considers it an integral part of her language. "I don't think that it is very informal, but I typically would still refrain from using it in very serious, job/academic papers or applications." Her reflection regarding her usage of *ya'll* shows some level of metaknowledge regarding variation. She is aware of the manner in which *ya'll* might index a less formal language variation, and, as Silverstein (2003) explains, that it can be used to show belonging to, association with, or stances towards a subset of a population (Eckert, 2012, p. 94).

Maria also reflects on smaller variances in her formality present in her overall less formal WhatsApp excerpts. She explains that she consciously chooses to write longer responses more formally, since they are often the result of a more serious subject matter. She considers longer responses to be more thought-out, compared to shorter ones that she describes as "quick-thought responses". Maria elaborates further by explaining that her choices of including/excluding apostrophes in contracted words, capitalization of *I* or *I'm*, and word reductions, are a mix of conscious/unconscious decisions. She states that she will "dumb-down [her] writing to dissipate tension, or move to another lighter subject", or they may happen as a reflection of her mood. She elaborates that if she is in a serious mood, she tries to be more grammatically correct, while if she is relaxing, she will shift her language subtly to reflect that. One example of this that she elaborates on is through her use of emoticons, explaining that she uses the emoticon *:)* in a more relaxed way as compared to emojis, with variations such as *:))* indicating the emphasis of some degree of comical approach. Maria's views on her writing are clearly indicative of some level of knowledge regarding identities at a micro-level, as discussed by Bucholtz & Hall (2009, pp 20-21), wherein roles and stances are enacted temporarily in relation to the specific interaction.

She elaborates on the repetitions of punctuation common to her WhatsApp excerpts, explaining that using multiples of a question mark *???* emphasizes a feeling of shock and questioning, whereas a single *?* indicates a more genuine or serious question "without any other emotional attachment to it." She further explains that she likes to seem excited but is uncomfortable with overuse of exclamation points in E-mails since they go against her desire to seem serious and formal on this platform. She expresses that she therefore attempts to balance these two points in her E-mails through frequent usage of single exclamation marks, thus showing both expressiveness and seriousness. Both of these can be seen as both affective or epistemic stances, or mixes of the two, as explained by Barton & Lee (2013, pp. 86-7).

4.4.2 Interview Findings from Richard

When asked how he views his own online writing style in English, Richard calls it "definitely formal". He elaborates that he makes sure to keep a formal tone and use proper spelling online in an effort to be taken seriously, though he also tries to avoid seeming "stuffy". His attention to accuracy and level of formality in his own language is possibly a result of English being his second language. Similarly to Maria, he claims that his formality is related to the context and situation in which the writing is occurring.

He explains that elements such as avatars give users of the Forum a visual identity, along with a closer sense of community, which leads to his tone being affected to a larger degree when posting on the Forum compared to Reddit, which he considers as "one huge chat room." He feels that Reddit is "disconnected", due to the large amount of varied discussion in posts and comments, and that there is no need for a persistent personality. He states he has a more solidified identity on the Forum due to consistently identifying elements, such as his avatar and join date, being visible in all his posts. He expresses that this leads him to, for instance, joking around and confiding more with members of the Forum, and that he has a more "consistent personality" on this platform. On Reddit, on the other hand, he considers the possibility that he is overall more informal and states that "it's possible I adopt a certain personality, mood, goal or style just for that thread." He states that he treats the Forum with more gravity, while being more light-hearted,

compassionate and "younger" on Reddit, describing the platform as "chaos", where his posts are "noise in the system". Despite these different attitudes towards each website, he states that his vocabulary across them is consistent. Richard's reflections on his attitudes toward Reddit and the Forum are clearly in line with the idea of micro-levels of identity as presented by Bucholtz & Hall (2009): Richard enacts smaller, fleeting roles depending on the specifics of the interactional context, which he seemingly does more often on Reddit than on the Forum, such as projecting a "younger" identity.

On the point of colloquialisms or expressions such as "that's how you roll", he describes them as giving him a feeling that he appears more proficient in English or more authentic. He states that he loves playing with language, and that he has a habit of writing down funny or memorable phrases. He also states that, when writing online, he is able to show aspects of his personality that he does not normally show in real life, and that language is a tool for this. He explains that his use of such expressions or "informalities" is part of adopting a certain character. He expresses that he feels that certain references or idioms indicate a masculine and carefree attitude, while others indicate childishness, curiosity, intelligence, etc. When discussing his use of non-standard or informal language in some of his posts, he states he has "sort of entered a different mode where I'm writing as someone who is not quite myself, and then it feels natural to use that kind of language." Richard is therefore aware, to some extent, of the indexicality of language, which pertains to the use of linguistic features to show belonging to, association with, or stances towards a subset of people to which the feature is associated (Silverstein, 2003 in Eckert, 2012, p. 94). Richard associates specific expressions or phrases with specific people or personalities, which he applies to a given situation in order to adopt a character.

When discussing abbreviations of bands and terms on the Forum, he explains that they are common practice on this platform. He elaborates that this is to save space and time, since all members are familiar with the abbreviations for names, terms, and titles, providing a sense of professionalism. Further insights into abbreviations relate to ones used in specific gaming communities, such as *OP* (overpowered) or *LFR* (looking for Raid). One major insight from Richard is that "a lot of times it would not make sense to use the full form." He explains that this is caused by several factors, including the abbreviation not directly mirroring the grammatical meaning of the full form, and that they not only describe the activity, but also an attitude towards it.[14] Finally, he comments that gaming communities are fond of abbreviations, and that his usage is directly influenced by the fact that "everyone" uses them. Use of abbreviations is one of the features that, according to Barton & Lee (2013), is part of categorizing CMC (p. 5). Richard's reflections regarding them can also be related to the concept of indexicality, as they are a clear way of showing belonging to a specific subset of people. This, in turn, becomes a way of displaying one of several identities, since these forms function, as Barton & Lee (2013) describe them, as strategies to signal a belonging to a sub-culture (p. 68), by using very specific forms that are exclusive to a relatively small subset of people.

---

[14] An example of this, Richard explains, is showing naïveté or delusion when referring to an actual 'Raid' (an activity in the videogame World of Warcraft) as an *LFR*, since the latter is considered a low-effort version of the former.
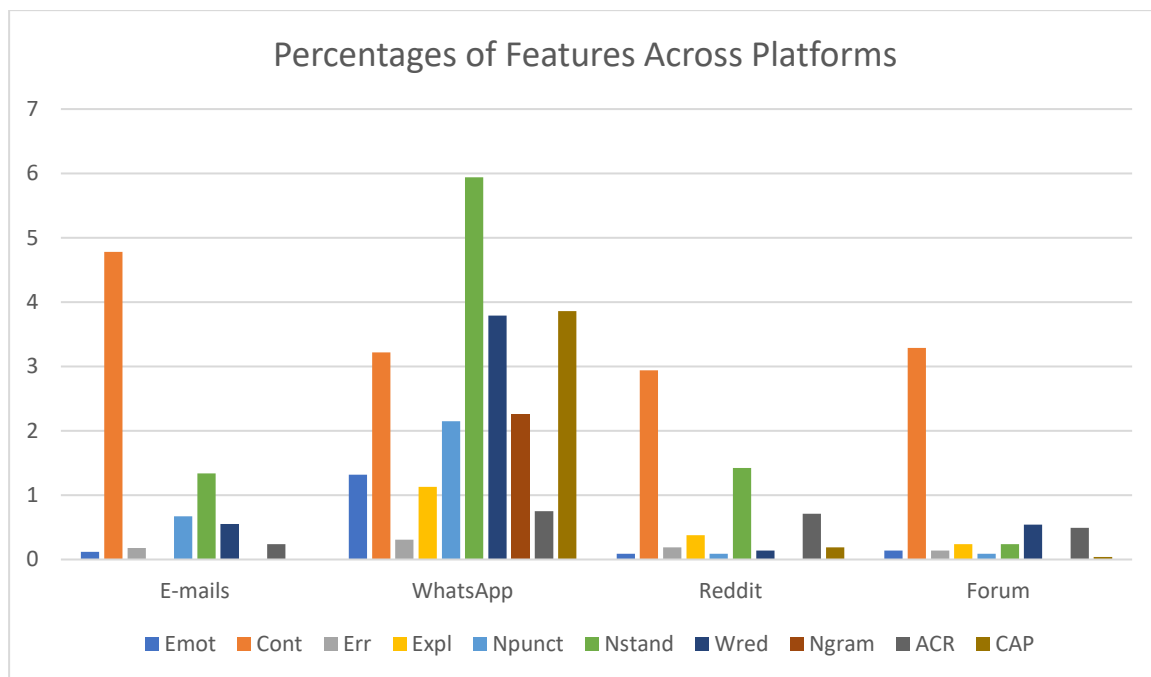
5 Discussion

