

University of Nottingham
UK | CHINA | MALAYSIA

#GlockeAktiv: Gender and Ethnicity Differences in German-language YouTube Comments

Supervisors: Professor Nicola McLelland, Dr Olivia Walsh

LOUIS ALEXANDER COTGROVE, BA (HONS)

Submitted to the University of Nottingham in partial requirement for the degree of Master of Arts by Research

September 2017

Student ID: **4280155**

Word Count: **19,998**

Table of Contents

1 Introduction 4

2 Research Context: Gender, Ethnicity, and Youth Online Language 7

2.1 The Relationship between Spoken and Written Language Online 7

2.2 Gender Performance in Spoken Language 9

2.2.1 Gender and Discursive Differences in Spoken Communication 11

2.2.2 Gender and Lexical and Syntactical Differences in Spoken Communication. 13

2.3 Gender Performance Online 16

2.3.1 Constructing and Identifying Gender Online 16

2.3.2 Discursive Differences in Online Communication 18

2.3.3 Gender and Lexical and Syntactical Differences in Online Communication .. 20

2.4 Youth Language Online and Offline 9

2.5 Migrant Background, Gender and the German Language 24

2.5.1 Migrant Background and Identity 25

2.5.2 Migrant Background and Gender 28

2.6 Research Gaps and Questions 30

3 Methodologies 32

3.1 Compiling the Corpus 32

3.1.1 General Parameters 34

3.1.2 Parameters for the Listicle Lifestyle Videos 35

3.1.3 Parameters for the Music Videos 35

3.1.4 Assigning Gender and Ethnicity to Commenters 37

3.1.5 Analysing the Parts of a Comment: Comment Construction Units 43

3.1.6 Removing Outliers from the Data Set 45

3.2 Linguistic Features to be Analysed 47

3.3 Statistical Methods and Analysis of the Features 52

4 Results and Discussion 55

4.1 General Differences across Gender 55

4.1.1 Comments and Comment Length and Gender (Hypotheses A-C) 55

4.1.2 Positive or Adversarial Nature of Comments & Profanity across Gender (Hypotheses D-F) 57

4.1.3 Conversation Length and Interaction across Gender (Hypotheses G and H) 58

4.1.4 Females use more linguistic structures that weaken the assertion in speech (Hypothesis I) 59

4.1.5 Males use more contractions in language than females (Hypothesis J) 60

4.1.6 Orthographical and Lexical Features of Online Communication across Gender (Hypotheses K-N) 61

4.2 The Effect of Ethnicity on Language Differences across Gender 66

4.2.1 Language Mixing (Hypothesis O) 67

4.2.2 Comments and Comment Length across Ethnicity (Hypotheses A and B revisited for intersection with ethnicity) 67

4.2.3 Positive or Adversarial Nature of Comments & Profanity across Ethnicity (Hypotheses D-F revisited for intersectional analysis) 69

4.2.4 Females use linguistic structures that weaken the assertion in speech across Ethnicity (Hypothesis I revisited for intersectional analysis) 70

4.2.5 Contractions and Abbreviations across Ethnicity (Hypothesis J revisited for intersectional analysis) 71

4.2.6 Orthographical Features of Online Communication across Ethnicity (Hypotheses K-N revisited for intersectional analysis) 72

4.3 Intersectionality 75

5 Conclusions 78

6 Appendix 84

7 Bibliography 84

1 Introduction

This study examines discursive and linguistic differences in the online communication of German-speaking young people based on a corpus of comments made underneath recent YouTube videos that are primarily focused on the 13-17 and 18-24 demographics[1] (henceforth: the YouTube Young Person Corpus, or YYPC). The study examines language differences that occur based on the gender and ethnicity of the commenter, specifically in an online setting. A whole generation in Germany and Austria has now reached adulthood exposed to online communication, following the invention of the World Wide Web in 1991, and owns personal devices, such as mobile phones, to access online sources and communicate online.[2] Germany is also becoming more diverse: 22% of the population had a migration background in 2016 (Statistisches Bundesamt 2016), up from 15% in 1991 (Bundesamt für Migration und Flüchtlinge 2009).

The interaction between computer-mediated communication (CMC), gender, ethnicity and youth language is explored for the first time in this study. Each of the four areas has its own well-established body of research, but the interaction between them has not been widely studied. There is much research on gender performance online, but little accounts for how both a person's gender and

[1] According to YouTube data and media profile of the YouTuber (i.e. features in teen magazines, such as BRAVO).

[2] 85.5% of 14-19 year-olds owned a smartphone in Germany in 2016, of whom roughly 92% used them to access the internet (BurdaForward 2016).

ethnicity, often referred to as intersectionality,[3] affect online language use. Some work has been done on how gender or ethnicity affects youth language, but none on online language. In YouTube comments, these areas coincide and this study is the first to investigate all of these factors together. By reviewing previous research on these areas, I have identified linguistic features that have been suggested to show differences between males and females,[4] and between those with a non-migration (i.e. those with German/Austrian heritage) background and those with a migration background. By using the newer data of the YYPC, collected between April 2016 and May 2017, this study updates and re-evaluates previous work, and suggests how a "typical" comment by members of the four subsets (e.g. a male or female with either a migration or a non-migration background) might look. For example, the study presents evidence that comments from females with a migration background most closely portray standard language.

The study is structured in five chapters. Following this introduction, Chapter 2 reviews previous work on sociolinguistic variation in German concerning the linguistic features and practices that may vary according to gender and ethnicity, the relationship between CMC and spoken language, and the specificities of youth language. Chapter 3 describes the research methodology, including corpus creation, data analysis, and the parameters to assign gender and ethnicity to the

[3] Intersectionality can also refer to other social factors, such as age or sexuality, but commonly refers to the combination of gender and race (cf. Crenshaw 1989).

[4] It is important to acknowledge a gender (and sexuality) spectrum and the non-binary nature of biological sex (such as intersex). However, nearly all past research assumes a binary paradigm of two genders as well as an ethnicity binary between a non-migration and a migration background. In this study, I refer to commenters as male or female to avoid confusion when talking about young people, who have not reached adulthood, but I am not equating gender and sex.

commenters. It also explains which features and methods were used to analyse the data, including the choice of statistical tests. The findings of the investigation are presented in Chapter 4, where I examine to what extent linguistic differences across gender and ethnicity identified in previous research are supported by the YYPC, and explore intersectional online linguistic differences, such as between females of different migration backgrounds, to contribute to this fledgling research area. The essay concludes in Chapter 5, summarising how online language production differs across gender, ethnicity, and the intersections within those two areas. The section also suggests areas for further research, such as exploring the intersection of the previous social factors with sexual orientation.

2 Research Context: Gender, Ethnicity, and Youth Online Language

Previous research has demonstrated the use of language as a gatekeeper to maintain social hierarchies in both an online and "real life" context, presented as a series of dichotomies – males and females, non-migrant and migrant, and old and young, where in each case the former group is the holder of power. In this section, I review the research on these areas, focusing on publications relating to a German-language context to avoid ascribing English-speaking phenomena to German, and examine the publications that explore where some of these concepts converge. In doing so, I identify the linguistic features that, it has been suggested, may reflect the differences in how these groups communicate, yielding the research questions that guide the investigation in this study.

I first outline why it is important to consider linguistic differences identified in both spoken and online research. I then investigate language differences between genders, before exploring research on youth language. Following that, I examine research on linguistic differences based on migration background, as well as the smaller body of research on intersectionality, particularly that of doubly-minoritised groups, such as females with a migration background. I conclude with an outline of some of the research gaps that form the investigation of this study.

2.1 The Relationship between Spoken and Written Language Online

It is important to recognise the parallels between spoken and online written language as there are many features from spoken language that are replicated in an online written medium, including language economy (cf. Storrer 2001; Siever 2006)

and an informal, oral writing style, as promoted by moderators of online forums (cf. Crystal 2006: 24-25). The representation of spoken language features using the written (or typed) word has been called *konzeptionelle Mündlichkeit* ('Conceptual orality'). The term was popularised following the work of Koch & Österreicher (1985), who adapted an investigation by Söll & Hausmann (1980) on the differences between written and spoken French, to form their *Nähe-Distanz Modell* ('Proximity-Distance Model'), which expressed the interplay between linguistic features (the 'conception') of spoken and written language and the medium in which they are produced (i.e. oral or graphical). Table 1 shows how a sentence or utterance might appear depending on the conception and medium, adapted from Kilian (2010: 69). Kilian (2010: 81-83) also identified several linguistic features of online language that were adapted from spoken language, such as the use of tag questions, prosodic spelling (i.e. the repetition of graphemes or punctuation to represent intonation and emphasis), and the use of emoji/emoticons to depict non-verbal communication or emotion. These features have all been investigated as markers of gender difference in online language, as shown in section 2.3.1. It is therefore important to study both spoken and online language research on gender and ethnicity language differences, as features from the former research area often appear in the latter.

		Conception	
		Spoken	**Written**
Medium	**Graphic (example)**	*nehm wa mal an* (e.g. in an E-Mail) **[1]**	*nehmen wir einmal an* (e.g. in an essay) **[2]**
	Phonetic (example)	[ne:mwɐma'an] (e.g. in a casual conversation between friends) **[3]**	/ne:mən‿wɪɐ‿ainmal‿an/ (e.g. at an academic colloquium) **[4]**

Table 1 Conceptual orality and literality examples. Adapted and translated from Kilian (2010: 69).

2.2 Gender Performance in Spoken Language

Research on the interaction between social gender and language production first gained traction in the 1970s in the Anglo-American sphere, with the rise of second-wave feminism, catalysed by Betty Friedan's *The Feminine Mystique* in 1963. Academics such as West & Zimmerman (1977) and Lakoff (1975) hypothesised that the relationship between the gatekeepers of social power and the English language is used to supress women from an early age. They claimed that this led to the development of "women's language" (Lakoff 1975: 47-48), which discourages the expression of opinion and the self for women, whereas men are allowed to freely express themselves linguistically. The "ultimate effect" of this, Lakoff claimed, is a catch-22 for women, as they "are systematically denied access to power, on the grounds that they are not capable of holding it as demonstrated by their linguistic behaviour" (Lakoff 1975: 48). The idea of "women's language" has since been criticised by later researchers for being essentialist and ignoring the "heterogeneity of [a] group women even within a single society, let alone worldwide" (Cameron 2005: 17), but it is still pertinent to the idea that differences in access to institutionalised power structures can lead to language differences between social groups. Such a position is held by Eckert & McConnell-Ginet (1992: 483), who write

that by privileging the language of a certain social group, their linguistic practices are established as societal norms, "toward which everyone else orients". At the same time, they argue, it is imperative not to "abstract[…] gender from other aspects of social identity" when investigating language differences, as this "leads to premature generalisation even about 'normative' conceptions of femaleness and maleness" (Eckert and McConnell-Ginet 1992: 471).

In the early research on the relationship between gender and the German language by Trömel-Plötz (1979, 1984b) and Pusch (1984), they claimed, in line with Lakoff, that German was androcentric, although it assumed other forms particular to the grammatical gender structure in German. Pusch (1984: 48-65) provides several examples, such as the use of the masculine inflected pronoun to refer to 'everybody', i.e. *jeder kennt das*, and the use of masculine plural forms to cover both genders, such as *Studenten*, rather than the feminine-only form of *Studentinnen*. The use of masculine forms of language to generalise, Pusch argues, alienates and subjugates the female population, and must be reformed or else will lead to the "*sprachliche Vernichtung der Frau*" ('linguistic extermination of woman') (Pusch 1984: 10-11). Whilst most German gender theorists agree with Trömel-Plötz and Pusch that the establishment of the masculine as the 'normal' in the German language side-lines women in society (cf. Schoenthal 1989; Braun and others 1998; Klann-Delius 2005), there is considerable disagreement as to how men and women may express themselves differently through language, or if indeed a genderlect exists. The various discursive (e.g. turn-length, turn-taking, and interruptions), lexical, syntactical features that previous research has claimed differentiate

between men and women's language in "offline" contexts will be explored in sections 2.2.1 and 2.2.2.

2.2.1 Gender and Discursive Differences in Spoken Communication

Trömel-Plötz (1979), reinforced by her later work (Trömel-Plötz 1984a) and by Hummel (1984, in Trömel-Plötz 1984b), investigated mixed-gender communication on mainstream television mixed-gender news panel discussions and reached four conclusions. Firstly, men talk for longer than women – Trömel-Plötz (1984a: 59) reported that men spoke for twice as long as women and Hummel (1984: 285) found that men spoke for 50% longer than women. Secondly, women faced more difficulty in having their voice heard in a conversation – in a larger conversational group, Trömel-Plötz (1984a: 60) found that men would hand the next conversational turn to other men. Thirdly, if a woman were speaking, men would be more likely to interrupt her, whilst the reverse was rare. Fourthly, men more often decided the topic of conversation, and tended to have the last word before seeking to change the subject. For Trömel-Plötz, gender is the single most important factor when examining conversational discourse ("Geschlecht bestimmt mehr als andere Faktoren unser Gesprächsverhalten") (Trömel-Plötz 1984b: 15). However, contemporaneous research by Wodak (1981) on a different conversational context did not corroborate all the above findings. In a study of transcripts of therapeutic discourse in Vienna, Wodak did not find that men spoke for longer than women, although she did find that men interrupted women more frequently in conversation. Wodak's (1981: 272-275) main conclusion was that men and women used differing linguistic and conversational strategies to describe similar problems. For example, when discussing an issue at work, women spoke about their feelings

regarding it and their problems of inferiority or identity, and often created a narrative by reproducing dialogue that happened between them and their boss, for instance. Men, on the other hand, did not use much interpersonal language, instead listing the pressures at work and then their symptoms (Wodak 1981: 273).