The purpose of this paper is to investigate variation in the usage of English in digital arenas and possible reasons for and meanings of such variation. Due to the limited number of participants, and the relative ephemerality of certain internet phenomena, the observations made cannot be generalized for larger groups of people. They are representative solely of the actual participants, within the timespan from which their data is collected. There are a lot of points one could discuss based on the data and theory. Due to limits in space, the following sections bring up a selection of the most salient ones.

The variations identified in the data came in the form of minor adjustments made in response to specific contexts that, to varying degrees, added up to larger shifts in expression of identity. In Maria's case, the differences between E-mails and WhatsApp messages showed very noticeable variations between the two platforms. Her E-mails were well-worded and grammatically correct, with zero expletives, zero capitalizations of full words, and zero non-grammatical contractions. Comparing this to her WhatsApp messages, her language varied in that she used many more non-standard forms of words and punctuation, used a number of expletives and capitalizations of full words, as well as a number of non-grammatical word reductions. Additionally, on WhatsApp, there were also variations identified that relied on specific contexts. This is discussed in section 5.3.

In Richard's case, the differences, or lack thereof, between Reddit and the Forum showed fewer overt variations between the two platforms. Overall, most of the variation between platforms resulted from Reddit excerpts containing more non-standards and acronyms. The total amount of acronyms on Reddit was 15, and 9 of them were counted from a singular text excerpt, as a result of the context allowing for these acronyms as standards among members of specific groups. The number of non-standards was a result of a single excerpt in which there were a number of italicized words and sentences. Furthermore, Richard's high amount of word reductions on the Forum compared to Reddit were the result of a few excerpts containing 2-4 of these, while the rest had 0-1. The excerpts with the highest number of word reductions were all ones relating to musical terms or reductions of band names. These points illustrate that variations done by Richard were not largely affected by whether he wrote on Reddit or the Forum, but rather the specific needs of the individual writing situations. A possible cause for this may be that the two platforms share several similarities, whereas E-mail and WhatsApp are more contrasting platforms.

The differences in the textual data between all four platforms are visualized in Figure 1 below (in order to account for differences in message sizes and word counts, the data in all five figures presented below has been converted to show features as percentages of the total word count for each platform).

Figure 1: Percentage of Features Across All Four Platforms

The following sections will explore some possible explanations of variations, through various factors that might influence them. In light of the above figure, which illustrates differences in percentages of features of platforms, the discussion of possible explanations will start with the potential influence of the platform.

5.1   The Platform

The gathered data is insufficient for showing if these differences are a direct result of the platform itself, especially in the case of Richard, whose data is overall quite similar on both his platforms. However, data from interviews with both participants shows that they are conscious of where their text is being written, and that they perceive themselves as accommodating their writing accordingly.

The flaws in Crystal's (2001) concept of "Netspeak" being its overreliance on common linguistic features within different platforms (Androutsopoulos, 2006, pp. 420) are reflected in Maria's data. Her E-mails certainly fit into a mold of common features, such as high formality through accurate and grammatically correct language, as well as the inclusion greetings and signatures, which are typically associated with writing E-mails. However, some of her E-mail excerpts notably deviate from this mold, for instance through her use of emojis in one of them, frequent usage of non-repeated exclamation points, and non-standard words such as *ya'll* or greetings in a local European language (even when E-mailing non-speakers of that language).[15] This example also relates to Barton & Lee's (2013) theorizing around CMC, in that it highlights the importance of contextual and social factors, and that one cannot separate any text from these. They also acknowledge that there are regularities within one mode (p. 5-6), which is clearly seen in the data as Maria's E-mails are undoubtedly more formal overall in comparison to her WhatsApp messages.
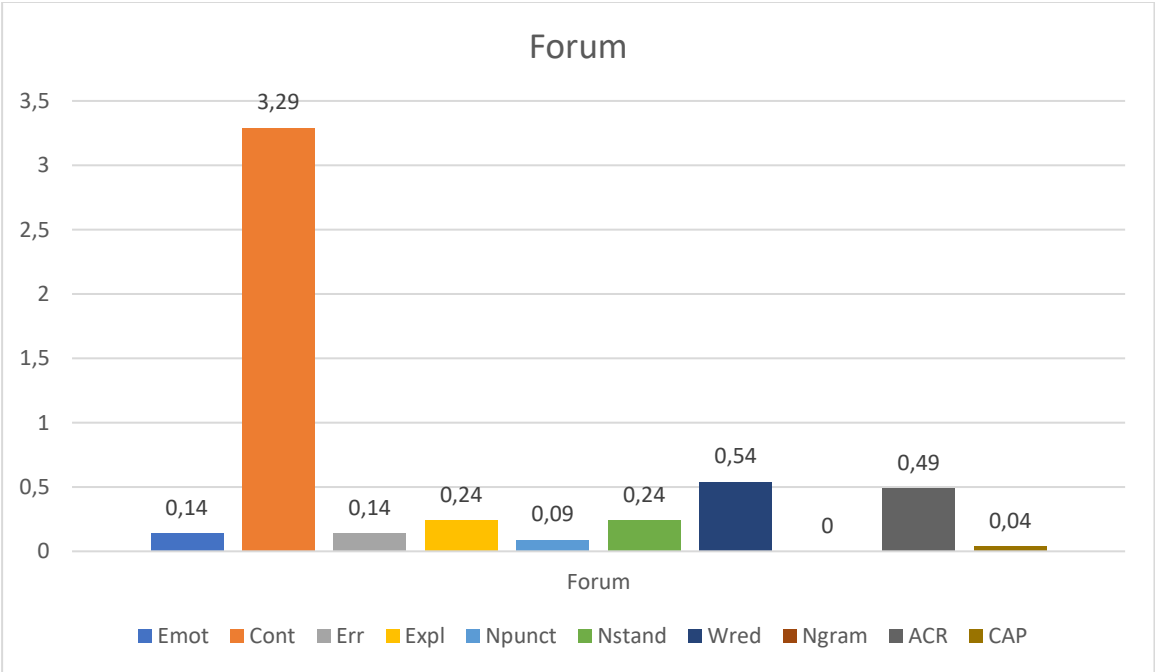
---