The strongest opposition to Trömel-Plötz and Hummel's hypotheses on discourse came from Hellinger & Bussmann (2003a: 162), who stressed the ultimate primacy over gender of other socio-economic factors such as age, social class, hierarchy, and membership of a community of practice in turn-allocation and speaking time.[5] This is broadly in keeping with Eckert & McConnell-Ginet's (1992: 471) invocation to use gender as a lens through which to examine "other social attributes such as class, ethnicity, or age", however, Hellinger & Bussmann's approach was not based on their own empirical research, which undermines their claims. Instead, they performed a meta-analysis of three investigations by Wodak (1981), Gräßel (1991), and Thimm (1990). They claimed for each of the three investigations that 'no clear distribution of these strategies [i.e. Trömel-Plötz and Hummel's four hypotheses] could be established' (Hellinger & Bussmann 2003a: 162), despite Gräßel (1991) clearly stating "Unterschiede im Sprachverhalten von Männern und Frauen sind vorhanden, und zwar in einer so großen Zahl, dass sie nicht übergangen werden können" ('Differences in the language behaviour of men and women exist, and indeed in such a large number, that they cannot be

[5] A community of practice, according to Meyerhoff (2002: 527-28), is a group of people who meet and talk about a particular purpose, such as a bridge club or local council. Often new linguistic repertoires develop out of these communities as a result of the extended interaction between the members, such as an inside joke surrounding the use of a particular word, which then takes on new semantic attributes.

dismissed') (Gräßel 1991: 283). Gräßel (1991: 283) demonstrated marked differences between how female and male speakers produce language, which correspond to the findings of Trömel-Plötz and Hummel, but attributed them to the relationship between speakers vis-à-vis their standing according to the power dynamic in society, such as the expert/layperson dichotomy. For example, in both Gräßel's and Hummel's discourse analyses, higher-status (i.e. middle/educated classes) speakers of both genders interrupted those from a lower status more frequently, with women from a lower socio-cultural background being interrupted the most. These findings suggest that the power dynamics of a society - in which gender plays an important role - control the structure of conversation, the allocation of speaking time, and interaction between participants. Although this study does not examine socioeconomic factors, the discursive features that have been mentioned appear in YouTube comments and form part of the investigation, outlined in section 3.2.

2.2.2 Gender and Lexical and Syntactical Differences in Spoken Communication

Earlier research on lexical and syntactical differences in language between men and women in the English language generally agreed that women attempt to speak closer to the standard (Milroy & Margrain 1980: 63-64), are politer than men e.g. use less profane language and more indirect ways of speaking (Lakoff 1975: 50), and more "hedge words", such as *I think*, *you know*, *sort of* and *perhaps* (Lakoff 1975: 56–57; Coates 2013: 88–90). Much of this research has been disputed, particularly regarding gender as the dominant factor surrounding linguistic markers of politeness and hypercorrectness (cf. Steil and Hillman 1993; Hellinger and Bussmann 2003), with Brown (1991: 105–6) emphasising social class as the main

influence on linguistic politeness. Gisela Klann-Delius (2005: 57) concludes in her meta-study of German and English-language research on politeness that "what is clear, is that politeness is not a universal indicator of women's speech and it does not always signal a subordinate position [author trans.]".

Within other German-speaking research, investigations on linguistic difference across gender are mostly qualitative, rather than quantitative, due to the detailed nature of the discourse analysis approach, and lexis and syntax-related findings are often relegated to addenda to the main argument or small sections at the end of chapters on broader topics. For example, Trömel-Plötz (1982: 45-53) devotes just nine pages of 200 to lexical and syntactical differences, largely based on English language work by Lakoff (1975). Trömel-Plötz lists seven lexical and syntactical variables that characterise *Frauensprache* ('women's language'):

1. Women use diminutives and euphemisms more often (e.g. *Süppchen* for *Suppe*)
2. Women use different or very few vulgarities, such as coarse language, *double entendres*, and swear words[6]
3. Women may have a different vocabulary, in line with their 'traditional roles' concerning 'childcare', 'home-making', and 'fashion'
4. Women use adverbs and the conditional such as, *scheinbar*, *vielleicht*, or *würde sagen* as an 'Abschwächung der Aussage' ('weakening of the assertion') (Trömel-Plötz 1982: 48)

[6] Trömel-Plötz also observes how many insults are specifically aimed at characteristics found undesirable in women, such as *Beißzange* ('shrew') or *Klatschbase* ('gossip'/garrulous), whereas the same characteristics (argumentative, talkative etc.) are not denigrated and are even celebrated amongst men (she provides the example that male academics are encouraged to be *Klatschbasen*).

5. Women use more tag questions, such as *oder nicht?* or *nicht wahr?* (cf. Holmes 1984 for similar claims in English)
6. Women are more likely to use such modal particles or adverbs, such as *eben* or *nur* that also weaken the assertiveness of a statement
7. Women are more likely to express requests, commands, and opinions as questions, such as 'Willst du nicht hierbleiben?', or by using the subjunctive, for example, 'Ich würde mich freuen, wenn du X machen könntest'

Trömel-Plötz does not provide any quantitative research or other background information on the data, which leads to some doubt regarding the extent to which this characterises the language of women. Trömel-Plötz's findings have at various times been disputed or corroborated, such as in Gräßel (1991: 291-94), who provides statistical evidence that women used more questions and the subjunctive to formulate a request and more modal particles (6. and 7. above), but 86 of the 95 investigated features showed no statistically significant usage differences between genders. However, older studies, despite the questionability of the results, are still a useful starting point to reexamine linguistic features in a fresh context, such as online youth language.

As posited in section 2.2, there does not seem to be an agreed spoken genderlect with distinguishable lexical and syntactical features that would define a general 'men's language' or 'women's language', according to Lakoff. However, that is not to say there are not locally created differences depending on the specific communicative medium or space, which demands that a researcher strictly define their research context to explore how gender is performed linguistically. What is still generally agreed is that any differences that are present (and this also applies in

a socio-economic context) could appear to exist for reasons of linguistic gatekeeping created by the power dynamics within society.

2.3 Gender Performance Online

Unlike research on spoken language, with its conversation and discourse analysis approach, research on online language focusses more on the lexical, syntactical, and orthographical features of language. The field is dominated by Anglo-American research, such as Herring (1996a, 1996b, 1996c, 2003) and Danet (1998) and there is little work on how these features differ across gender in German. Research on online communication in German tends to focus on multilingualism (discussed in section 2.5), the relation of online language to standard German (discussed in section 2.2), and dialect (section 2.5.2). In the following sections, I discuss how gender is constructed and performed online, before exploring which features we might expect to find in online language in an English (and one Italian) context.

2.3.1 Constructing and Identifying Gender Online

The advent of computer-mediated communication was welcomed as a step towards neutralising gender differences in society, and by extension, linguistic differences. Balka (1993) quoted politicians of the 1980s, such as Tony Benn, who claimed 'computer networking technology' would lead to a more open, democratic, and equal society. She hypothesised that 'computer networks [could lead] to feminist social change' by allowing previously excluded populations access to power structures, such as means of communication. She supplied positive examples of feminist or other female-only mailing lists that provided spaces for women to communicate, but also alluded to the problems that designated women-only spaces had in dealing with unwanted male participants. From a discourse perspective,

Balka presented results from the 'Femail' mailing list that suggested that both men and women contributed to discussion in proportion with the gender demographic of the subscribers – in an environment where 74% identified as women and 25% as men, 71% of messages were from those women and 26% from men.

Despite the hopes of academics like Balka that online spaces would help reduce gender power differences, research has shown that these power differences instead manifest themselves in different ways to make up for the absence of face-to-face interaction in text-based communication. The prevailing finding of research in this field is that most online communicators, as found by Kendall (1998: 150) in her research on gender performance in an online chatroom, reflect offline sociocultural norms and expectations in their online choice of language, conversational topic, and username. Kendall goes on to state that 'offline identities affect [the] views of other's identity performances. Gendered […] identities continue to have salience in online interactions, with power relations often operating in much the same ways as they do offline' (Kendall 1998: 150). This is useful when investigating linguistic differences between genders online, as, in pseudo-anonymous spaces like You Tube comments, we are reliant on the assumption that commenters are replicating not only the language associated with a particular gender identity but also the performativity and discourse surrounding the interaction between gendered performances. This view is shared by Danet (1998), who explored gender performance in web chat, particularly men who posed as women and vice versa. Danet posited that gender experimentation did not help to break down the perception of a binary, rather it reinforced it, as those 'textual cross-dressers' resort to stereotypical and exaggerated linguistic production to

perform. Herring & Zelenkauskaite (2009) found similar results on an interactive Italian-language television show, where viewers texted in messages to appear on screen. They found that the language conformed to existing stereotypes: women produced language that 'symbolise[d] feminine qualities', whilst men wished to appear 'dominant' (discussed below in sections 2.3.2 and 2.3.3). Through such linguistic behaviour, the writer would gain or consolidate 'social capital' and therefore appear attractive to a potential partner (Herring & Zelenkauskaite 2009: 26).

2.3.2 Discursive Differences in Online Communication

This section outlines how sociocultural gender norms are manifested in discourse analyses on online communication. Herring (1996c) found that gender stereotypes are present in online communication, just as in spoken conversation. She initially conducted informal ethnographic research on early university discussion lists after noticing different reactions to disagreement in conversation from participants of different genders. Men, she observed, participated far more in discussions with high levels of adversarial language, whereas women 'committedly avoided' such interaction, instead preferring more cooperative environments (although disagreement still existed in these forums) (Herring 1996c: 147). Her later research (Herring 2003: 207-10) collated these gender differences into a list of five observations, which reflect, as in wider offline society, the prevalence of male-dominated discourse:

1. Men type longer messages than women
2. Men begin and close discussion topics, whilst women mostly contribute
3. Men assert their opinions as facts, whilst women use more justification

4. Men use cruder and more antagonistic language (referred to as 'flaming'); women use politer language and refrain from swearing
5. Men are more adversarial in their disposition towards fellow participants, whilst women are more supportive. Herring (2003: 208) provides an example of a man using sarcasm in reply to a statement with which he disagreed, whereas a female respondent stated that she disagreed but also complimented the original poster

Herring's five observations mirror those on gender in spoken German discourse by Hummel (1984) and Trömel-Plötz (1979, 1984b). This suggests we can expect to see similar language production differences within German computer-mediated communication, although no such research has yet been undertaken.

Not all subsequent research has corroborated Herring's observations. Baron (2004) investigated the Instant Messaging conversations of American college students and found that women produce longer messages and take more conversational turns than men, which seemingly contradicts Herring's first observation. The women's longest messages were longer than the men's longest messages, and men tended to produce more one-word turns than women (Baron 2004: 408-09). However, this study only used conversations between same-gender pairs, unlike the other research in this field. The differences posed by Baron suggest that perhaps in a mixed setting, men think they must uphold the existing sociocultural power balance (cf. Holmes 1992: 166), and therefore attempt to dominate the conversation by "talking" for longer. Alternatively, in Baron's work, we see Herring's fifth observation in action: the men's adversarial conversational style leads to shorter conversations and conversational turns in a same-sex

environment. A similar phenomenon can be observed in Herring & Zelenkauskaite (2009), where the nature of the continuous message stream meant that the writers of the text messages could not easily interact with each other and form conversations. In this environment, the majority of the texts were sent by women and were longer (Herring & Zelenkauskaite 2009: 19-20), which again does not support Herring's first observation. Despite the research not being completely congruent, we could hypothesize that we will see longer and more numerous messages from men in a mixed environment but shorter exchanges in male-male interactions, although the tone and register of language used within those exchanges is likely to remain similar to spoken language. This is examined in sections 4.1.1 and 4.1.3.

2.3.3 Gender and Lexical and Syntactical Differences in Online Communication

Besides discourse and conversation practices, there has been research into how gender differences are reflected in the lexis, syntax, and orthography of online communication, particularly in how these features might deviate from the standard, although there is disagreement about which forms they might take. Research agrees that women repeat exclamation marks (known as punctuation prosody, the repetition of punctuation marks for emphasis) (Waseleski 2006: 1020; Herring & Zelenkauskaite 2009: 23) and letters within a word (grapheme prosody, e.g. *hiiiiiii*) (Herring & Zelenkauskaite 2009: 23) more than men. However, while Witmer & Katzman (1997: paragraph 16) and Baron (2004: 416) found that women used more emoticons than men, Wolf (2000: 831) found no significant difference. Baron (2004: 415) claimed that men use more contractions, such as *can't* or *won't*, and CMC language, i.e. *LOL* or *OMG* (excluding emoticons), but Herring & Zelenkauskaite

(2009: 22) argued that the reverse is true and these are features of women's online language. They claimed that non-standard orthography from women is part of a linguistic performance that 'symbolise[s] feminine qualities of emotionality, sociability, and playfulness' (Herring & Zelenkauskaite 2009: 26). Herring & Zelenkauskaite also found that men's CMC language represented a more standard version, which is 'symbolic of masculine association with the dominant societal institutions'. This is seemingly at odds with research in spoken language that found that men use more non-standard language, such as slang and dialect, than women (cf. Labov 1966, 1972). Instead, it is more accurate to say that their findings indicated that men's language in text messages reflected spoken norms and appeared to favour efficiency over creativity or conveying emotion. As a result of the disunity in the academic field regarding these features, it is important to see how all of the above features apply to the YYPC, if indeed they appear at all.

2.4 Youth Language Online and Offline

This section examines the features of youth language and how they differ from standard spoken and online language, before discussing the role of gender in adolescent language production. There is less pressure for youth language to adhere to standard language, or conversely more pressure to use 'slang' or 'youth language' - Holmes (1992: 183) goes as far as to say using slang is the 'linguistic prerogative of young people'. The majority of research agrees that, for young people, the importance of identity construction in this transitional and developmental stage is paramount (Androutsopoulos & Georgakopoulou 2003; Dürscheid & Brommer 2009; Tagliamonte 2016a; Tagliamonte 2016b), and much of this is reflected in their language production, according to Eckert (2003: 381).

The majority of German research on online language focuses on general linguistic features such as the extent to which online language displays conceptual orality (Kilian 2001; Storrer 2001; Schlobinski 2004; Wirth 2005), how typing affects language production (Siever 2006; Albert 2013), or regional differences (Siebenhaar 2006; Androutsopoulos 2013). However, few studies investigate youth language online. Kleinberger Günther & Spiegel (2006: 107-13) list a range of lexical, syntactical and orthographical features of youth language on the internet, such as shouting capitals and punctuation prosody, but my research indicates that these are no different to features found across all kinds of online language. Dürscheid (2006: 126) concurs, querying whether the non-standard features found in youth language are genuinely produced as part of performing youth identity or whether they are a product of the online medium.