[15] 'Non-repeated' punctuation refers to singular instances, such as *!* or *?*, as compared to repetitions of the same punctuation mark, such as *!!!* or *??*
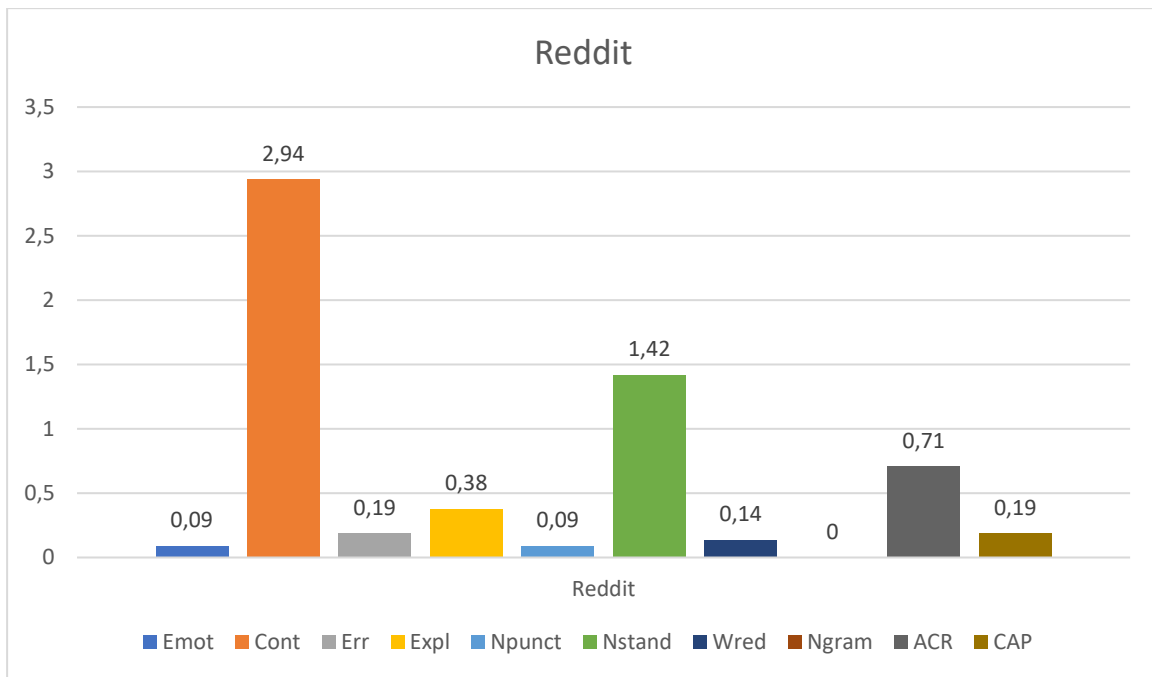
5.2   Identity

Eckert's (2000) view on variation as a social practice is reflected in the findings of this research through, for instance, Richard. He responds to social meanings carried by certain idiomatic phrases, as he associates them with proficient and authentic language, and uses these phrases to construct a unique social meaning when he inhabits a specific character or identity. He also shows group membership and his relation to practices common within these groups and communities through his use of very local word-forms, such as abbreviations for specific bands or more well-known gamer-language such as *LFR* and *OP*. This is also in line with Schilling's (2013, pp. 327-28) definition of linguistic variation, as Richard both knowingly and unknowingly projects various attributes that are contingent on the variations he chooses. These two approaches coalesce into the Third-Wave view of sociolinguistic variation, as described by Eckert (2012). Variation constructs a style by indexing ideological associations with a subset of people. This style is then deployed and displayed, with some degree of awareness, in order to show belonging to this specific subset. The style is therefore one part of a larger identity, wherein this identity is local and restricted to specific interactions. By using variations that are linked to communities of gamers, such as *OP*, Richard can show both members and non-members that he associates himself with such communities. The lines between which variations fit which contexts are, however, quite fuzzy. Both *LFR* and *OP* can be used within a specific World of Warcraft-community successfully to show belonging to an identity, but *LFR* may not function to this regard for other videogames where the term is not used. *OP*, on the other hand, may function on a more general level in gaming related contexts. Thus, there likely exists a number of identities with minute differences between them, each deployed in unique and specific situations and communities according to the writer's wishes for the expression of their identities.

As Barton & Lee (2013) point out: online, people find themselves relying solely on the written word to communicate their identities, since physical cues are limited (p. 68). Thus, variation might play an even bigger role in digital written forms, as it becomes the primary way of signaling an identity or a belonging, with a greater level of control as we can choose to obscure other aspects of ourselves. Like the adage goes: "On the Internet, nobody knows you're a dog." Again using Richard as an example, we see how this occurs through the attention he gives his identity within the musical Forum, where he is highly conscious of his more consistent identity/set of identities and therefore his style and tone in relation to the other members. This is also seen in his data from Reddit, where he enacts smaller and more transient identities that are guided by and shape the specific contexts. This relates to Bucholtz & Hall's (2009) concept of micro-levels of identity, which play out differently on each of the platforms. While Richard's language is more consistent across his platforms than Maria's language is across hers, we see that he enacts more and smaller identities on Reddit than on the Forum. Richard states that he feels less connected to the posts and the community on Reddit. On the Forum, however, he feels he has a more solidified identity/set of identities as a result of a larger sense of community, and that he is more recognizable by other members of the Forum, leading to him describing himself as more "serious" and "self-conscious" when writing on this platform. He also has a few topics he is uncomfortable with discussing on the Forum, but that he feels he can talk more openly and freely on Reddit, thus allowing him to adopt more varied and unique styles on the latter platform.
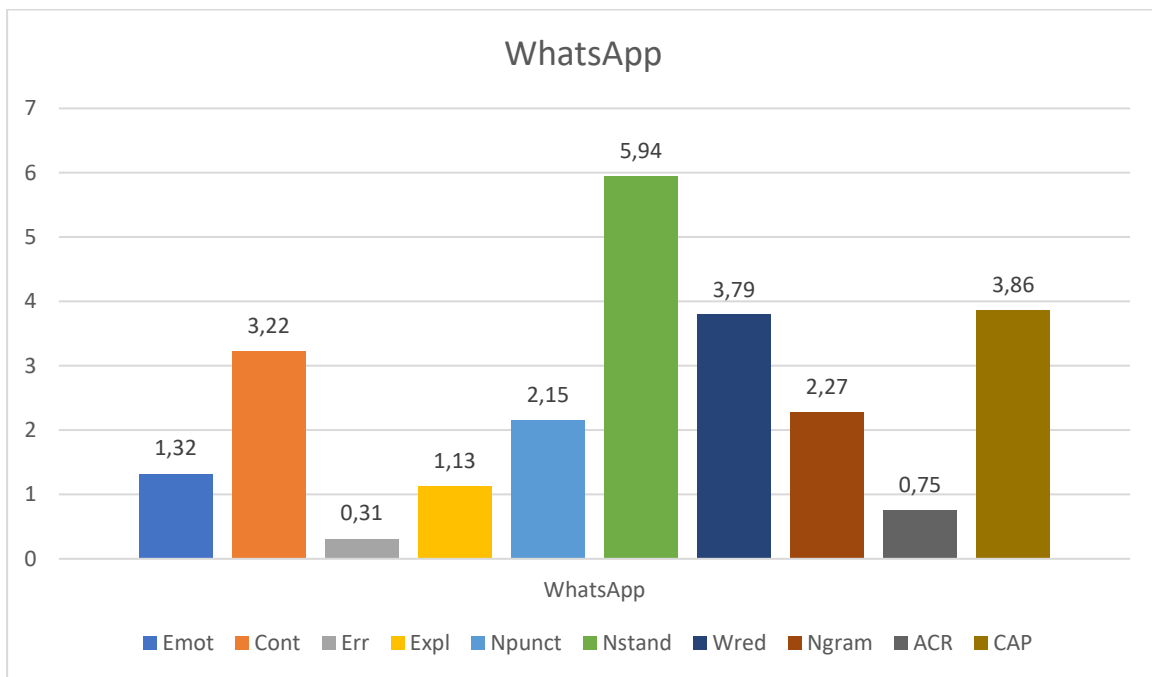
The differences are reflected in the data (Figures 2 and 3) where we see that his posts on Reddit display more variability in certain features. His use of non-standards, acronyms, and expletives are slightly higher on Reddit, while data from the Forum is slightly more even throughout. This indicates that, while his language certainly varies on the Forum, it varies to a smaller degree than on Reddit, seemingly as a result of his willingness to explore and project several identities on Reddit, rather than sticking to a select few identities he is more comfortable with sharing from on the Forum. The same difference in expression of identities across platforms is the case for Maria, though with much more readily apparent differences across her text excerpt data (Figures 4 and 5). Here, too, we note much higher amounts of features in her WhatsApp data when compared to her E-mail data. Her E-mails are largely formal, since she wishes to inhabit an identity that projects a serious and professional tone. Her language on WhatsApp, however, is more characterized by expressions of micro-levels of identity that rely more on the specifics of the interaction. Thus, her E-mails are more consistently formal, whereas on WhatsApp she might dip in and out of identities whose level of formality depends on the current topic of conversation.
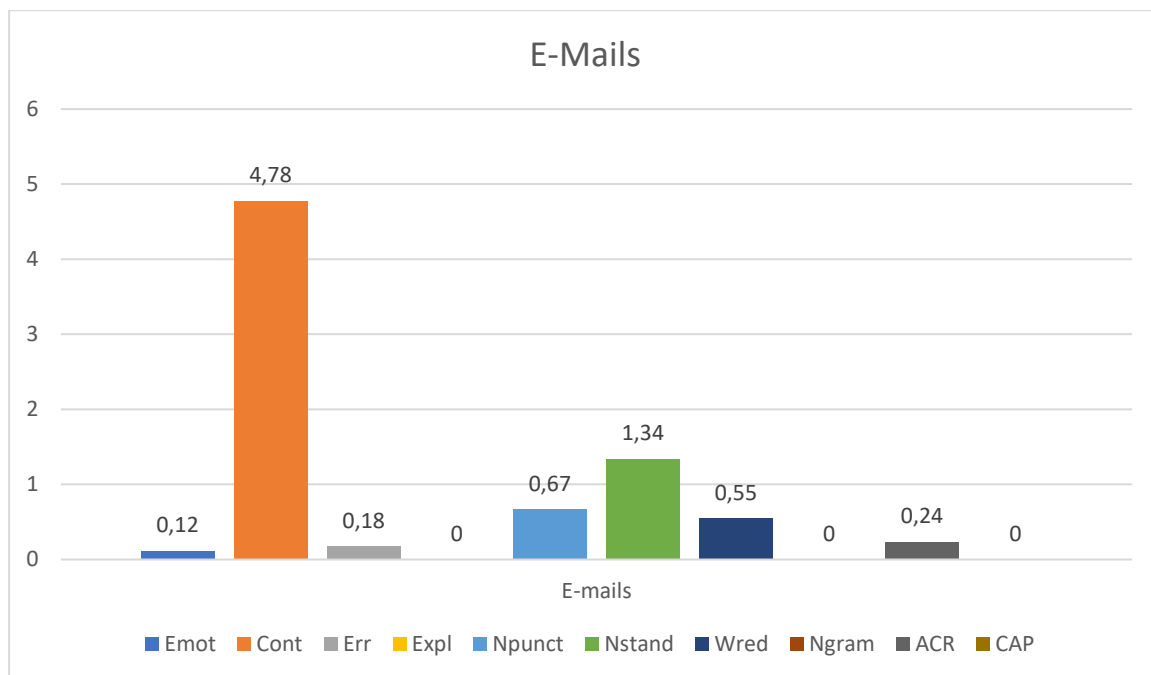

Figure 2: Richard's Features on the Musical Forum

Figure 3: Richard's Features on Reddit

Figure 4: Maria's Features on WhatsApp

Figure 5: Maria's Features in her E-mails

5.3   Subject Matter

Both participants state that the subject matter of an interaction is a factor for how and why their language varies. In Maria's case, this is evident in her WhatsApp data, where the majority of her formal messages are ones where the subject matter is serious, such as the mention of her sick pet, while her informal messages occur in less serious situations, like when she is asking her friends what their plans for the day are. In her interview, Maria claims that the subject matter affects the length of her messages on WhatsApp. She further explains that longer messages are more thought out and therefore include more standards, thus raising their level of formality. Shorter messages, on the other hand, are "quick-thought responses", and therefore have fewer standards, thus lowering their level of formality. These points are shown in the data as well, where the more formal WhatsApp messages are longer, occurring in situations where the subject matter is more serious. Thus, some of Maria's identities include ones that are expressed in reaction to the seriousness of the subject matter.

As for E-mails, Maria does not make specific mention in her interview of the subject matter affecting her in the same way as her WhatsApp messages, however the data consisted mostly of E-mails regarding formal documents and discussions of work and school. The few instances of E-mails that are addressed to close friends and family do not deviate very much from her other E-mails, despite their subject matters differing, with the exception of one of them including emojis. Regardless of length of the E-mail, the level of formality was consistently high. Thus, it seems as if the subject matter in E-mails is not as large of a factor for her usage of variation, or the subject matters in her E-mail data was more consistent. One of the reasons behind this could be that the format of E-mail itself, which Maria likens to letter-writing, is generally formal, and that this outweighs the effects of subject matter on level of formality. It is also possible that she does not employ the same habit of "quick-thought responses", and instead puts some amount of thought into all her E-mails, leading to higher levels of formality throughout.

46

Richard's data shows a similar pattern where the subject matter affects his enactment of various identities. This is true both for his Reddit and Forum excerpts, such as examples wherein he inhabits a knowledgeable identity when the topic is wrestling, in which he will use terms, acronyms, and references that show a level of familiarity with and knowledge of the topic. Richard does not, however, go through the same degree of variation of formality as Maria does. Whereas Maria might use far more non-standards in non-serious conversation, Richard largely keeps a more consistently standard level in his language. This can be attributed to the differences between each participant's respective platforms. Richard's variations mostly come in the form of technical terms or highly local abbreviations that relate to the topic, but he rarely uses non-standard grammaticality or punctuation. The reason for this might be similar to reasons for Maria's formality in her E-mails. Much like she aims to be consistently formal to a degree that the subject matter has little effect, it might also be the case that Richard endeavors to maintain a certain standard in his language, such that he avoids being excessively informal. This is supported by his interview data, in which he states that he is intentionally formal and proper when writing online, in an effort to be taken seriously, but without being "stuffy". The full tables in appendices 3-6 also show this, as the level of formality on Reddit and the Forum varies between excerpts, while WhatsApp and E-mails are more consistently low/high throughout.

5.4   Audience Design

Richard explains that he is consistently formal on both his platforms due to the fact that he considers both of them to be public. Maria's platforms, however, are private, in the sense that the only people intended to read her messages and E-mails are herself and her recipient(s), which she states allows her to be comfortable in her language. These points all relate to Bell's (1997) theory of audience design, in which shifts in language are made to be more alike that of the person they are talking to (p. 244). More generally, it refers to the 'to whom and where' of an interactional context affecting the performance of identities and usage of variation. Both participants state that this 'to whom and where' is a large factor for them in CMC. Richard's usage of specific acronyms is certainly an example of this, as it is clear he cannot use *LFR* the same way outside of certain communities, at the risk of not being understood.

In her WhatsApp messages, Maria's own explanation makes it clear that, while she is conscious of her audience, the subject matter is a larger factor for her modulating her overall formality. Maria also states that she will consciously make efforts to "dumb-down" her own writing in an effort to dissipate tension. Thus, she knowingly adapts her own language to intentionally modify the current tone of the interaction. In order to do so successfully, she is aware of who she is talking to, and thereby aware of the specific linguistic elements needed to shift the tone in a way that is effective towards this individual. This is shown in one of her excerpts in which she offers advice and comforts a third party by alternating usage of *you* and *u*, showing that the matter is serious but not overly so, through being supportive in her message while remaining relatively lighthearted through her variations.

Maria's usage of emoticons is also a clear expression of audience design, since she shows in her data and states in her interview that the usage of emoticons is a direct response to whether the other person has used them. This is clear in her WhatsApp data as well as her E-mails, in which the one E-mail that includes emojis does so as a direct result of the third party using them in the preceding E-mail. Maria's formality in E-mails is also explainable through audience design, since she describes the recipients of her E-mails as more socially distant individuals, which she regards as a factor for writing more formally.