In any case, we can expect youth language online to exhibit the non-standard features that have been investigated in more general CMC research, such as grapheme prosody, emoticons/emoji, and other orthographical creativity, but perhaps at a higher frequency due to the tendency of young people to use language further from the standard. We can also expect to see a youth-specific lexis and syntax. Androutsopoulos (2004: 180-81) provides examples of these features in five groups (examples in brackets are his own from 2004):

1. Adapted Anglicisms, particularly Germanised verbs with particles (e.g. *auschillen* from 'to chill out')
2. Innovative word building patterns, such as suffixing *-i*, or *-o* (e.g. *Grufti* [Goth], *Normalo* [Normie])
3. Intensifiers in the first position of compound nouns (e.g. *Killer-, Bomben-*)

4. Compound nouns ending in 'identity words', such as *–junkie, -freak* that demonstrate affinity to a concept (e.g. *Schokoladenjunkie* ('chocoholic'))
5. Other lexical coinages or lexical changes (e.g. the semantic change of *geil* from 'horny' to 'cool')

Androutsopoulos's findings tally with previous research by Schlobinski (1995: 319-21), who stressed the exaggerated nature of youth language, as well as its inherent creativity that serves to 'distinguish […] "young people" from the rest of society' (Schlobinski 1995: 319). However, as Schlobinksi illustrates with several examples from sketches and jocular translations of classic German literature into supposed youth language, it is important to distinguish popular perceptions and stylisations of youth language (sometimes held 'responsible for the "decline of the German language"': Schlobinski (1995: 317)), from how it is actually produced. Spitzmüller (2006: 34-35) lists several articles from German mainstream media outlets in the 1980s that all lament the supposed inability of German youth to speak "proper" German. A decade later, he finds the same papers saying the same thing, although there is a marked shift to criticising the supposed influence of *Kanak Sprak*, a derogatory name for the multiethnolect spoken by some young inhabitants of Germany with a migration background (Spitzmüller 2006: 43-44). The relation between youth and ethnicity is considered below in section 2.5.

The transitory nature of youth identity means that there is never any one youth language in Germany. The language spoken by young people changes rapidly and varies wildly depending on geographic location and youth subculture. The statement by Schlobinski (1995: 326) that there are as many youth languages as there are youth groups is supported by Watzlawick (2006: 290), who lists many

lexical and syntactical differences between youth groups, such as heavy-metal fans, skinheads, and Hip-Hoppers. One of her questions asked which word they would use to describe a positive experience. The answers varied significantly (for example, *hammer*, *heftig*, or *krass*, respectively), but there were significant overlaps: most of the groups used the word *geil* or a variation thereof, e.g. *affengeil*. We can therefore expect young people to produce a wide variety of non-standard language online, some of which may be particular to the YouTube or online context. There may also be variations in what kind of online-specific language, depending on gender or ethnicity, or even within the specific context of the YouTube video, e.g. the genre of music video.

2.5 Migrant Background, Gender and the German Language

Holmes (1992: 166) states that 'sex [gender] differences in language are often just one aspect of more pervasive linguistic differences in the society reflecting social status or power differences.' Indeed, language is not just used as a gatekeeper to enforce a gender hierarchy, but also affects a social hierarchy based on ethnicity. In the USA, for example, language is used to enforce the 'linguistic supremacy' of white people over African-Americans (and others), according to Samy Alim (2006: 59). Denigrating the African-American sociolect restricts access to power structures; linguistic gatekeeping forces African-Americans to conform to 'White norms' in order to attempt to avoid discrimination (Samy Alim 2006: 66), but in doing so then face the accusation of 'acting white' (Bergin & Cooks 2002: 113), and risk alienating themselves from their peer group.

Similar attitudes exist in Germany and Austria towards those with a migrant background. Labourers, often from Turkey or the Mediterranean and Balkan states,

were initially invited to Germany in the 1960s for temporary work, but many were needed for several years and so they brought their families and made Germany their home. Many of their descendants live in undesirable inner-city housing developments, often referred to as 'ghettos' (Rollefson 2013: 227), with other migrants from similar socioeconomic backgrounds. The relatively dense concentration of migrants and their families in urban areas has led to the creation of various urban ethnolects of German, previously called *Gastarbeiterdeutsch* ('Foreign Worker German') (Barbour & Stevenson 1998: 214) but also referred to as *Türkendeutsch* (Turk-German) (Androutsopoulos 2001: 321) or *Kiezdeutsch* (''Hood German') (Wiese 2009: 782). Foreign Worker German originally referred to the 'imperfect' acquisition of German by recent arrivals to the country, and as such, is no longer an accurate term to use to describe the language of those with migrant heritage. Many were born and raised in Germany or Austria, and often hold the citizenship of one of those countries, as well as having fluent German in their linguistic repertoires. Therefore, I will refer to the ethnolects collectively as *Kiezdeutsch*, as that is the term used by the speakers of this multiethnolect (Wiese 2009: 783), and which is also not linked to any one particular migration background, indicating only the urban location of the speakers.

2.5.1 Migrant Background and Identity

Having a migration background adds an extra level to the internal negotiation of one's identities. Children and young adults are most affected, having the added pressure of forming an identity as a young person (Bierbach & Birken-Silverman 2007; Keim 2007b; Lytra et al. 2010; Prokopiou & Cline 2010). The language choice and behaviour of children from migrant families therefore becomes an 'act of

identity' (Le Page & Tabouret-Keller 1985: 178-82), which is used to create and project many overlapping aspects of their mixed cultural background.

Two studies that focus on the language of groups of young individuals from a migrant background in Mannheim are Bierbach & Birken-Silverman (2007), who investigated a mixed-gender group of Sicilian-German young people, and the ethnographic studies by Keim (2007a, 2007b) on an all-female group of Turkish-German teenagers. Both examine how the tension between social pressures, e.g. from parents, mainstream institutions, and their peer group, is important in creating and maintaining identities, and how this is represented in the language used by the groups. The Turkish-Germans, who referred to themselves as the 'Powergirls', communicated using a mix of Turkish and Mannheim-accented German, whereas the Sicilian-Germans used a code consisting mostly of Sicilian-Italian dialect, interspersed with Mannheim-accented German, but standard Italian and English were also used to humorous or playful effect. This reflects the sociocultural forces that shaped the group members' identities: Sicilian and Turkish were the languages of the family and common culture for the groups, whereas German represented the language of mainstream society, from which they often felt marginalised (Bierbach & Birken-Silverman 2007: 125). Keim states that German institutions do not support bilingualism, which made integration more difficult and alienated many with migrant heritage, especially children in the school system. This in turn could lead to a vicious cycle of poverty and marginalisation, where the Powergirls, and many like them, were often forced to remain within their immediate communities in an environment of German-Turkish mixing (Keim 2007b: 157).

Indeed, mixing these languages within the same speech act[7] is often impenetrable to outsiders. For the Powergirls, Keim (2007b: 162-63) suggested that language choice had particular discursive and sociosymbolic functions: specific constellations and topics often favour one language over another. For example, in times of high emotion, Turkish was favoured. However, for both groups, an 'in-group' minority language or dialect might be used in conversation with a member of an 'out-group' (e.g. German and adult versus migrant background and child) to subvert existing power structures, because the 'out-group' member is reliant on the 'in-group' for social or behavioural cues due to the language barrier. This is especially appealing for groups or individuals who feel marginalised by society.

The groups differed in how they were perceived by society and the affinity they had with wider social trends. For example, Bierbach & Birken-Silverman (2007: 128-30) argued that the Sicilian-Germans in Mannheim also identified with a 'glocal' (global and local simultaneously), 'modern urban youth culture', characterised by the use of English as an acrolect[8] to emphasise worldliness or coolness. However, the use of English did not form part of the identity and language of the Powergirls; rather, they remained local, with their language shaped by German, the Turkish of their families, and the *Kiezdeutsch* of their neighbourhoods (Keim 2007a: 448-60). Caglar (1998: 244-245) discussed how the perception of the ethnic group by mainstream society shapes their actions and constructions of identity – Turkish-

[7] Language mixing, here, means the use of two or more languages within the same utterance. It has also been referred to as codeswitching (Heller and McClure 1988), translanguaging (García & Wei 2014: 2), or codemeshing, heteroglossia, metrolingualism, and polylanguaging (Lewis, Jones & Baker 2012: 650). For the purposes of this study, the phenomenon will be referred to as language mixing.

[8] An acrolect, as opposed to a basilect, is a language variety that holds overt prestige, such as Received Pronunciation vs Norwich Working Class Speech (cf. Trudgill 1972).

Germans are also alienated based on skin colour and religion, unlike the Sicilian-Germans. Whereas the Turkish community of London enjoy a (mostly) positive reputation as a migrant group (cf. Lytra et al. 2010), the ethnicity and religion of the Powergirls were seen as incongruent with German values (Keim 2007: 159). This research reaffirms the importance of identity in shaping the linguistic patterns of an individual, especially the relationship which that individual has with mainstream institutions and language practices.

2.5.2 Migrant Background and Gender

Identity can have a significant effect on language behaviour, and may be particularly salient in an intersectional group, such as females with a migration background. The studies by Bierbach & Birken-Silverman (2007) and Keim (2007a), also examined language differences between male and female individuals with a migration background. The Sicilian-German male youths analysed in Bierbach & Birken-Silverman (2007) were far more predisposed to use Sicilian dialect than their female counterparts. Dialect is important to construct an authentic male identity, according to the authors (Bierbach & Birken-Silverman 2007: 139-140). By contrast, the teen girls actively rejected their community language whilst in dialogue with their peer groups, instead only using it when 'back home' to talk to elderly relatives (Bierbach & Birken-Silverman 2007: 139). However, they still self-identified as Sicilian-German and remained in the peer group. Early research from Labov (1972: 243) and Trudgill (1972: 182-83) suggested that such language behaviour is due to the 'less secure' social position of women, particularly lower-middle class women, where 'it is more necessary for women to secure and signal their social status linguistically [by using prestige forms]' (Trudgill 1972: 182). Bierbach & Birken-Silverman (2007: 140)

disagree implicitly with this judgement: they state that the boys' linguistic performance contains such conscious Sicilian-German language mixing as part of a demonstration of masculinity, to compete with other males for the 'internal group status and […] the girls' attention'. The Sicilian-German girls, on the other hand, did not need to demonstrate such linguistic creativity, as they have access to a form of power by choosing how they award their attention to the boys in the group. Similar projections of masculinity occurred amongst the wider Turkish-German population through the use of Turkish lexis and morphology (such as Ø production of articles or prepositions (Auer 2003: 258-59)) that have been adopted into *Kiezdeutsch*, according to Androutsopoulos (2001), Auer (2003) and Tekin & Colliander (2010). Many examples of *Kiezdeutsch* lexis in Tekin & Colliander (2010: 51-52) refer to a masculine interlocutor, such as in the terms of address that are common in *Kiezdeutsch* (e.g. *Alder*, *lan*, *moruk*), which may indicate that the use of *Kiezdeutsch* is more popular amongst males. Interestingly, the language of the Powergirls in Keim (2007a: 386-415) does not conform to these gender expectations. They consciously used typically Turkish-German male language patterns, such as coarse language, Turkish slang and frequent language mixing, whilst discussing typically feminine topics, such as fashion, to dissociate themselves from all expected sociocultural and linguistic behaviour and create their own subculture. This stands in constrast with what they see as opposing and harmful forces: their traditionalist parents, a mainstream society that does not understand them, and other Turkish-Germans who either harass them or who conform to what they see as harmful stereotypes (i.e. the hypermasculine male or the ultra-religious female).

Transcending or rejecting stereotypes is viewed as important in explaining

linguistic behaviour, regardless of ethnic background. Auer & Dirim (2003: 233-34) provide a case study involving three young female teenagers of German background, who consciously learned Turkish and spoke it regularly, particularly using Turkish set phrases or words, such as *para* ('money'), in place of German. They reported that such language production 'was a means of being accepted and liked' and enabled the girls to feel part of the local 'multi-cultural [...] youth-cultural scene', whilst simultaneously rejecting the perceived 'prejudice' (the girls' own words) that they claimed is often shown to migrants by ethnic Germans.

Given the pseudo-anonymity of YouTube, it will be interesting to see which of the features discussed here are replicated in the YYPC, such as the potential differing attitudes towards typically male language practices between females of different migration backgrounds.

2.6 Research Gaps and Questions

As discussed above, there are several points of consensus and disagreement as to linguistic differences between genders and ethnicities, in online language and in the interaction between these variables. The following investigation explores these phenomena in a new context compared to previous research and explores the research gap concerning the intersection of these variables.

The disagreement on how language is used across gender and ethnicity illustrates the need to add to our knowledge by examining these features in a new context – that of You Tube comments. The same applies to discursive features, such as turn length or "speaker" attitude. Research is also still needed to understand how the gender differences interact with different migration backgrounds, and the differences in language practices across ethnicity, as there is very little research on

the language of those with a migration background online, apart from on language mixing (cf. Androutsopoulos 2006, 2008, 2010, 2015).

These research gaps provide the impetus for the research questions that guide this essay. In particular, we must consider whether linguistic features considered indicative of male and female gender performativity are present in an online youth context. If so, which linguistic features reflect the differences between males and females with a non-migration background, and does that differ for males and females with a migration-background? The following section explains the methodology for investigating these questions, outlining which features are investigated and how the corpus was compiled.

3 Methodologies

This chapter first explains how I collated and compiled the corpus of comments for the YouTube Young Person Corpus and methods to assign gender and ethnicity to commenters. The second part of the chapter presents my hypotheses based on previous research and the linguistic features that were examined to test them. Finally, I explain the statistical tests and models used to test the hypotheses. An overview of the YouTube Young Person Corpus is presented below in Table 2.