5.5   Stance

Maria's usage of *ya'll* on both of her platforms is a good example of an expression of stance as defined by Jaffe (2009), in which stances are ways of indexing specific and shared cultural structures of feelings and norms that may be used as a way of drawing social boundaries that are core to social differentiation and categorization. Stances are thus usable by individuals in order to lay claim to their own particular identities and statuses, as well as evaluating those of others (p. 7). It is likely that Maria's usage of *ya'll* is primarily a stance indexing a cultural norm associated with a subset of speakers of American English and a way to claim an identity. She states that it is an integral part of her language, and since it is found across both platforms and in all levels of formality, it is likely a very core part Maria's expression of her identities.

Barton & Lee (2013) state that stance can be affective or epistemic, where the former signals the writer's feelings, while the latter is a way of signaling knowledge or belief towards a statement (pp. 86-7). Maria also expresses stance, both affective, epistemic, or a mix of both, through her use of punctuation marks. She states that she wishes to project excitement, which is primarily an affective stance, without sacrificing formality in her E-mails, and therefore chooses frequent but singular exclamation points. On WhatsApp, however, she may use several repetitions of the same exclamation point or question mark, noting that the number of repetitions, or lack thereof, add some meaning to them, such as *???* expressing shock (primarily affective) and *?* expressing a genuine questioning (primarily epistemic). Regarding her E-mails, she notes that a single exclamation point is a nice balance between the expressiveness and seriousness she wishes to convey, showing how she uses both affective and epistemic stances through slight variation of features.

Maria explains in her interview how her mood, which can be expressed through affective stances, is a major factor for her use of variation and projection of various identities. Generally, she claims, her usage of non-standard forms, such as *u* over the standard *you*, comes as a result of a more relaxed mood. As explained in section 5.3, shorter messages include fewer standards and are used during less serious topics, and longer messages include more standards and are used during more serious topics. There is a seeming link in the available data between a relaxed mood and more frequent and shorter messages. Maria is thus expressing stance not only through single variations in words, but through larger modifications to her method of communication as a whole, since she can express an affective stance of relaxedness by communicating in shorter messages, or a stance of seriousness by using longer one. This is also a way of aligning with others as described by Du Bois (2007, p. 163) by setting the tone of an interaction for both the sender and recipient(s) through intentional usage of message length.

Although both Maria and the recipient's level of awareness of such stance expressions is unclear, it is apparent that Maria uses very specific terms to signal her own feelings as well as beliefs towards her statements and recipients through such small variations.

Richard also expresses stance in his data, through his intentional choice of various identities in his Reddit posts. He is seemingly aware of his own tendency to enter characters or, as described by Chambers (2013), project attributes (pp. 327-8), such as 'intelligent' or 'curious', wherein he will make adjustments to his language. In doing so, he also enacts stances, both within the discourse as epistemic 'I know this' examples, but also at a more meta level, by enacting ideological stances towards the language used by specific people. He is explicit about this, stating that he associates certain idioms or phrases with a higher degree of authenticity, or as better signals of proficiency in the language. This point may also result from Richard being an L2 speaker of English, such that he wishes to enact a stance of proficiency in his second language. Overall, Richard often makes use of what Barton & Lee (2013) refer to as an "expert voice" (p. 91) by frequently using highly accurate and informative language, especially in posts where the subject matter is on one of his primary interests or hobbies. A good example of this is one of his posts where he discusses wrestling, wherein he projects an expert voice by using phrases, terms, names, and concepts that index a high degree of familiarity and knowledge with the subject matter. This too is an alignment with other subjects, in which Richard enacts an "expert voice" as a result of his wish to educate or clarify things to the third party. Whereas both participants enact numerous stances, Maria's are seemingly tied more to her mood and the subject matter, whereas Richard's are a result of the subject matter and his intention towards authenticity.

6 Conclusion

6.1 Summary

This paper set out to answer how written English varies within a selection of online spaces, and what motivates this variation. In order to achieve this, I collected, processed, and analyzed authentic texts from willing participants in order to identify a number of salient features used within them. The findings here were that Maria's language was very varied across her platforms. Her language in her E-mails was consistently formal with a high number of standards, whereas her language on the Instant Messaging app WhatsApp was overall informal with few standards, with some messages being at a middle level of formality. In Richard's case, both his platforms were mostly formal with several standards, with occasional excerpts from both platforms having a middle level of formality.

In order to gain insight on why these variations were chosen, interviews were conducted with both participants. Some primary insights from both participants were that they are conscious of the situation they are writing in, as well as to whom, and that they will accommodate their variation usage accordingly. Maria elaborated on her conscious methods for raising and lowering her formality through variations, as well as explaining that her usage of emoticons relies primarily on her recipient's degree of usage. Richard also discussed his own intentionally accurate and formal language as a way to seem authentic, as well as describing some of the smaller differences between Reddit and the Forum that affects smaller variations across the two.

In conclusion, written English online varies through a set of features that index a variety of meanings and communities. Participants deploy these features with varying degrees of awareness but are primarily motivated by the context the writing occurs in, their audience, and their mood. Furthermore, they are motivated by which of their identities they wish to express, and which affective and epistemic stances they wish to enact.

6.2 Suggestions for Further Research

Further research approaches have several avenues for expanding or challenging the findings presented here. One approach might be to expand the number of participants within a set of platforms, in order to explore whether the findings here are indicative of common trends, or deviations from a different pattern. Another approach might be to investigate multiple participants that are all part of the same platforms, to deepen the insight into the importance of micro-level factors such as context, mood, and audience. Finally, an interesting approach could be to investigate spoken English online through similar methods as ones used here, in order to explore whether findings are similar or wholly different when the data is spoken rather than written.

References

Akselberg, G. & Mæhlum, B. (2014). Sosiolingvistisk Metode. In Mæhlum, B., Akselberg, G., Røyneland, U., Sandøy, H. *Språkmøte. Innføring i Sosiolingvistikk.* (p. 73-87) Oslo: Cappelen Akademisk Forlag. (2. Ed.)

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), 419-438.

Androutsopoulos, J. (2010). Localising the global on the participatory web: Vernacular spectacles as local responses to global media flows. In N. Coupland (Ed.), *Handbook of Language and Globalization.* Oxford: Wiley-Blackwell

Barton, D. & Lee, C. (2013). *Language Online. Investigating Digital Texts and Practices.* (Routledge)

Bell, A. (1997). Language style as audience design. In Coupland, N. & Jaworski, A. (Eds), *Sociolinguistics.* Macmillan Education UK.

Brady, A. (2011). Opportunity Sampling. In Jupp, V. (ed.), *The SAGE Dictionary of Social Research Methods.* London: SAGE publications, Ltd. http://methods.sagepub.com/base/download/ReferenceEntry/the-sage-dictionary-of-social-research-methods/n137.xml

Bucholtz, M. & Hall, K. (2009). Locating Identity in Language. In Llamas, C. & Watts, D.J.L. (Eds.), *Language and Identities*. (Sociolinguistics).

Chambers, J. (2003). *Sociolinguistic theory: Linguistic variation and its social significance* (2nd ed., Vol. 22, Language in society). Oxford: Blackwell.

Crystal, D. (2001). *Language and the Internet.* Cambridge, U.K.: Cambridge University Press.

Dalen, M. (2004). *Intervju som Forskningsmetode. En Kvalitativ Tilnærming.* Oslo: Universitetsforlaget

Du Bois, J.W. (2007). The stance triangle. In Englebretson, R. (Ed.), *Stancetaking in Discourse* (pp. 139-182). Amsterdam/Philadelphia: John Benjamins Publishing Company

Eckert, P. (2000). *Linguistic variation as social practice: The linguistic construction of identity in Belten High* (Vol. 27, Language in society). Malden, Mass: Blackwell.

Eckert, P. (2004). The meaning of style. In *Texas Linguistic Forum,* 47, 41-53.

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, 41(41), 87-100.