Feature	Number
Number of comments collected	1 855
Outliers removed as per section 3.1.6	4
Number of comments used in YYPC	1 657
Total number of words in YYPC	18 338
Comments from men	593
Unique male commenters	508
Comments from women	1 063
Unique female commenters	974
Comments unable to be classified according to gender	194
Comments from non-migration background	532
Unique non-migration background commenters	473
Comments from migration background	405
Unique migration background commenters	345
Comments unable to be classified according to ethnicity	850
Total number of conversations	116
Total number of comments in conversations	551

Table 2 Key Statistics of the YouTube Young Person Corpus (YYPC)

3.1 Compiling the Corpus

A YouTube comment is one entry underneath a YouTube video that is produced by one person by entering text into the comment box and pressing enter. Other visitors to the same page can also leave separate comments or reply to other people's existing comments by clicking the reply button. A typical comments section is laid out in Figure 1. The comments were collected from 18 videos and compiled to form the 'YouTube Young Person Corpus' (YYPC) of 20,919 words in 1855

comments. Twelve of the videos were 'listicle'-style videos from the 'Lifestyle' genre,[9] selected from four demographics:

a. Three male YouTubers (two with a German, and one with an Austrian ethnic background)

b. Three female YouTubers with a German ethnic background

c. Three male YouTubers with a migrant background to Germany

d. Three female YouTubers with a migrant background to Germany

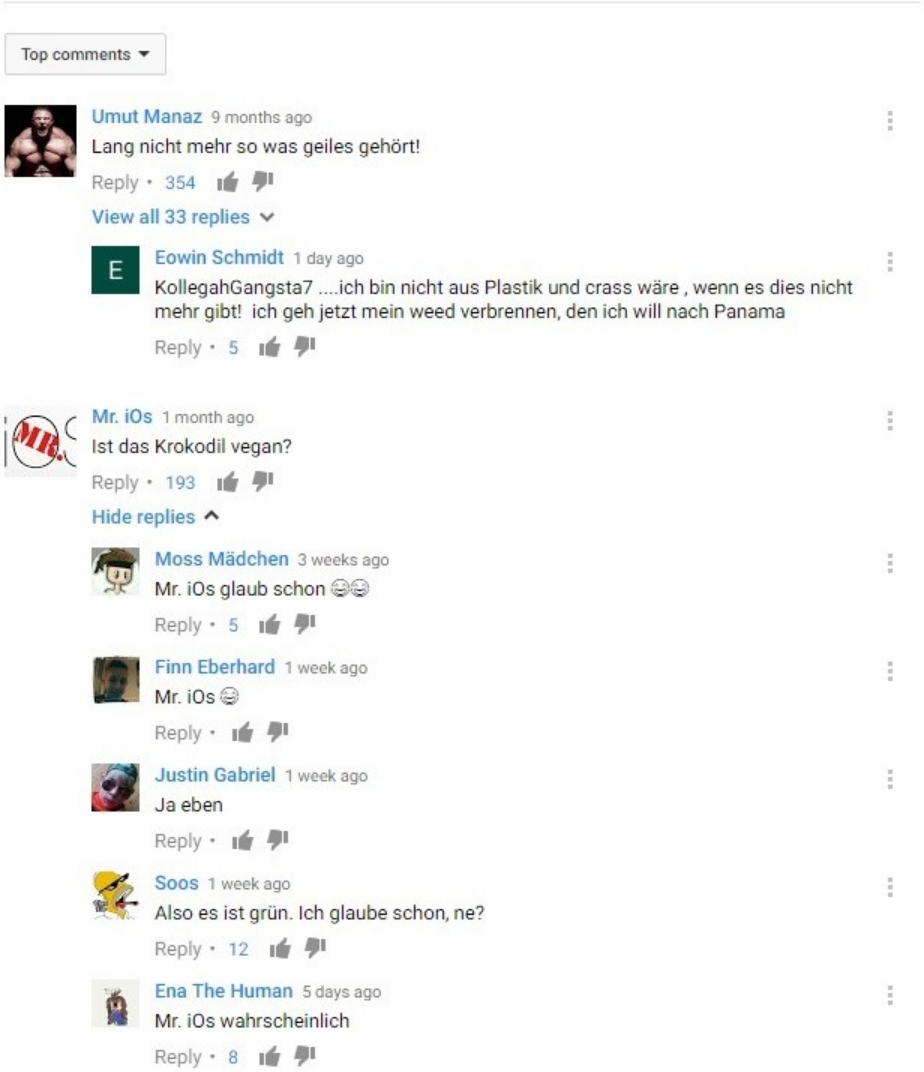


Figure 1 Typical Layout of the Comments Section of a YouTube Video

[9] A listicle, a portmanteau of 'list' and 'article', is a video format where the YouTuber presents content in the form of a list, such as Michael Buchinger's 'Hass-Liste für den Mai 2017', where he goes through several unrelated issues that have annoyed him in May.

Six of the videos were music videos from three different genres: two each from Mainstream Hip-Hop/Rap, Niche Hip-Hop/Rap, and Mainstream Pop. Selecting videos from two areas and from several demographics and genres ensured a wide variety of commenters and varied language practices that covered a broad cross-section of German youth culture. Listicle videos, by covering a variety of topics, further expand the conversational stimuli for commenters, which prevents comments from fixating on a specialised issue that may contain specific vocabulary. Further criteria for the collection of the videos are listed below, detailing the general parameters for the collection of the videos, as well as the specific parameters that apply to the listicle videos and the music videos.

3.1.1 General Parameters

1. The video was uploaded to YouTube by an official source responsible for the creation of the video, for example, a YouTube verified account
2. German was the main language of the video
3. Each video was created by a different person, which ensured a broader range of language
4. The video was uploaded between April 2015 and May 2017 to ensure the videos reflected contemporary youth culture and would target the youth demographics
5. The videos were aimed at the 13-17 and 18-24 demographics, established by analysing the following:
    a. If the YouTube featured in teen magazine features (e.g. BRAVO)
    b. Age-appropriate content matter (e.g. pranks or toys)
    c. YouTube data on viewership
    d. YouTube channel suggestions (see Figure 2)
        i. Channels suggested from YouTube's algorithm that highlights channels with similar content and demographics

ii. Channels suggested by the YouTubers' on their "homepage"

6. All comments and metadata (profile picture, username of commenter etc.) collected from between May 2016 and May 2017, as indicated by YouTube's on-page time-stamp next to the commenter's username, were collected. This ensured the YYPC reflected the most up-to-date language
7. A contiguous block of 150 comments or 1300 words, whichever was the larger, was taken to be analysed and formatted
8. All comments that could not be suitably analysed were then removed, including:
    a) Comments written entirely in a language other than German
    b) Comments entirely consisting of ASCII art
    c) Double-posted comments (i.e. duplicates)
    d) Copy and pasted song lyrics and transcripts

3.1.2 Parameters for the Listicle Lifestyle Videos

1. The YouTubers were either verified by YouTube's "Blue Tick" scheme, acknowledging their influential status or had over 100,000 subscribers to their channel

3.1.3 Parameters for the Music Videos

1. The video had either:
    a) over 10m views (Mainstream Hip-Hop/Rap and Mainstream Pop)
    b) fewer than 2.5m views & 10,000 likes per day average, indicating a more underground scene (Niche Hip-Hop/Rap)

Table 3 shows the videos selected using these criteria and the video reference code, which was assigned to all comments, along with a number, to easily identify under which video they were written, e.g. DF1 – 009 refers to Svenja Lang - *cool macht mehr von diesen viedeos* [sic].

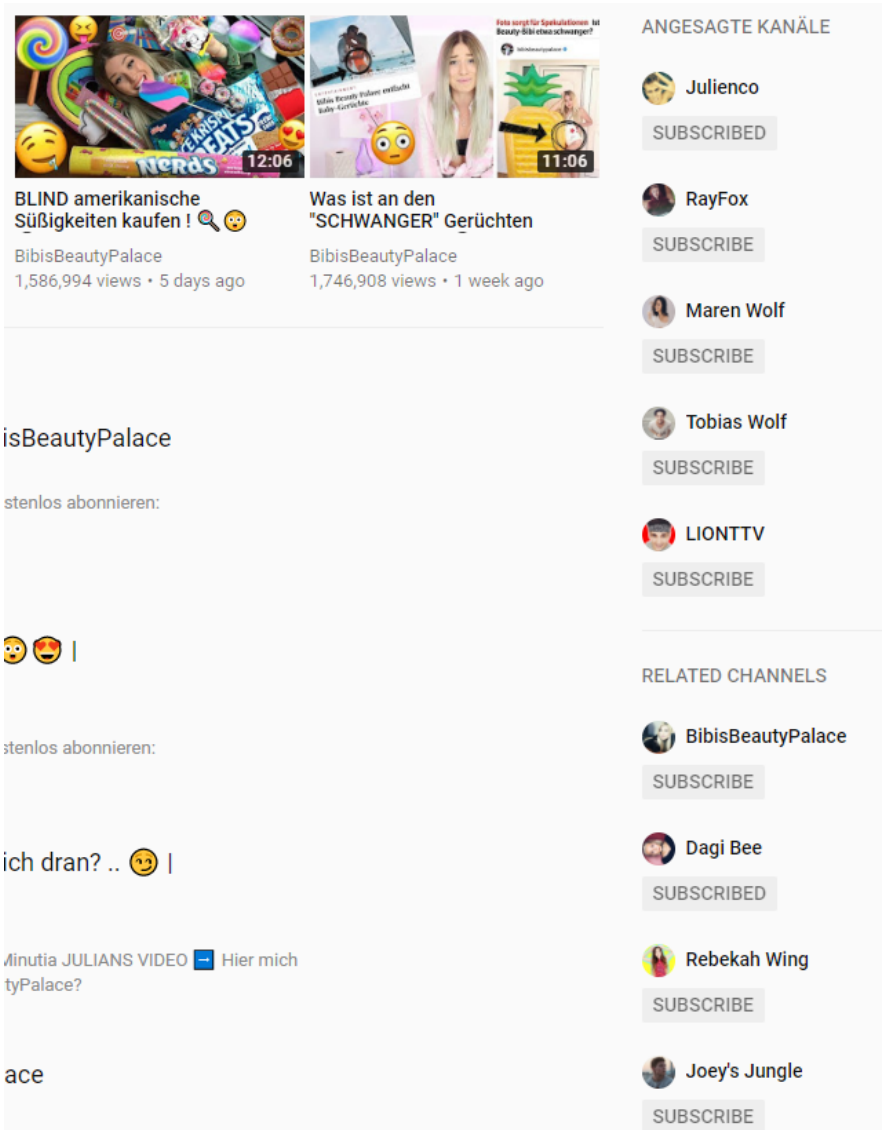


Figure 2 Section of YouTube "homepage" of BibisBeautyPalace with suggested channels on right hand side ('Angesagte Kanäle' from Bibi, 'Related Channels' per YouTube algorithm)

No.	Video Name	YouTuber/Artist	Video Reference Code
	Lifestyle 'Listicle' Videos		
	Male, Non-migration background		
1	Ich bin KEIN Kätzchen!!! _ Michaels Hass-Liste für den Mai 2017	Michael Buchinger	GM1
2	DINGE, die man im HOTELZIMMER NICHT TUN sollte	Joey's Jungle	GM2
3	TOP 10 - so krank ist BIBI eingeschlafen 🤒😂	JulienCo	GM3
	Female, Non-migration background		
4	10 PERIODE HACKS die JEDES MÄDCHEN kennen sollte! ♡	Bonnytrash	GF1
5	AMERIKANISCHE SPIELSACHEN , die ihr garantiert nicht kennt 🙄	BibisBeautyPalace	GF2
6	.. Meine schlimmsten HATER KOMMENTARE 😳 #4	DagiBee	GF3
	Male, Migration Background		
7	7 ARTEN ZU KÜSSEN - TEIL 2!! 😉 mit Freundin	Simon Desue	MM1
8	Die 10 DÜMMSTEN Gesetze der WELT !	Tugay	MM2
9	TÜRKE vs AFRIKANER 2	Der Ömsen	MM3
	Female, Migration background		
10	5 TOP & FLOP Drogerie Produkte	Hatice Schmidt	MF1
11	50 DINGE, die ihr schon immer über mich wissen wolltet	Ischtar Isik	MF2
12	Die 5 Größten FEHLER beim RASIEREN!! - INTIMZONE / Eingewachsene Haare / Rasierpickel	Nihan	MF3
	Music Videos		
	Mainstream Hip-Hop/Rap		
13	Sex ohne Grund	Ali Bumaye ft. Shindy	MR1
14	Palmen aus Plastik	Bonez MC & RAF Camora,	MR2
	Niche Hip-Hop/Rap		
15	Wir sind friedlich	SXTN	NR1
16	Fluchtwagen glänzen	Capital Bra ft. King Khalil	NR2
	Mainstream Pop		
17	Auf uns	Andreas Bourani	MP1
18	Pocahontas	AnnenMayKantereit	MP2

Table 3 List of Videos that form the YouTube Young Person Corpus

3.1.4 Assigning Gender and Ethnicity to Commenters

To examine the relationship between gender, ethnicity and online language use, a gender and ethnicity was ascribed, where possible, to each commenter. Ascribing a gender or ethnicity identity to someone without consulting them must be justified. Both gender and ethnicity are sensitive topics, and this essay does not claim to have assigned the "true" gender or ethnicity for any one individual. Rather, I have

identified commenters by how their online profile performs a gender or ethnicity, although research by Kendall (1998: 150) suggests that nearly all online avatars reflect the gender roles of the person behind them. The investigation adapts Butler's (2006: 185-93) framework suggesting that how one displays or "does" gender is the result of a collection of sociocultural forces that create a gender identity, which in turn exerts pressure on an individual to repeatedly act (or talk, for example) in a particular manner, such as the expectation that girls should play with dolls. This is referred to as gender performativity. Gender performance, on the other hand, refers to consciously conforming to or transgressing the expectations of performativity. The pseudonymous nature of online communication offers the opportunity for individuals to perform gender in different ways to what one might expect from them in an offline setting. This means that much of online research, when discussing results of gender differences, also potentially includes those who wish to 'transgress' and more or less consciously produce language in the style of another gender.[10]

A similar framework can be applied to ethnicity (Clammer 2015: 2159), where the performative aspect is often referred to as 'ethnifying' (García 2010). However, the presentation and performance of ethnicity is far more nuanced to frame than gender, and while this study attempts a basic categorisation of commenters into binary ethnic groups (migrant heritage and non-migrant heritage), I acknowledge that the methodology could be more sensitive. However, the scope of this thesis would not allow a more rigorous approach.

[10] Research does exist on gender crossing and trans identities online (Kendall 1998; Huffaker and Calvert 2005), but there are insufficient data to include this aspect in this study.

For clarity, it is worth repeating that the 'real-life' gender or ethnicity of the commenters is not assigned, rather the gender or ethnicity performance according to their username and associated metadata of their account. I assigned a gender to the commenter using the steps below in sequence:

1. YouTube encourages users to sign up for their YouTube account with their Google+ name, which, like the other social media website, Facebook, is assumed to contain a person's real name, e.g. *Claudia Müller*. My first step to establishing the gender of the author of a comment was to see if their name was traditionally male or female. Through step one, 842 comments were identified as female and 448 as male.
2. Next, if that could not be unequivocally established, such as in the case of gender-neutral names, I examined the profile picture to see if the commenter in question presented as male or female; all users are required to have a profile picture that is displayed on their channel homepage and next to every comment that they leave. An example of a profile picture is given in Figure 3. Through step two, 76 comments were identified as female and 35 as male.



Figure 3 Example of Comment Gender Identified through Profile Picture

3. For users who did not use their real names as their usernames, instead preferring a variation (e.g. *Emily das Erdbeer*, *DieEchteFiona*) or representation of their interests (e.g. *Abbas_Gamer*, *Noemi.Potterhead*), I assigned a gender based on the first name, verified wherever possible by

their profile picture. Through step three, 121 were identified as female and 67 as male.

4. Some usernames contained a gendered noun (e.g. *AlbanERxx722* or *Marieposa Prinzessin*), which was used to ascribe a gender. Through step four, 13 were identified as female and 28 as male.
5. I examined the content of the comment. Where they explicitly identified as male or female, this was used to classify their gender performance (e.g. schni.travels.the.world, *ich mag clue auch aber iwie stimmt es bei mir nie.meine tage gabe* [sic] *ich soooo unregelmäßig* ('I like Clue [a period-tracking app] too but somehow it never works with me. I have my period soooo irregularly')). Through step five, 14 were identified as female and 17 as male.
6. In case of any remaining doubt concerning which gender the commenter was performing, the comment was excluded, even where there were some strong indicators for a particular gender performance, such as a username indicating stereotypically male or female interests, such as gaming (*eXtasY cs go*) or fashion (*fashionaddicted 4ever*). There were 194 comments (10.5%) for which gender could not be assigned. A breakdown of how commenters were assigned gender is presented in Table 4.

Assignment of gender to comments	Male	Female
Full name	448	842
Profile picture	35	76
First name + profile picture	67	121
Gendered noun in username	28	13
Gender in comment content	17	14
Total	595 (35.8%)	1066 (64.2%)
Comments not able to be assigned gender	194	

Table 4 Assignment of Gender to Comments

A similar process was applied to determine the ethnicity of the commenters:

1. For commenters with usernames representative of a full name, such as those linked to Google+ accounts, the commenter was classified as non-migration if the surname or first name was typically Germanic and as migration background if not, e.g. *Jennifer Hoffmanns* for a non-migration background and *Dilara Dogan* for a migration background. As the comments were all in German, it was considered reasonable to assume that a commenter with a German surname also had a German-speaking background. By contrast, an English-language study might have issues here due to widespread migration to English-speaking countries (particularly the USA) and its status as a world language. The definition of non-migration in this essay is a broad one: if a commenter is performing as though they do not have migration heritage, coupled with a name that is Germanic, they will be considered as such. This classification therefore does not distinguish between commenters who are second or third generation migrants from commenters with a longer history of German citizenship, as the focus of the study is the performance of ethnicity (cf. García 2010). If a commenter with multiple or intersecting ethnic identities wishes to overtly highlight their migration background, they are categorised as such, such as *Nico W*. One might assume from his Germanic name that he would identify as having a non-migration background but he writes “Ich bin auch russe […]” (‘I am also a russian’), thus identifying with the migration background group at this juncture in time. He might not identify as Russian at all in ‘real life’ or in comments under other videos, however, in this particular space, he

performs having a migration background. Through step one, 462 commenters were identified as having a non-migration (ethnic German) background, and 349 were identified as having a migration background.

2. There is also a phenomenon amongst German speakers, particularly females with a non-migration background, of splitting their first name in two and using the second half as a surname, ostensibly for reasons of privacy, e.g. *LE NA* or *Sa Ra*. Through step two, 38 were identified as having a non-migration background and four a migration background.
3. If a commenter used an ethnic determiner in the username, this was taken as indicative of their ethnicity, such as *Leila Kurdin*. Through step three, eleven commenters were identified as having a non-migration background, and 28 were identified as having a migration background.
4. If the username was not conducive to establishing the ethnicity of the commenter, the comment content sometimes indicated an ethnicity performance, which was used to classify the comment, e.g. *Aly Suchti* – …ich bin die einzige Türkin in meinem alter [*sic*] in meinem Umfeld… ('I am the only Turkish girl in my social circle'). Through step four, 21 commenters were identified as having a non-migration background, and 25 were identified as having a migration background.
5. In case of any doubt concerning which ethnicity the commenter was performing, the comment was excluded, such as in the case of Polish surnames, which are common amongst those who might consider themselves "ethnically" German/Austrian. This inevitably excludes some commenters with maternal German/Austrian heritage if they took their

father's surname, but such inaccuracies are unavoidable in such a study. Overall, 723 (43.9%) comments could not be assigned an ethnicity. Although a significant percentage of comments could not be assigned an ethnicity, the corpus was still relatively large, at 10,341 words. A breakdown is provided in Table 5.

Assignment of Ethnicity to Comments	Non-Migration	Migration
Full name	462	349
Split first name	38	4
Ethnic determiner in username	11	28
Ethnicity in comment content	21	25
Total	532 (56.7%)	406 (43.3%)
Comments not able to be assigned ethnicity	732	

Table 5 Assignment of Ethnicity to Comments

3.1.5 Analysing the Parts of a Comment: Comment Construction Units

As mentioned in section 2.1, online language is often representative of a written form of spoken language, which means that if we are to subdivide a comment into smaller parts for analysis of e.g. average comment length, using individual words as the constituent part of a comment would not accurately convey the conceptually oral nature of online language. Comparing comments based on word length often cannot convey the complexity of a comment or conversely, identify similar comments, as it does not account for the variability in the number of words that can be used to produce a semantically similar utterance. For example, comment MF1 – 013, *Deine Haare sehen so toll aus* ('Your hair looks so great') and comment MF1 – 033, *Deine Haare sind wunder wunder wunder wunder wunder wunderschön* ('Your hair is wonder wonder wonder wonder wonder wonderful') both semantically convey the same message, but differ in the number of words (6 vs 9 words). To better compare the complexity and lengths of online comments and account for their often speech-like nature, I adapted the concept of Turn

Construction Units (TCUs) from conversational analysis for online language, which I call Comment Construction Units (CCUs).

According to Schegloff (2007: 3-4), TCUs are 'sentences or clauses more generally, phrases, and lexical items'; Hutchby & Wooffitt (1998: 48) state that 'turn-construction units […] broadly correspond to linguistic categories such as sentences, clauses, single words (for instance, *Hey!* or *What?*) or phrases', or indeed, can function as a complete utterance. Although the definitions may seem nebulous, they account for the phonetic and discursive nuances in the context of a conversation, which Schlegoff calls the 'packaging' of a TCU. As Hutchby & Wooffitt (1998: 48) warn, one 'cannot take a prescriptive stance on this question [the definition of a TCU], because what a turn-construction unit consists of in any situated stretch of talk is a [conversation] members' problem.' That is, it differs based on the interactive behaviour of its participants and the synchronous nature of a spoken conversation.

Internet-based messaging, such as the comment section of a YouTube video, is an asynchronous form of communication, where interruptions do not exist. This allows us to define a CCU, the corresponding online version of the spoken TCU, following a set of criteria, which ensures a level of consistency for analysis. The criteria for establishing the boundary between CCUs were developed as part of a previous pilot investigation (Cotgrove 2017; unpublished) using YouTube comments and are listed below. In the illustrative examples, the end of a CCU is indicated by a double slash ('//'):

- At the end of a self-contained part of speech (noun phrase, verb phrase etc.), e.g. *ohhh kukuuu // 1 million klicks // braaaa //*

- At places of standard punctuation, e.g. the comma separating clauses in *Isso,// halt 'n Fakt*
- Emoji/emoticons used as punctuation that typically mark the end of a clause, e.g. *Dieser* [sic] *Lied hat mir geholfen mein flickflack zu lernen // Ich hatte Angst […]*
- Clause boundaries (relative, subordinate, coordinate etc.) regardless of whether standard punctuation is used, e.g. *Dieser Lied* [sic] *hat mir geholfen // mein Flick Flack zu lernen*[11]
- Before coordinating conjunctions introducing a new set of noun/verb phrases, rather than used to list chains of nouns, e.g. *Ich hatte Angst // aber ich hab mir gedacht ich zieh einfach durch*
- At the end of noun phrase used as an address, e.g. *Du dreckiger Bastard // so ein Rassist*

3.1.6 Removing Outliers from the Data Set

This section examines if there are any comments that could be classified as outliers due to their length. This would ensure that only normal data (in a statistical sense) are analysed in the corpus. To measure comment length in this investigation, I have used the CCU, as it more accurately describes the length of comment on YouTube, as explained in section 3.1.5. I have provided a graph in Figure 4 of the distribution of comments based on the number of CCUs to see if there are any comments that seem overly long. As the distribution of CCUs per comment in the YouTube Youth Corpus is heavily positively skewed, i.e. non-parametric (confirmed by a Skewness & Kurtosis test and a Shapiro-Wilk test (for both, $p \leq 0.0001$)), we should not use mathematical means to calculate outliers, such as excluding data points outside the

[11] Although the dependent clause cannot function on its own, it is counted as a separate CCU to illustrate that the comment is more complex (and longer) than just a main clause.

3-sigma (3σ) boundary (Levon 2010: 80).[12] As all comment lengths are contiguous (i.e. there is at least one comment per CCU length, from 1-17), all comments will be used for analysis (excluding the comments removed per section 3.1.1).

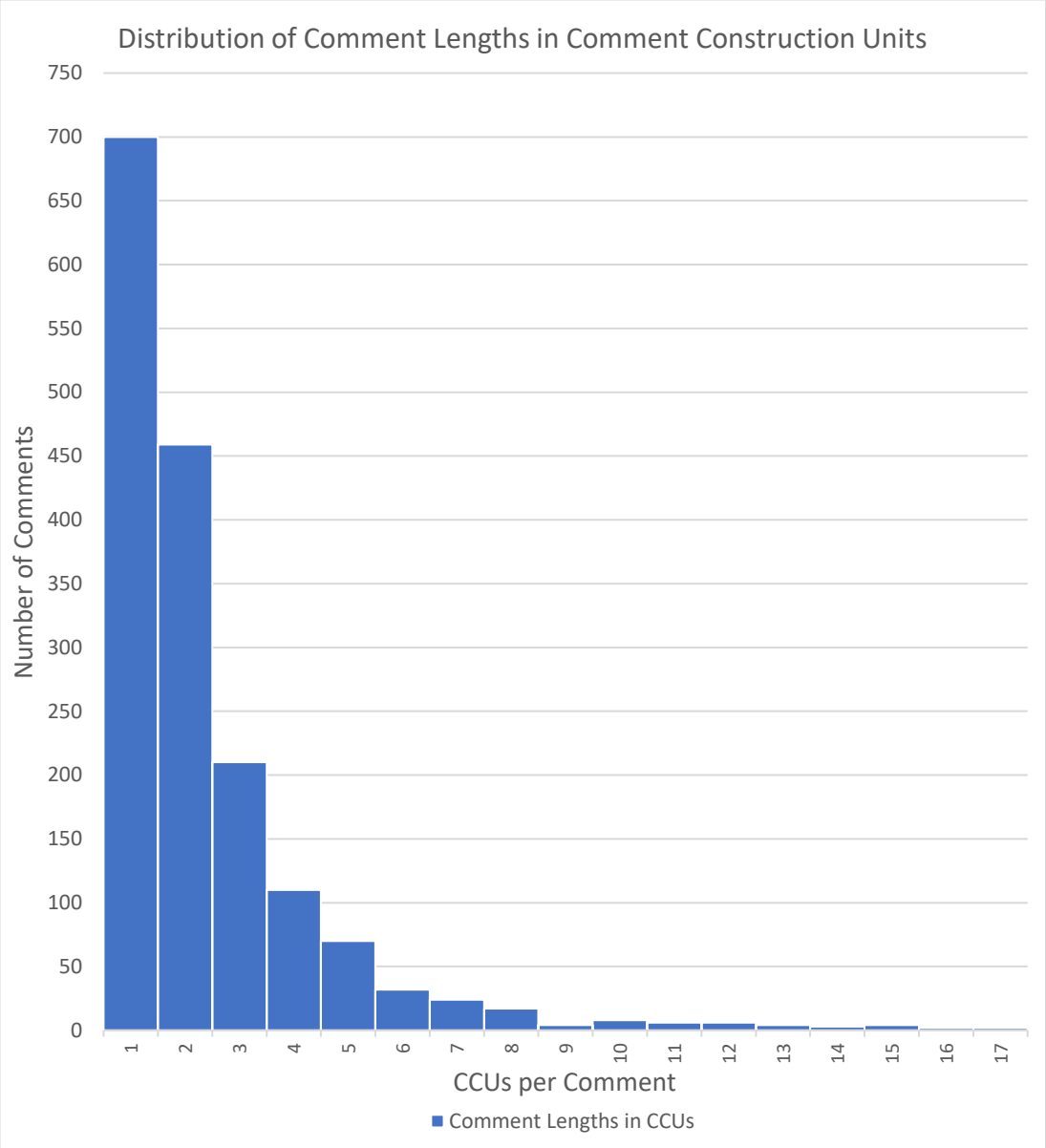


Figure 4 Distribution of comment lengths in CCUs

[12] The 3σ rule, an empirical mathematical rule, posits that in a normal distribution of any data, roughly 99.7% of the data is within three standard deviations of the mean (μ) and can be considered representative of the data as a whole (i.e. a p value of ≤0.01) (Levon 2010: 80). For the YYPC, only 97.9% of the data points fall within 3σ due to the skewness.