Herring, S.C. (1996). *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives. Pragmatics and Beyond* series. Amsterdam: John Benjamins.

Herring, S.C. (2001). Computer-mediated discourse. In Schiffrin, D., Tannen, D., & Hamilton, H.E. (Eds.) *The Handbook of Discourse Analysis,* (pp. 612-634). Malden, MA: Blackwell Publishers.

Herring, S.C. (2004). Slouching toward the ordinary: Current trends in computer-mediated communication. *New Media & Society,* 6(1), 26-36.

Horvath, B. M. (2018). Ways of Observing. Studying the Interplay of Social and Linguistic Variation. In Mallinson, C., Childs, B., van Hek, G. (eds.) *Data Collection in Sociolinguistics.* (2018). Routledge. (2. ed)

Hårstad, S. (2010). *Unge språkbrukere i gammel by. En sosiolingvistisk studie av ungdoms talemål i Trondheim.* (PhD thesis) Norwegian University of Science and Technology, Trondheim.

Jaffe, A. (2009) Introduction: The Sociolinguistics of Stance. In Jaffe, A. (Ed.), *Stance: Sociolinguistic perspectives* (pp. 3-28) (Oxford studies in sociolinguistics). New York: Oxford University Press.

Jupp, V. (2011). Volunteer Sampling. In Jupp, V. (ed.), *The SAGE Dictionary of Social Research Methods.* London: SAGE publications, Ltd.

Kvale, S. & Brinkmann, S. (2009). *Det kvalitative forskningsintervju.* Oslo: Gyldendal Akademisk (2. ed)

Lüders, M. (2015). Researching social media: Confidentiality, anonymity and reconstructing online practices. In Fossheim, H. & Ingierd, H. (eds.), *Internet research ethics.* Oslo: Cappelen Damm Akademisk.

Mallinson, C. (2018). Research Design. In Mallinson, C., Childs, B., van Hek, G. (eds.) *Data Collection in Sociolinguistics.* (2018). Routledge. (2. ed)

McCulloch, G. (2019). *Because Internet: Understanding the New Rules of Language.* New York, NY: Riverhead Books.

NESH. (2019). *Forskningsetisk veileder for internettforskning*. Oslo: De nasjonale forskningsetiske komiteene. (2. ed)

Quist, P. (2009). Sociolingvistik i bevægelse. In Gudiksen, Hovmark, Quist, Scheuer & Sletten (Eds.). *Dialektforskning i 100 år*. Copenhagen: Nordisk Forskningsinstitut.

Schilling, N. (2013). Investigating Stylistic Variation. In Chambers, J & Schilling, N. (Eds.), *The Handbook of Language Variation and Change* (pp. 325-349). Oxford, UK: John Wiley & Sons.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication,* 23, 193–229

Thurlow, C., & Mroczek, K. (2011). Introduction. In Thurlow, C., & Mroczek, K. (Eds.), *Digital Discourse: Language in the New Media.* Oxford University Press.

Unicode.org (2019). *FAQ.* Retrieved May 25 2020 from https://home.unicode.org/basic-info/faq/

Utdanningsdirektoratet. (2019). *Læreplan i Engelsk* (ENG01-04). Retrieved May 28, 2020, from https://data.udir.no/kl06/v201906/laereplaner-lk20/ENG01-04.pdf

Wetzel, L. (2018). Types and Tokens. In Zalta, E.N. (Eds.), *The Stanford Encyclopedia of Philosophy.* Retrieved May 28, 2020 from https://plato.stanford.edu/archives/fall2018/entries/types-tokens/

Widerberg, K. (2011). Historien om et kvalitativt forskningsprosjekt. Oslo: Universitetsforlaget (3. ed).

Appendices

Appendix 1: Letter of Information and Consent

Appendix 2: General Questions Included in Interviews for Both Participants

Appendix 3: Maria's Table for E-mails

Appendix 4: Maria's Table for WhatsApp

Appendix 5: Richard's Table for Reddit

Appendix 6: Richard's Table for the Musical Forum

Appendix 7: Relevance for the Teaching Profession

Would you like to participate in the research project

"Variation in English within Online Space"?

This is a query for you regarding whether you would like to participate in a research project where the purpose is to investigate variation in the usage of English in digital arenas and possible reasons for this variation. In this document you will be provided with information regarding the goals for the project and what it would entail for you as a participant.

Purpose

The purpose of the project is, as mentioned, to explore variation in English in different areas on the internet, in three cases comprising both native speakers and non-native speakers. This entails a look into how written English is produced by each participant on a selection of arenas (websites/applications) they agree to share from. To gain further insight, the project is interested in relevant background information on participants, to map which factors might be contributing to shaping such variation found on digital arenas. The project's thesis is thus formulated as seeking to explore how variations within English play out in written form online, and with which behind-the-scenes factors that drive them, in light of three "case"-studies. The project is a Master's Thesis, conducted as part of a 5-year programme for Teacher's Education in English at the Norwegian University of Science and Technology – NTNU.

Who is responsible for the research project?

Norwegian University of Science and Technology – NTNU, The Department of Language and Literature, supervisor Annjo Klungervik Greenall, and student Isak Enger are responsible for the project.

Why are you being asked to participate?

You are being asked to participate as you are highly proficient in English, are a native speaker or non-native speaker, and are, or have recently been, an active user of the language on digital arenas. If you consent, you will be one of two to three participants in the project.

What would my participation involve?

If you choose to participate in the project, it would involve you providing a selection of self-produced text in English from a variety of websites/applications as data for the project. If you consent to participating in the project, a short interview will be conducted to get an overview of which websites and applications you most frequently use, and that

will thereby be sources of data for the project. Data you send will be short and long excerpts of text in the form of screenshots of conversations, posts on forums, etc. If you feel that some text excerpts are too private and would not like to share them with the project, you are free to exclude and replace them with other relevant excerpts. Where context is necessary in order to fully understand the excerpts of text, you are asked to briefly provide this as part of the data. In cases where excerpts include a response from a non-participating person, you are asked to completely anonymize any identifying or private information regarding this third party. This includes anything that might reveal their name, face, location, age, residence, etc. You are also to anonymize any sensitive information that the third party shares in the conversation, such as information regarding criminal activity, illness, or any other information that might be harmful to the third party. If you are in doubt regarding whether or not something needs to be expunged, you may ask the project coordinators without revealing the information itself. If you 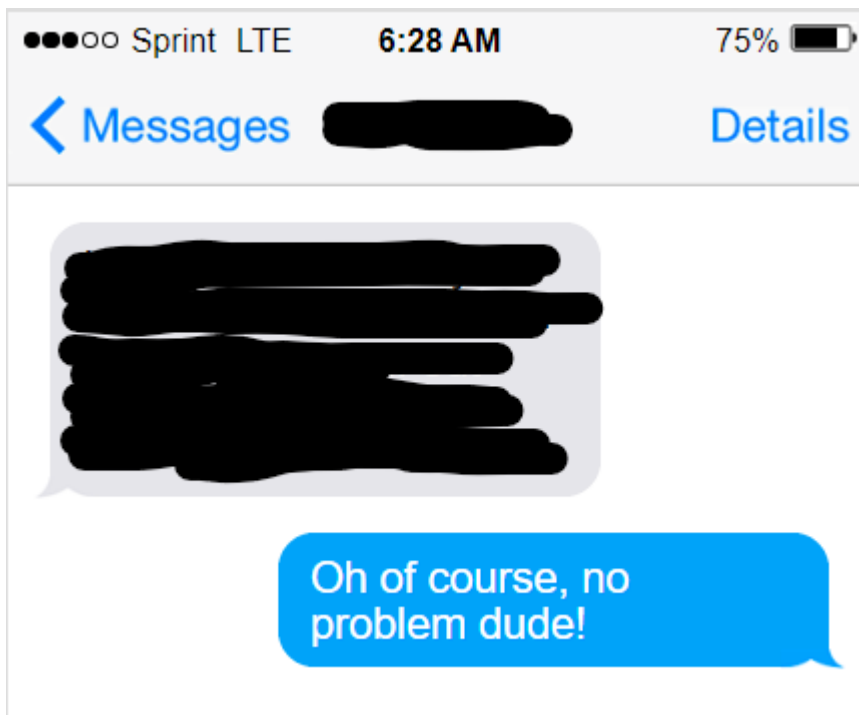are unable to do so, or are still unsure, we ask that you do not send us such an excerpt. Anonymizing information needs to be done such that none of the aforementioned details are revealed, but need not be a cumbersome process. Simply using digital software to draw a large line across the information is sufficient. In some situations the removal of information may obscure the overall context. We ask that, in such cases, you briefly reintroduce the context without revealing the sensitive information. For example, if the data you send looks like this:

It should be sent as such:

Context: Third party was unable to bring their textbook to class, and asked if the participant could share hers with them.