Linguistic Features Investigated		
Hypothesis and evidence from previous scholarship		**Linguistic Feature Used to Test Hypothesis**
A	There is no difference in the number of times males and females speak (cf. Gräßel 1991; Cameron 2003; vs. Pusch 1984; Trömel-Plötz 1984a)	1. Mean number of comments produced by commenter
B	Males speak for longer than females/Males produce longer comments than females (cf. Hummel 1984; Herring 2003)	2. Length of a comment (measured in Comment Construction Units) 3. Words per Comment Construction Unit
C	Males produce more one-word comments than females (cf. Baron 2004)	4. Number of comments consisting of one word
D	Females are more positive and supportive in their language than males (cf. Herring 2003)	5. Positive response of comment
E	Males are more adversarial in their language than females (cf. Trömel-Plötz 1979, 1984a; Herring 2003)	6. Adversarial response of comment
F	Males use more swear words and other profane language than females (cf. Herring 2003)	7. Profanity
G	Females have longer conversations than males (cf. Herring 2003; vs. Baron 2004)	8. Length of a conversation
H	Males favour all-male conversations (cf. Herring 2003; vs. Baron 2004)	9. Proportion of same-gender conversations
I	Females use more linguistic structures that weaken the assertion in speech (cf. Trömel-Plötz 1982; Coates 2013)	10. Hedge words and phrases
J	Males use more contractions in language than females (cf. Baron 2004; vs. Herring & Zelenkauskaite 2009)	11. Contractions
K	Females use more emoticons than males (cf. Wolf 2000)	12. Emoji & emoticons
L	Young men and women use a similar amount of neologisms & CMC terminology (cf. Androutsopoulos 2004)	13. Neologisms associated with computer-mediated communication
M	Females use more grapheme prosody than males (cf. Herring & Zelenkauskaite 2009)	14. Grapheme prosody
N	Females use more punctuation prosody than males, specifically repeating exclamation marks (cf. Herring & Zelenkauskaite 2009; Waseleski 2006)	15. Punctuation & emoji prosody
O	Males with a migration background use more language mixing than females with a migration background	16. Use of community language in a comment

Table 6 List of Features Analysed in the Online Language Variation Investigation

3.2 Linguistic Features to be Analysed

From the research reviewed above in Chapter 2, I identified 15 hypotheses on linguistic differences between either genders or migration backgrounds, which form the investigation. These are listed in Table 6, along with the researcher(s) who suggested it (or disputed it, as indicated by 'vs.'), and the linguistic features that are

used to test the hypothesis. Below, I provide examples of the features and explain the parameters for identifying and measuring them in the YYPC.

Number of Comments produced by population group

This feature is measured by counting the number of comments produced by males, females, those with a migration, and non-migration background, or a combination of gender and ethnicity.

Length of a Comment (Measured in Comment Construction Units)

This feature is measured by counting the number of CCUs contained within a comment as explained in section 5, as they provide a better representation of the length of a comment than the number of words. There was an overall average of 2.4 CCUs per comment in the YYPC.

Words per Comment Construction Unit

This feature is measured by dividing the number of words in a comment by the number of CCUs. It provides an insight into the complexity of the CCUs, which can differ across gender and ethnicity and provides us with an insight into the internal structure of a comment.

Number of comments consisting of one word

This feature is measured by counting all comments that consist of one word; excluding emoji and the names of other commenters (such as when replying to a previous comment, cf. Figure 1).

Positive Comment

To qualify as a positive comment, the comment must contain supportive words and/or demonstrate an overall positive tone, either towards the video or a fellow commenter, such as compliments and the use of smiling emoji, e.g. GF1 – 058 *Bibi*

du bist so geil 😍 *und Julian du bist so witzig* 😂 ('Bibi you are so great 😍 and Julian you are so funny 😂'). If there was any doubt or if a commenter expressed conflicting emotions in the same comment, it was not counted. There are 572 (34.4%) positive comments in the corpus.

Adversarial Comment

The same applies to counting comments of an adversarial nature. The comment must directly challenge the video or fellow commenter, use insulting or derogatory language or otherwise be undermining in tone, e.g. MM1 – 041 *Wegen unnötige Videos Deabo* ('Due to unnecessary videos unsubbed [unsubscribed from the channel]'). There are 223 (13.4%) adversarial comments in the YYPC.

Profanity

This feature covers any word that is considered a swear word or highly offensive in German. The most common examples in the data are: *ficken, scheiß/e, Hure/Hurensohn, Arsch, drecks* (and variants thereof)*, Fotze, Nutte,* and *amk*.[13]

Length of Conversation

A conversation in the YYPC is defined as one thread of comments between more than one commenter (as illustrated in Figure 1). Conversations on YouTube are never technically finished as future commenters are able to reply to threads long after they have started, but, as this applied to conversations and participants affecting both genders and migration backgrounds, it should not skew the results of

[13] *Amk* is an acronym short for *amına koyayım*, a Turkish phrase that can be translated as 'f**k your p***y' and is used in the YYPC by commenters from migration backgrounds.

the YYPC. This feature is measured by counting the number of comments in conversations where participants all had the same gender or migration background.

Proportion of Same-Gender Conversations

This feature is measured by finding the percentage of all-male conversations within conversations with a male participant and the percentage of all-female conversations within conversations with a female participant.

Hedge Words and Phrases

This feature is measured by counting the hedge words and phrases adapted from Trömel-Plötz (1982: 45-53) for an online setting, as listed in Table 7.

Hedge Words and Phrases		
eben	man könnte	scheinbar/anscheinend
eben so	meiner Meinung nach	vielleicht
ge?	meinst du nicht auch	wäre schön, wenn
gell?	ne?	wenn du
glaube ich	nicht wahr?	willst/solltest du nicht
ich weiß nicht, ob	nicht?	woll?
ist das nicht	nur so	würde mich freuen
ist es nicht so	oder?	würde sagen

Table 7 List of hedge words and phrases looked at in YYPC

/Contractions

The words that were counted as contractions are taken from Durrell (2003: 3-51) and Fagan (2009: 246-68) and encompass all examples where a word has been shortened, as in the case of *habe* to *hab'* or two or more words combined, such as *mit dem* to *mitm*. Further examples from the YYPC include *drauf* (*darauf*), *ne* (*eine*), *ausn* (*aus den*), and *isso* (*es ist so*).

Emoji & Emoticons

This feature includes any of the Unicode characters and combinations that are displayed as small graphical entities when inserted into a comment, such as 😂 or

😍, or any combination of punctuation and characters that are used to add a paralinguistic effect (except prosodic effects), such as :(to indicate sadness through a sad face, or XD to indicate mirth through a laughing face.

Neologisms associated with Computer-Mediated Communication

Any word, phrase or acronym that has been coined or repurposed to be used in computer-mediated communication is counted under this feature, excluding general vocabulary like *computer*, *mouse*, *monitor* etc., but including names (and derived verbs in the case of *googeln*) of prominent internet companies and paralinguistic features such as hashtags and @-ing. Examples include the following: *klicken, Google, Youtube, glockeaktiviert, like, subscribe, abo/abonnieren, Facebook, event, site/seite, Instagram, hdgdl,*[14] *omg, wtf, #, @, kanal, gg*, and *views*.

Grapheme Prosody

This feature includes all examples where one or more graphemes was repeated for exaggeration or to represent spoken intonation, e.g. GF1 – 013 *Julian ist so duuuuuuuuuummmmmmm* (‘Julian is so duuuuuummmb’). Each word where such a feature occurred was counted as one example of grapheme prosody, but if a word contained more than one prosodic letter, as in the case above, it only counted as one, as it represents one speech act.

Punctuation & Emoji Prosody

This feature contains all examples where punctuation marks, emoji and emoticons are repeated for exaggeration, similar to grapheme prosody i.e. GM3 – 002 *JUNGGESELLENABSCHIED?!????!?!?* (‘STAG DO?!????!?!?🤨’). As with features that

[14] (*Ich*) *habe dich ganz doll lieb*

contain grapheme prosody, each example in the same comment must pertain to different CCUs, but different agents of prosody, e.g. punctuation marks and emoji, are counted separately as they have different paralinguistic effects. An example of such is GF1 – 163 *Stoppt mal bek 6:36 zu geil!!!!*😂😂😂 *dieser Blick!!!!!* ('Stop [the video] at [sic] 6:36 too good!!!!😂😂😂 that look!!!!!😮'), which has three counts of punctuation prosody, two sets of exclamation marks and one set of emoji.

Use of community language in a comment

To measure language mixing (as explained in section 2.5.2), I counted every time a community language, i.e. a language spoken by a significant migrant community, such as Turkish, Arabic, Russian, or languages of the former Yugoslavia (but not English), was used in a CCU (multiple foreign words within the same CCU were counted as 1). I have grouped all the community languages together, as there were not enough data points to examine the languages individually. For example, the following comment has three instances of language mixing (bold, CCUs indicated by '//'): ***inshallah*** *dein* ***baba*** *sieht das video dann kommt* ***terlik*** 😂*-spaß* ('**God willing** // your **father** sees the video // then the **slipper** will come [i.e. he will hit you with the slipper] // 😂 –joke').

3.3 Statistical Methods and Analysis of the Features

Each of the 16 features were analysed across seven sets (six of two groups and one of four groups), listed below:

1. Comparing males and females
2. Comparing those with a non-migration background and those with a migration background

3. Comparing males and females with a non-migration background
4. Comparing males and females with a migration background
5. Comparing males with a non-migration background and males with a migration background
6. Comparing females with a non-migration background and females with a migration background
7. Comparing all four of numbers 3. – 6.

To compare the use of the features across the six pairs, I applied at least one of three tests of statistical significance. Each of the three tests provides a different angle from which to analyse the difference between the feature and group in question:[15]

1. A Chi-square test shows whether there is a statistically significant difference in the overall use of a feature between two groups e.g. *do females use more emoji than males?* (When the value of a variable was less than five, I used the Fischer's exact test in place of a Chi-square test.)
2. A two-sample proportion test shows whether there is a difference in the proportions of two groups who used a feature e.g. *do more females use emoji in comments than males?*
3. A Mann-Whitney U test, the non-parametric equivalent of a t-test, compares two means. This tests whether there is a statistically significant difference in the average frequencies of a feature e.g. *do females use more emoji per comment than males, when an emoji is used?*

[15] The three tests were chosen based on the guidelines for statistical tests by Ott & Longnecker (2015) and Dörnyei (2007).

To test whether there was a statistical significance in the variation of the use of a feature between all four gender and ethnicity groups, I also used a Chi-square test. The results of these tests are presented in Chapter 4.

4 Results and Discussion

This chapter is divided into three parts, each detailing a different aspect of the results. The linguistic differences between genders are analysed in section 4.1, the linguistic differences across ethnicity in section 4.2, and the results of language variation from an intersectional perspective is section 4.3. For each of the sub-sections, the relevant statistical tables are included at the end for reference.

4.1 General Differences across Gender

4.1.1 Comments and Comment Length and Gender (Hypotheses A-C)

The YYPC data concerning the number of comments produced by each gender (Hypothesis A) and that males use more one-word comments than females (Hypothesis C) support the earlier findings presented in 3.2 and contradict previous findings on the comparative length of the comments (Hypothesis B). The findings of the YYPC are explained below for each of the corresponding hypotheses, presented in Table 8.

Hypothesis A: There is no difference in the number of times males and females speak

In the YYPC, males wrote more comments on average than females (1.16 vs 1.09 comments, $p \leq 0.05$, Mann Whitney U test). This supports the findings of Pusch (1984) and Trömel-Plötz (1984a), although the weakly significant p-value could indicate that a larger study might support Gräßel (1991) and Cameron (2003), who found there was no significant difference between how often males and females spoke.

Hypothesis B: Since evidence suggests males speak for longer than females, males will, on average, produce longer comments than females

Hypothesis B was tested by comparing the mean length of a comment in CCUs and the mean number of words per CCU. The mean number of CCUs in a comment did not differ significantly between genders (2.5 for females versus 2.3 for males, Mann Whitney U Test). The average female CCU contained significantly more (0.8 more) words than males' ($p \leq 0.0001$, Mann Whitney U Test). One possible explanation for this is the unembellished nature of male writing, as suggested by Herring & Zelenkauskaite (2009: 25-26), particularly in qualifying adjectives and adverbs. To explore this, I examined how often males and females qualified *geil*, a popular adjective used in youth language, with words such as *so*, *mega*, and *zu*. Of 30 uses of *geil* by females, 57% (n=17) were modified compared to 28% of males (7 of 25), which is weakly statistically significant ($p \leq 0.05$, two-sample proportion test).[16] This suggests that females do indeed embellish their comments with qualifying adjectives and adverbs. Further research in this regard is needed, however.

Hypothesis C: Males produce more one-word comments than females

10.4% (62 of 594) of comments by males consisted of one-word compared to 6.0% (64 of 1063) by females, statistically significantly higher ($p \leq 0.01$, Chi-square test). This supports the results of previous research and is consistent with the above suggestion that males' writing is less embellished than females'. Typical one-word comments from females consisted of greetings, such as *hi* (cf. GF1 – 064) or adjectives such as *nice* (cf. GF2 – 081). Males also used such adjectives in one-word

[16] From all the instances of *geil*, only three comments used the word more than once in a comment (two males and one female).

comments but also favoured the use of the hashtag *#glockeaktiv* (14 instances to females' 5) to indicate that they had subscribed to the YouTube channel after watching the video.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
A: Females write more comments than males	1.09 comments/ commenter	1.16 comments/ commenter	≤0.05	Yes	Mann Whitney U test
B: The mean comment written by a female contains more CCUs than by a male	2.5 CCUs/ comment	2.3 CCUs/ comment	N/A	No	Mann Whitney U test
B: Females write more words per CCU on average than males	4.7 words/CCU	3.9 words/CCU	≤0.0001	Yes	Mann Whitney U test
B: Females qualify *geil* more than males	58% (n=17)	28% (n=7)	≤0.05	Yes	Two-sample proportion test
C: Males use more one-word comments than females	64 of 1063 (6%) comments	62 of 594 (10%) comments	≤0.01	Yes	Chi-square

Table 8 Results from Hypotheses A-C: comments and comment length

4.1.2 Positive or Adversarial Nature of Comments & Profanity across Gender (Hypotheses D-F)

This section explains how the YYPC data supports Hypotheses D, E, and F, which investigate the positive nature of females' comments, the adversarial nature of males' comments, and the use of profanity by gender, respectively. The results are presented in Table 9.

Hypothesis D: Females are more positive in their language than males

37.1% (n=396) of comments made by females were positive, compared with 29.6% (n=176) of males. This supports the hypothesis (p≤0.05, Chi-square) adapted from Herring (2003).

Hypothesis E: Males are more adversarial in their language than females

Hypothesis E was also adapted from Herring (2003) as well as from Trömel-Plötz (1979, 1984a), and the YYPC data fully support their findings (p≤0.0001, chi-

square):22.7% (n=135) of the 595 comments made by males take an adversarial stance to either another commenter or the video, compared to only 8.4% (n=90) of comments made by females.

Hypothesis F: Males use more profanity and swear words than females

The results from the YYPC strongly support Hypothesis F. Compared to only 1.9% of females (n=20), 8.1% of males (n=48) used profanity in their comments (p≤0.0001, two-sample proportion test) and males' language contained over six times as much profanity per 1,000 words (10.65 words/1000 vs 1.69 words/1000, p≤0.0001, Mann Whitney U test). These results suggest that the findings by Lakoff (1975: 47) that males are not as 'ostracised' as females for using rough language applies online, as well as in spoken language.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
D: Women leave more positive comments than males	n=396 (37.1%)	n=176 (29.6%)	≤0.05	Yes	Chi-square
E: Males leave more adversarial comments than females	n=90 (8.4%)	n=135 (22.7%)	≤0.0001	Yes	Chi-Square
F: Males use more profanity than females	n=22 (0.17%)	n=60 (1.1%)	≤0.0001	Yes	Chi-Square
F: More males use profanity than females	n=20 (1.9%)	n=48 (8.1%)	≤0.0001	Yes	Two-sample proportion test
F: Males' language contains more profanity per 1,000 words	1.69	10.65	≤0.0001	Yes	Mann Whitney U test

Table 9 Results from Hypotheses D-F: positive and adversarial comments, profanity

4.1.3 Conversation Length and Interaction across Gender (Hypotheses G and H)

The YYPC data on the gender differences between conversation length (Hypothesis G), and the proportion of same-gender conversations (Hypothesis H) contradict the findings in section 3.2. The YYPC data is explained below and presented in Table 10.

Hypothesis G: Females have longer conversations than males

All-male conversations, as defined in section 3.2 have an average of 6.97 comments and all-female conversations have an average of 2.62 comments per conversation. This does not support Hypothesis G, rather suggests that the opposite is true ($p \leq 0.01$, Mann-Whitney U test). One possible explanation is the adversarial nature of males' comments and replies, which encourages further responses, such as in Conversation 92 (in Appendix), where a flippant response to a comment sparked a 42-comment angry exchange.

Hypothesis H: Males have more same-gender conversations than females

The YYPC does not support Hypothesis H, finding no statistically significant difference between the proportion of all-male (n=33, 46.5%) and all-female conversations (n=45, 54.2%) to the total number of conversations with a participant of that gender. A closer look at the conversations did not reveal any reasons for the results. Perhaps a larger data set with more conversational data is needed.

Findings based on Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
G: All-female conversations are longer on average than all-male conversations	2.62 comments/ conversation	6.97 comments/ conversation	≤0.01 (the other way)	Yes	Mann Whitney U test
H: Males have more same-gender conversations than females	54.2% (n=45)	46.5% (n=33)	N/A	No	Two-sample proportion test

Table 10 Results from Hypotheses G and H: conversational interaction and length

4.1.4 Females use more linguistic structures that weaken the assertion in speech (Hypothesis I)

The YYPC data does not support Hypothesis I, instead demonstrating that very similar proportions of males (3.8%, n=23) and females (3.6%, n=38) use hedge

words, as shown in Table 11. Contrary to the findings of Milroy & Margrain (1980) and Trömel-Plötz (1982: 48-49), males (4.3 words/1000) use slightly more hedge words than females (2.9 words/1000) but this difference is not statistically significant. Hedge words and phrases were not a common feature in the YYPC, and perhaps this is a feature of conceptual orality that is not transferred to YouTube comments.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
I: Females use more hedge words than males	n=38 (0.3%)	n=24 (0.4%)	n/a	No	Chi-Square
I: More females use hedge words than males	n=38 (3.6%)	n=23 (3.8%)	n/a	No	Two-sample proportion test
I: Females' language contains more hedge words per 1,000 words than males	2.9 words/ 1000	4.3 words/ 1000	n/a	No	Mann Whitney U test

Table 11 Results from Hypothesis I: hedge words

4.1.5 Males use more contractions in language than females (Hypothesis J)

Baron (2004: 415) found that more males used contractions, whilst Herring & Zelenkauskaite (2009: 26) found that more females used contractions. The results of the YYPC do not support either result, instead showing no statistically significant differences between the two groups, as illustrated in Table 12, although the data shows that slightly more females in the YYPC used this feature. However, as noted in 2.1, the use of such spoken language features is common to online language, regardless of gender.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
J: Male use more contractions than females	n=120 (0.9%)	n=49 (0.8%)	N/A	No	Chi-Square
J: More males use contractions than females	n=101 (9.5%)	n=45 (7.6%)	N/A	No	Two-sample proportion test
J: Males' language contains more contractions per 1,000 words than females'	9.2 words/ 1000	8.7 words/ 1000	N/A	No	Mann Whitney U test

Table 12 Results from Hypothesis J: Contractions

4.1.6 Orthographical and lexical features of online communication across gender (Hypotheses K-N)

This section examines four hypotheses on the use of specifically online lexical and orthographical linguistic features.

Hypothesis K: Females use more emoticons/emoji than males

The YYPC confirms the hypothesis that more females (49.2%, n=524) use more emoticons/emoji than males (28.6%, n=170) ($p \leq 0.0001$, two-sample proportion test), which is presented in Table 13. Females (3.1 emoji/comment) also use more emoji in a comment than males (2.8 emoji/comment) ($p \leq 0.0001$, Mann Whitney U test), so it is unsurprising that a higher proportion of characters (letters, spaces, punctuation etc.) typed by females is comprised of emoji/emoticons (2.1%, n=1623 vs 1.4%, n=482 for males) ($p \leq 0.0001$, Chi-square). The use of emoji has also increased over time. Compared to Wolf (2000: 831), where there were 30.6 emoticons per 100 comments, the figure in the YYPC is 127 emoji/emoticons per 100 comments,[17] a statistically significant increase of over a third ($p \leq 0.0001$, two-sample proportion test). Reasons for such changed linguistic behaviour most likely include rising ownership of devices that use online communication, ease of access (often smartphones suggest emoji to the user),[18] as well as the proliferation of emoji throughout popular culture; an emoji has been included as the Oxford

[17] Unfortunately, Wolf only included total emoticon numbers and did not count if a comment contains multiple emoticons. Assuming each comment contains one emoticon, the YYPC contains a higher proportion of comments that contain emoji/emoticons (30.6% vs 41.7%) ($p \leq 0.001$, two-sample proportion test).

[18] MacRumors. 'iOS 10: Available Now', ≤https://www.macrumors.com/roundup/ios-10/> [accessed 01.08.2017]

Dictionary Word of the Year 2015,[19] and emoji are the central premise in a 2017 $50m Sony film.[20]

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
Females use more emoji than males	n=1623 (2.1% of all characters)	n=482 (1.4% of all characters)	≤0.0001	Yes	Chi-Square
More females use emoji than males	n=524 (49.2%)	n=170 (28.6%)	≤0.0001	Yes	Two-sample proportion test
Women use more emoji in a comment than males	3.10 emoji/ comment	2.84 emoji/ comment	≤0.0001	Yes	Mann Whitney U test
More emoji are used in 2017 than in 2000	2000 data: 30.6 emoji/comment 2017 data: 127 emoji/comment		≤0.0001	Yes	Two-sample proportion test
There is proportionally more emoji use by commenter in 2017 than in 2000	2000 data: 30.6% (77/251) 2017 data: 41.8% (694/1658)		≤0.001	Yes	Two-sample proportion test
Males use more emoji in all-male conversations than mixed conversations	Mixed=0.08 emoji/comment All-male=0.47 emoji/comment		≤0.01	Yes	Mann Whitney U test
More males use the Face With Tears of Joy emoji than females	33% (n=173)	45% (n=77)	≤0.05	Yes	Two-sample proportion test

Table 13 Results from Hypothesis K: emoji/emoticon use

A further finding of Wolf (2000: 832) was that males used more emoji in a mixed environment than in all-male company. The YYPC data suggests the reverse: males used emoji six times more in all-male conversations than in mixed settings (0.47 emoji/comment vs 0.08 emoji/comment) (p≤0.01, Mann Whitney U test). This may indicate a change in how emoji are perceived within the frame of gender performativity; the wide range of emoji available means males, who previously were expected to avoid expressing the emotional context through paralinguistic

[19] Oxford University Press. 'Oxford Dictionaries Word of the Year 2015 is…', Oxford University Press, ≤http://blog.oxforddictionaries.com/2015/11/word-of-the-year-2015-emoji/> [accessed 11.06.2017]

[20] British Board of Film Classification. 'The Emoji Movie', ≤http://www.bbfc.co.uk/releases/emoji-movie-2017> [accessed 01.08.2017]

means, now have an outlet, as is demonstrated by the use of the 'Face With Tears of Joy' (😂) emoji. Males use this more often than females (45%, n=77 vs 33%, n=173) (p≤0.05, two-sample proportion test), who instead appear to prefer a wider range, such as various heart emoji, and the 'See-No-Evil Monkey' (🙈). With more emoji approved yearly by the Unicode Consortium, including elves, merpeople, and redheads, we can expect to see a diversification of emoji use, although it remains unclear if males and females will use similar amounts.

Hypothesis L: Males and females use similar amounts of CMC terminology and neologisms

The YYPC data suggest that more males (13.1%, n=78) use CMC vocabulary than females (10%, n=107), which does not support Hypothesis L, although the correlation is weak (p≤0.05, two-sample proportion test). The YYPC data, also presented in Table 14, shows that males use more CMC vocabulary than females (19.3 words/1000 vs 11.8 words/1000), but again the correlation is weak (p≤0.05, Mann Whitney U test). Perhaps CMC vocabulary use differs across internet media, and a corpus comprised of Facebook or Twitter comments would show different results (particularly the use of hashtags on Twitter (cf. Cunha et al. 2012)), but within this sample of YouTube comments, there are no significant differences.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
Men use more CMC than females	n=153 (0.12%)	n=109 (0.19%)	≤0.05	Yes	Chi-Square
More males use CMC than females	n=107 (10%)	n=78 (13.1%)	≤0.05	Yes	Two-sample proportion test
Men's language contains more CMC per 1,000 words than females'	11.8 words/ comment	19.4 words/ comment	≤0.05	Yes	Mann Whitney U test

Table 14 Results from Hypothesis L: CMC terminology

Hypothesis M: Females use more grapheme prosody than males

The YYPC data does not support Hypothesis M; there is no significant difference in the proportions of male or female commenters who use grapheme prosody, nor with respect to instances per 1,000 words nor average use of grapheme prosody per comment, as shown in Table 15. This is perhaps surprising as all previous research found that females used more grapheme prosody than males, such as in Herring & Zelenkauskaite (2009: 23), where females used grapheme prosody roughly three times as frequently as males. In the YYPC, there are gender differences as to what kind of grapheme cluster is repeated. For example, females repeat the *o* of *so* more than three times as frequently as males (36% (n=35) vs 10% (n=5)) ($p \leq 0.01$, two-sample proportion test) (e.g. MM1 – 070 *Deine Freundin ist soooooooo hübsch*), while males tend to exaggerate their laughter, using *haha*, twice as often (57% (n=27) vs 30% (n=29) ($p \leq 0.01$, two-sample proportion test) (e.g. MM3 – 132 *Ich kann nicht mehr ahhahahhahahahahhahahhahahahha*). Herring & Zelenkauskaite (2009: 22) did not measure for *haha*, instead only counting where extra single graphemes had been used. If we are to replicate this previous study and exclude the prosodic instances of *haha* from both populations, females' use of grapheme prosody becomes significantly more frequent than males' ($p \leq 0.01$, Chi-square), which tallies with previous research. However, I would argue that the exaggerative nature that grapheme prosody intends to convey warrants the inclusion of such grapheme clusters in the investigation.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
Females' language contains more grapheme prosody per 1,000 words than males'	8.4 words/ 1000	7.5 words/ 1000	N/A	No	Mann Whitney U test
More females use grapheme prosody than males	n=86 (8.1% of females)	n=42 (7.1% of males)	N/A	No	Two-sample proportion test
Females use more grapheme prosody on average in a comment than males	1.14 times/ comment	1.12 times/ comment	N/A	No	Mann Whitney U test
More males use *haha than females**	30% of females (n=26)	57% of males (n=24)	≤0.01	Yes	Two-sample proportion test
More females use *soo than males**	36% (n=31)	10% (n=4)	≤0.01	Yes	Two-sample proportion test
Females use more grapheme prosody than males (excl. *haha*)	6.6% of females (n=70)	3% of males (n=18)	≤0.01	Yes	Chi-Square

Table 15 Results from Hypothesis M: grapheme prosody

Hypothesis N: Females use more punctuation prosody (specifically repeating exclamation marks)

The YYPC data strongly confirms Hypothesis N. More females (26.5%, n=283) use punctuation prosody than males (19.4%, n=115) (p≤0.001, two-sample proportion test) and females also use more punctuation prosody overall (n=396, 37.1% of comments vs n=136, 22.9%) (p≤0.0001, Chi-square), as shown in Table 16. However, this includes all punctuation marks, emoji and emoticon prosody. The original research by Herring & Zelenkauskaite (2009) and Waseleski (2006) only found a significant difference in the use of exclamation marks, whereas in the YYPC this is not the case: both males (10.9%, n=16) and females (13.9%, n=31) produced similar proportions of exclamation mark prosody. This may indicate a shift in how males and females signal exaggeration: instead of repeating punctuation, they are using emoji prosody. The YYPC seems to suggest so: 58.5% (n=233) of all punctuation prosody instances were repetitions of emoji. The use and repetition of emoji is an

important (and regular) feature of today's online language, employed to the same effect as regular punctuation.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
More females use punctuation prosody than males	26.5% of females (n=283)	19.3% of males (n=115)	≤0.001	Yes	Two-sample proportion test
Females use more punctuation prosody in a comment than males	1.40 times/ comment	1.18 times/ comment	≤0.001	Yes	Mann Whitney U test
Females use more exclamation mark prosody than males	n=31 (13.9%)	n=16 (10.9%)	N/A	No	Chi-Square

Table 16 Results from Hypothesis N: punctuation and emoji prosody

4.2 The Effect of Ethnicity on Language Differences across Gender

I now turn to the results concerning language use and ethnicity, beginning with language mixing , before examining the variables discussed above for gender, and whether the overall differences also pertain to both those with a non-migration (i.e. ethnic German) background, and to those with a migration background. I have assigned an abbreviation to each of the four populations for expediency when referring to them, as shown in Table 17. If there is no mention of a relationship between two subsets for a particular variable, there was no statistically significant difference in the YYPC. The conversation interaction variable (as defined in section 3.2) could not be analysed across ethnicity as there were insufficient data points for conversations with commenters from only one ethnic group (non-migration background n=7, migration background n=3). Commenters partook in mostly mixed conversations or it was not possible to identify the ethnicity of some of the participants (only 56.1% could be assigned an ethnicity).