Data will consist of text produced after 2015. You will be asked to send data totalling 1500-2000 words per digital platform, meaning, for instance, 1500 words from a forum and 2000 words from WhatsApp chats, totalling 3500 words per participant. It is important to note that the word-count counts only for text you yourself have produced. Replies, added context, notes, or other text not part of the original text excerpt do not count towards the total word-count. You are not expected to provide this amount of data immediately after agreeing to participate. You are expected to provide the total 3-4000 words over a duration of time, so that you are given ample opportunity to find and select which excerpts you would like to send. However, you are asked to have provided the full amount of data by the end of February 2020 so as to give the project enough time to read through and analyse it.

If you choose to participate in the project, you will also be asked to take part in an interview designed to ask you to reflect on your own thoughts and attitudes towards usage of English online. These interviews will be semi-structured, meaning the questions will be prepared beforehand but there is room for follow-up questions. They will not ask any questions that go outside the areas described in the "Your privacy" section below. Interviews will be done via e-mail. You will be sent a list of questions that you will be asked to answer as extensively and as openly as possible. Based on your response you may receive a follow-up e-mail asking you to elaborate on some points or statements. The entirety of your response(s) within the e-mail(s) will be used as data for the project. Thereby, if you have any questions surrounding the project, its design, the questions you are being asked etc., and would like to have these questions be excluded from your data, please send them separately from the response to the e-mail interview.

Participation is voluntary

Participation in the project is voluntary. If you choose to participate, you can at any time withdraw consent without providing any reason. All information gathered about you will

then be disposed of. There will be no negative repercussions for you if you choose not to participate or withdraw from the project at a later date.

Your privacy – How we store and use your information
We will only use information on you for the purposes described in this document. We will process information confidentially and in accordance with privacy law.
- Only the student and supervisor will have access to your information.
- In order to secure your information your name and information will be assigned a code that will be stored separately from other data. Data and information will be stored on dedicated storage units that will be stored separately and locked away when not in use.

The project will not make you identifiable. All data on you will be stored safely, and sensitive information will be anonymized in the text itself. Sensitive information entails information that might be revealing or that might make you identifiable, such as your name, your place of residence within a country, or names of your friends or family. For instance, if the name of your address is included in one of your excerpts in a manner like this:

Participant: Sure, you can pick me up from Edvard Bulls Veg 1!

It should be rewritten in the following way:

Participant: Sure, you can pick me up from [Address]!

Information about you that will be part of the project includes excerpts from text you have written (your name and username(s) will be anonymized), information on your country of residence, countries you have previously resided in, level of education, social background, areas of interest, political affiliation, and general age.

What happens to your data when the research concludes?

The project is expected to conclude by June 30th, 2020. At the conclusion of the project, data and information you have provided, as well as any records of interviews will be deleted and destroyed.

Your rights

As long as you are identifiable by the data, you have the right to:

- Insight into what information about you has been registered
- To correct information about you
- To delete information about you
- To be provided with a copy of information about you (data portability), and
- To file a complaint to the Data Protection Officer or The Norwegian Centre for Research Data regarding the processing of your information

What gives us the right to process information about you?
We process your information based on your consent.

Per the request of the Norwegian University of Science and Technology – NTNU, The Department of Language and Literature – The Norwegian Centre for Research Data AS has concluded that the processing of personal information in this project is done in accordance with privacy law.

Where can I find out more?
If you have questions regarding the study, or wish to invoke your rights, please get in touch with:

- Norwegian University of Science and Technology – NTNU, The Department of Language and Literature through student:
  Isak Enger
  Phone: +47 40496094
  E-mail: isake@ntnu.no

  Or the project supervisor:
  Annjo K. Greenall
  Phone: +47 73596790
  E-mail: annjo.k.greenall@ntnu.no

- Our Data Protection Officer
  Thomas Helgesen
  Phone: +47 93079038
  E-mail: Thomas.helgesen@ntnu.no

- NSD - The Norwegian Centre for Research Data AS, via e-mail
  (personverntjenester@nsd.no) or by phone: 55 58 21 17

Kind regards

Annjo Klungervik Greenall   Isak Enger

(Researcher/supervisor.)   (Student)

Declaration of consent

I have received and understand the information regarding the project *Variation in English within Online Space* and have had the opportunity to ask questions. I consent to:

- Participating in data collection in the form of personal information and excerpts from self-produced texts
- Participating in interviews regarding my own views and attitudes towards English online.

I consent to having my data processed until the conclusion of the project, ca. June 30th. 2020

(Signed by participant, date)

Appendix 2: General Questions Included in Interviews for Both Participants

Introductory questions – Please include as much "necessary" information as possible.

- What is your age?
- In what country do you currently reside?
    - Where have you previously resided? (please state duration to the best of your ability)
- What is your highest level of education?
- Describe your social background (how you grew up, social life, current social status, living situation, etc.)
- What are your primary areas of interest? (Such as art, movies, literature etc.)
    - Briefly describe what attracts you to these interests
- What political affiliation would you say you adhere to?
- To what degree, if any, have your habits with using written English, changed over the past 4-5 years?

Attitudes towards language online

- Give a short overview of a typical day for you online, in cases where you exclusively use written English to communicate with other users.

- On which websites/applications are you most active? Which were you most active on over the past 5 years? Why these in particular?

- How would you describe the various ways people use English on online platforms where you are a regular participant/observer? ("observer" meaning you read other posts or messages without participating in them yourself)?

- How would you, in your own words, describe the way you use language when you write online?

- Do you consider yourself a frequent user of emoticons or emoji when writing online? Why do you use fewer/more/about the same amount as your peers?

- Would you consider your overall writing-style in English online as formal or informal? Why?

- To what degree is your level of formality affected by the website/application and you are on, as well as the ongoing topic/context?

- To what degree is your writing a reflection of the way you speak English?

- Do you feel that your gender affects the way you write online?

- How would you describe the difference(s) in your own writing when comparing [Platform 1] and [Platform 2]?

- Why do you think you write this way on [Platform 1]?

- Why do you think you write this way on [Platform 2]?

Appendix 3: Maria's Table for E-mails

The topmost row denotes the features counted, as described in section 4.1.

The leftmost column denotes which excerpt the features are counted from, for instance, 'E-M1' refers to 'E-mail excerpt number 1', and 'FO8' refers to 'Forum excerpt number 8'. The numbers underneath each excerpt denotes the amount of words contained within the excerpt, such that E-M1 contains 86 words and FO8 contains 246 words.

The second to bottom row is a totally tally, where the leftmost column shows the combined word total for all excerpts from each platform.

Cells at each intersection of a row and column contain the number of tokens of each feature, with the exception of the 'Formal' column, which shows the formality rating given to that excerpt.

The bottom row denotes what percentage each of the features were of the total amount of words.

E-Mails	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	CAP	Formal
E-M1 86	0	7	0	0	0	1	0	0	0	0	H
E-M2 26	0	0	0	0	1	0	0	0	0	0	H
E-M3 105	0	3	0	0	0	0	0	0	2	0	H
E-M4 57	0	7	0	0	2	3	2	0	0	0	M
E-M5 52	0	4	1	0	0	1	0	0	0	0	H
E-M6 44	0	2	0	0	0	3	0	0	0	0	H
E-M7 200	0	11	0	0	1	4	2	0	0	0	H
E-M8 125	0	7	1	0	2	0	0	0	0	0	H
E-M9 44	0	2	0	0	0	1	0	0	0	0	H
E-M10 112	0	8	0	0	1	2	1	0	0	0	H
E-M11 126	0	6	1	0	0	0	0	0	0	0	H
E-M12 24	0	1	0	0	0	1	1	0	0	0	H
E-M13 113	0	3	0	0	2	0	0	0	0	0	M
E-M14 116	0	6	0	0	1	0	2	0	1	0	M

E-M15 125	0	5	0	0	0	2	0	0	0	0	H
E-M16 134	0	7	0	0	0	3	0	0	0	0	H
E-M17 60	2	5	0	0	1	1	1	0	1	0	M
E-M18 82	0	4	0	0	0	0	0	0	0	0	H
Total 1631	2	78	3	0	11	22	9	0	4	0	H
% of total WC	0.12	4.78	0.18	0	0.67	1.34	0.55	0	0.24	0	

Appendix 4: Maria's Table for WhatsApp

WhatsApp	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	CAP	Formal
WA1 26	0	2	0	1	0	5	2	2	0	1	L
WA2 43	0	1	0	0	2	7	6	0	0	0	L
WA3 36	0	3	0	0	2	7	4	1	2	2	L
WA4 29	0	1	0	0	3	3	1	1	0	5	L
WA5 36	0	1	0	0	0	10	6	2	1	1	L
WA6 118	0	5	0	0	2	5	5	1	0	2	L
WA7 28	0	0	1	1	0	5	1	0	1	6	L
WA8 33	0	1	1	0	2	6	4	0	0	1	L
WA9 47	0	2	0	0	0	1	0	0	0	0	M
WA10 65	1	0	0	0	0	3	3	0	1	1	L
WA11 51	1	3	0	0	1	2	1	0	1	1	L
WA12 78	1	4	0	0	2	2	1	0	0	1	L
WA13 53	2	2	0	0	3	3	1	0	0	1	M
WA14 43	1	2	1	2	1	0	0	1	0	6	L
WA15 55	1	2	0	1	2	6	5	2	1	3	L
WA16 29	0	2	0	0	1	3	2	1	0	0	L
WA17 69	0	1	0	0	0	2	0	4	1	0	L
WA18 48	1	0	0	1	0	0	0	2	0	1	L
WA19 56	0	2	0	1	1	3	2	2	0	0	L
WA20 55	1	1	1	0	1	1	1	2	0	1	L
WA21 49	1	2	0	0	3	3	3	1	0	0	L
WA22 54	5	1	0	2	0	5	3	1	3	4	L
WA23 35	0	1	0	2	1	2	2	1	0	12	L

WA24 58	1	3	0	0	0	1	1	3	1	0	L
WA25 50	3	4	1	0	1	2	2	0	0	1	L
WA26 52	2	1	0	0	1	1	1	1	0	1	L
WA27 50	0	2	1	1	0	2	0	2	0	0	L
WA28 86	0	1	0	2	1	2	1	1	0	0	M
WA29 82	0	1	0	2	1	2	2	2	0	8	L
WA30 66	0	0	0	3	3	0	0	3	0	2	L
Total 1580	21	51	5	18	34	94	60	36	12	61	L
% of total WC	1.32	3.22	0.31	1.13	2.15	5.94	3.79	2.27	0.75	3.86	

Appendix 5: Richard's Table for Reddit

Reddit	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	Cap	Formal
Red1 238	0	12	0	0	0	12	0	0	0	0	H
Red2 303	1	7	1	3	0	4	0	0	2	0	M
Red3 226	0	5	1	0	0	11	1	0	9	0	H
Red4 191	0	5	1	1	0	0	0	0	0	2	M
Red5 309	0	9	1	0	0	0	2	0	0	0	H
Red6 388	0	8	0	1	1	0	0	0	3	0	H
Red7 134	1	0	0	0	1	1	0	0	1	1	H
Red8 64	0	1	0	0	0	1	0	0	0	0	H
Red9 47	0	1	0	1	0	1	0	0	0	1	M
Red10 203	0	8	0	2	0	0	0	0	0	0	H
Total 2103	2	62	4	8	2	30	3	0	15	4	H
% of total WC	0.09	2.94	0.19	0.38	0.09	1.42	0.14	0	0.71	0.19	

Appendix 6: Richard's Table for the Musical Forum

Forum	Emot	Cont	Err	Expl	Npunct	Nstand	Wred	Ngram	ACR	CAP	Formal
FO1 156	1	7	0	0	0	0	1	0	1	0	H
FO2 281	1	12	0	1	1	2	1	0	1	0	M
FO3 146	0	4	0	0	0	0	0	0	0	0	H
FO5 178	0	6	0	2	0	0	0	0	3	0	H
FO6 139	0	5	0	2	0	0	0	0	3	0	M
FO7 23	0	1	0	0	0	1	0	0	0	0	H
FO8 246	0	6	0	1	0	0	2	0	0	0	M
FO9 160	0	5	1	0	0	0	4	0	2	0	H
FO10 189	0	5	0	0	1	0	2	0	0	1	M
FO11 288	0	8	1	0	1	1	0	0	1	0	H
FO12 225	1	8	1	0	0	1	1	0	2	0	H
Total	3	67	3	5	2	5	11	0	10	1	H

FO4 is absent, as a result of it being particularly long excerpt that was cut so as to not exceed the word count per participant.

Appendix 7: Relevance for the Teaching Profession

The Internet's presence in the lives of teens and young adults today hardly needs introduction. They spend a large amount of time online, and with the language of many parts of the Internet being English, they are exposed to staggering amounts of language that varies in several aspects. Through this research, I have looked into variation through a specifically online lens, identifying what motivates such variation and how it may take shape. While restricted to the two cases discussed, the insights gained can serve as valuable knowledge to bring into the classroom, both as a teacher and as a tool for students.

By 10th grade, and by the end of Upper Secondary, students of Norwegian schooling are expected to use digital resources in their language learning, creation of texts, and interactions (Utdanningsdirektoratet, 2019, pp. 8-12). The Norwegian Directorate for Education and Training (2019) also states specifically that language learning is to occur in meetings with English language texts, where texts are understood as both oral and literate, and through various means, including the digital (p. 3). Thus, awareness of the many forms English can take online, as well as reflections around what motivates such variation, are important tools for how a teacher can approach this topic within the classroom. It allows for the teacher to carry these aspects of online language into the structuring of discussions and tasks that intend to explore this facet of English language use. Bringing up some of the points discussed here, such as the importance of context, and audience, are important for students to be aware of in their own reading and discussion of English online, both within and outside school-related contexts. Likewise, discussions of formality as dependent on context is an important point to relay to students. In teaching them to become proficient users of English, they benefit from increased awareness surrounding the fact that different situations may call for different language.

It can also be helpful for them more generally see the many various forms English inhabits, and be made aware that they are all equally authentic examples of language. English classrooms tend to be focused on a certain 'correctness' of the language. From the perspective of teaching of grammar and syntax, this is clearly with good reason. However, real life language, especially on the Internet, is not as concerned with proper grammar and syntax as some may be led to believe. Thus, seeing, exploring, and discussing language that falls outside normative examples often employed can spur more creative usage and reflection from students on their own language use. Knowledge of indexicality, for instance, may be an impetus for discussions on ideological associations within language.

Finally, the process of writing this paper has helped solidify, in my mind, the value of good and thorough feedback over longer writing processes. This is something I will undoubtedly carry with me into my profession, and I will strive to provide all my students with the same quality of feedback as I have received from my supervisors and peers through the creation of this research.