Population Subset	Acronym
Commenters with a non-migration background	GB (German Background)
Commenters with a migration background	MB (Migration Background)
Men with a non-migration background	GM (German Male)
Women with a non-migration background	GF (German Female)
Men with a migration background	MM (Migration Male)
Women with a migration background	MF (Migration Female)

Table 17 Acronyms for YYPC population subsets

4.2.1 Language Mixing (Hypothesis O)

The YYPC data support the hypothesis that males with a migration background (MM) (6.8% of CCUs, n=36) use more language mixing than females with a migration background (MF) (1.1% of CCUs, n=4) ($p \leq 0.0001$, Fisher's exact test) and also that a larger proportion of MM (11.2% of comments, n=28) use language mixing than MF (1.9% of comments, n=3) ($p \leq 0.0001$, two-sample proportion test). These results are presented in Table 18. Perhaps surprisingly for a corpus dealing with young people and videos that specifically dealt with ethnicity-related topics (such as video 9: *Türken vs. Afrikaner 2*), there were no instances of crossing (cf. Cutler 1999; Rampton 2014), i.e. language mixing by those with a non-migration background.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical Test
O: MM use more language mixing than MF	1.1% of CCUs (n=4)	6.8% of CCUs (n=36)	≤0.0001	Yes	Fisher's Exact
O: More MM use language mixing than MF	1.9% of comments (n=3)	11.2% of comments (n=28)	≤0.0001	Yes	Two-sample proportion test

Table 18 Results from Hypothesis O: language mixing

4.2.2 Comments and comment length across ethnicity (Intersectional Analysis of Hypotheses A and B)

The gender split for commenters of a non-migration background mostly tallied with the general results; there was no significant difference between the amount of comments between migration backgrounds (GB 1.12 vs MB 1.17

comments/commenters) and between females and males from either migration background (GF 1.11 vs GM 1.14 comments/commenters, MF 1.18 vs 1.15 comments/commenters). There was also no significant gender difference between the number of CCUs per comment, and GF wrote more words per CCU than GM ($p \leq 0.05$, Mann Whitney U test) as shown in Table 19. For those with a migration background, females from a migration background also wrote more words per CCU than males ($p \leq 0.05$, Mann Whitney U test). There was no difference in the mean length of a comment in CCUs between those with a non-migration background (GB) and those with a migration background (MB) (regardless of gender), although GB (4.6 words/CCU) did write more words per CCU than MB (4.0 words/CCU) ($p \leq 0.01$, Mann Whitney U test), as shown in

Findings based on Hypothesis	GB	MB	p-value	Statistically Significant?	Statistical test
A: MB write more comments than GB	1.12 comments/ commenter	1.17 comments/ commenter	N/A	No	Mann Whitney U test
B: The mean length of a comment by GB contains more CCUs than by MB	2.5 CCUs/ comment	2.2 CCUs/ comment	N/A	No	Mann Whitney U test
B: GB write more words per CCU on average than MB	4.6 words/ CCU	4.0 words/ CCU	≤0.01	Yes	Mann Whitney U test

Table 20. According to the YYPC, both gender and migration background impact the number of words used in a CCU. Further research within these intersections is needed to establish the reasons for this difference.

Findings based on Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
A: GM write more comments than GF	1.11 comments/ commenter	1.14 comments/ commenter	N/A	No	Mann Whitney U test
A: MM write more comments than MF	1.15 comments/ commenter	1.18 comments/ commenter	N/A	No	Mann Whitney U test
B: The mean comment written by GF contains more CCUs than by GM	2.5 CCUs/ Comment	2.5 CCUs/ Comment	N/A	No	Mann Whitney U test
B: GF write more words per CCU on average than GM	4.7 words/ CCU	4.1 words/ CCU	≤0.05	Yes	Mann Whitney U test
B: MF write more words per CCU on average than MM	4.3 words/ CCU	3.9 words/ CCU	≤0.05	Yes	Mann Whitney U test

Table 19 Results from Hypotheses A and B (Intersectional): comment numbers and length

Findings based on Hypothesis	GB	MB	p-value	Statistically Significant?	Statistical test
A: MB write more comments than GB	1.12 comments/ commenter	1.17 comments/ commenter	N/A	No	Mann Whitney U test
B: The mean length of a comment by GB contains more CCUs than by MB	2.5 CCUs/ comment	2.2 CCUs/ comment	N/A	No	Mann Whitney U test
B: GB write more words per CCU on average than MB	4.6 words/ CCU	4.0 words/ CCU	≤0.01	Yes	Mann Whitney U test

Table 20 Results from Hypotheses A and B (Ethnicity): number of comments and comment length

4.2.3 Positive or Adversarial Nature of Comments & Profanity across Ethnicity (Intersectional Analysis of Hypotheses D-F)

In the YYPC, the adversarial nature of the comments did not vary according to the ethnic background of the commenter; males of either background left proportionately far more comments that were adversarial in nature (both p≤0.0001, Chi-square), as shown in Table 21. However, for both ethnic groups, there was no difference in the proportion of positive comments, which weakens the result of the

general hypothesis that females leave more positive comments than males. The use of profanity also did not vary based on ethnic background: more males from both groups overwhelmingly used profanity in their comments than females. Indeed, there was only one comment in the whole corpus where a female from a migration background used profanity, which was a reply to a (rhetorical) question as to what kind of person would keep a comb in their hair - her response, *deine mutter in ihrer muschi amk HAHAHAH* ('your mother in her [vagina] amk[21] HAHAHAH') (MM3 – 135). Such a comment is uncharacteristic from a female with a migration background.

Findings based on Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
D: GF leave more positive comments than GM	38.5% of comments (n=157)	40.3% of comments (n=50)	N/A	No	Chi-Square
D: MF leave more positive comments than MM	37.6% of comments (n=59)	24.9% of comments (n=62)	N/A	No	Chi-square
E: GM leave more adversarial comments than GF	10.3% of comments (n=42)	35.5% of comments (n=44)	≤0.0001	Yes	Chi-Square
E: MM leave more adversarial comments than MF	5.1% of comments (n=8)	19.3% of comments (n=48)	≤0.0001	Yes	Chi-square
F: More GM use profanity than GF	2.0% of comments (n=8)	13.7% of comments (n=17)	≤0.0001	Yes	Two-sample proportion test
F: More MM use profanity than MF	0.6% of comments (n=1)	7.6% of comments (n=19)	≤0.001	Yes	Two-sample proportion test

Table 21 Results from Hypotheses D-F (Intersectional): positive and adversarial comments, profanity

[21] Turkish swear word, for a translation, see section 3.2, under 'Profanity'.

4.2.4 Females use linguistic structures that weaken the assertion in speech across ethnicity (Intersectional Analysis of Hypothesis I)

The YYPC data does not support the findings of Hypothesis I, when subdivided across ethnicity, as demonstrated in Table 22. For those with a migration background there are no significant differences and for those with a non-migration background, it is males who use more hedge words and phrases, although with weak significance (p≤0.05, Chi-square), rather than the females, as suggested in previous research. As with the more general data, hedge words are not a frequent feature of the online language of the YYPC.

Findings based on Hypothesis	Females	Males	p-value	Statistically Significant?	Statistical test
I: GM use more hedge words than GF	0.03% comments (n=12)	0.12% comments (n=14)	p≤0.05	Yes	Chi-Square
I: MM use more hedge words than MF	0.04% comments (n=6)	0.02% comments (n=5)	N/A	No	Chi-square

Table 22 Results from Hypothesis I (Intersectional): hedge words

4.2.5 Contractions and Abbreviations across Ethnicity (Hypothesis J revisited for intersectional analysis)

There are no differences across gender for Hypothesis J, which corresponds to the overall results. However, more males with a non-migration background (14.5%, n=18) use contractions than males with a migration background (5.6%, n=14) and they use more contractions per 1,000 words (11.8 contractions/1000 words vs 8.3 contractions/1000 words), although both test results show a weak significant difference (p≤0.05, two-sample proportion and Mann Whitney U test), as shown in Table 23. One possible explanation for the difference is that males with a non-migration background use contractions in spoken language more often and this is

transferred to their online language, especially since a common feature of *Kiezdeutsch* is to drop articles and prepositions (Wiese 2009: 792).

Findings based on Hypothesis	GB	MB	p-value	Statistically Significant?	Statistical test
J: More GM use contractions than MM	14.5% of commenters (n=18)	5.6% of commenters (n=18)	≤0.05	Yes	Two-sample proportion test
J: GM use more contractions per 1,000 words than MM	11.8 words/ 1000	8.3 words/ 1000	≤0.05	Yes	Mann Whitney U test

Table 23 Results from Hypothesis I (Intersectional): contractions

4.2.6 Orthographical features of online communication across ethnicity (Intersectional Analysis of Hypotheses K-N)

This section analyses the four orthographical features and the corresponding differences across gender and ethnicity, as discussed in section 4.1.6, which are emoji, CMC terminology, grapheme prosody, and punctuation & emoji prosody.

Hypothesis K: Females use more emoji/emoticons than males – Intersectional Analysis

Emoji use differentiated by background corresponds with the overall results. For those with a non-migration background, females (61.7% emoji/CCU, n=636) use more emoji per CCU than males (45.9% emoji/CCU, n=141) (p≤0.0001, Chi-square), and a higher proportion of females use emoji (44.4%, n=183 vs 33.9%, n=42) (p≤0.05, two-sample proportion test), as shown in

Table 24. For those from a migration background the results are mixed;

Findings based on Hypothesis	GB	MB	p-value	Statistically Significant?	Statistical test
K: GB use more emoji than MB per CCU	58.1% emoji/CCU (n=777)	43.4% emoji/CCU (n=388)	≤0.0001	Yes	Chi-Square
K: More GB use emoji than MB	42.1% (n=224)	42.1% (n=224)	≤0.05	Yes	Two-sample proportion test
K: GF use more emoji than other subsets	GF: 125% of expected use (n=636) GM: 76.1% of expected use (n=141) MF: 90.8% of expected use (n=177) MM: 76.2% of expected use (n=211)		≤0.0001	Yes	Chi-Square

proportionally more females (46.5% of comments, n=73) than males (24.5% of comments, n=61) with a migration background use emoji (p≤0.0001, two-sample proportion test). However, if males with a migration background use emoji in a comment, they use more than females (3.4 vs 2.4 emoji, p≤0.001, Mann Whitney U test), which also goes against both the overall trend of the YouTube corpus data and previous research.

A higher proportion of those with a non-migration background use emoji (42.1% vs 35.3%, ≤0.05, two-sample proportion test) as well as using more emoji per CCU (58.1% vs 43.4%, p≤0.0001, Chi-square), as displayed in Table 25. If we compare emoji use amongst all four population groups, females with a non-migration background use significantly more emoji than the other three subsets (p≤0.0001, Chi-square). Emoji can therefore be said to be characteristic of a typical online comment from a female with a non-migration background.

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
K: GF uses more emoji than GM per CCU	61.7% emoji/CCU (n=636)	45.9% emoji/CCU (n=141)	≤0.0001	Yes	Chi-Square
K: More GF use emoji than GM	44.4% of comments (n=183)	33.9% of comments (n=42)	≤0.05	Yes	Two-sample proportion test
K: More MF use emoji than MM	46.5% of comments (n=73)	24.5% of comments (n=61)	≤0.0001	Yes	Two-sample proportion test
K: MM use more emoji in a comment than MF	2.4 emoji/ comment	3.4 emoji/ comment	≤0.001	Yes	Mann Whitney U test

Table 24 Results from Hypothesis K (Intersectional): emoji/emoticons

Findings based on Hypothesis	GB	MB	p-value	Statistically Significant?	Statistical test
K: GB use more emoji than MB per CCU	58.1% emoji/CCU (n=777)	43.4% emoji/CCU (n=388)	≤0.0001	Yes	Chi-Square
K: More GB use emoji than MB	42.1% (n=224)	42.1% (n=224)	≤0.05	Yes	Two-sample proportion test
K: GF use more emoji than other subsets	GF: 125% of expected use (n=636) GM: 76.1% of expected use (n=141) MF: 90.8% of expected use (n=177) MM: 76.2% of expected use (n=211)		≤0.0001	Yes	Chi-Square

Table 25 Results from Hypothesis K (Ethnicity): emoji/emoticons

Hypothesis L: Males and females use similar amounts of CMC terminology and neologisms – Intersectional Analysis

The use of CMC terminology, when broken down by ethnicity, does not demonstrate any statistically significant differences between genders, although, like emoji, it is more frequent amongst those with a non-migration background (169% of expected use, n=91 vs 53.8% of expected use, n=43, p≤0.0001, Chi-square), as shown in Table 26. Males with a non-migration background demonstrated the highest use proportionally of this feature within the four subsets but without statistical significance.

Hypothesis	GB	GB	p-value	Statistically Significant?	Statistical test
L: GB uses more CMC terminology than MB	169% of expected use (n=91)	53.8% of expected use (n=43)	≤0.0001	Yes	Chi-Square

Table 26 Results from Hypothesis L (Ethnicity): CMC terminology

Hypothesis M: Females use more grapheme prosody than males – Intersectional Analysis

The differences in the use of grapheme prosody between males and females of either migration background corresponded to the general findings, i.e. there are no significant differences. This is also true of ethnicity; there are no significant differences in the use of grapheme prosody between those of a non-migration and migration background.

Hypothesis N: Females use more punctuation and emoji prosody than males – Intersectional Analysis

The differences in the use of punctuation prosody between males and females of either migration background mostly corresponded to the general findings, and are presented in Table 27. Although similar proportions of males and females with a non-migration background use punctuation prosody, females use more of it (109% of expected vs 71.4% of expected, p≤0.05, Chi-square), whereas the reverse is true for those with a migration background. Males and females use a similar amount of punctuation prosody but a higher proportion of females use it (27.4%, n=43 vs 17.3%, n=43, p≤0.01, two-sample proportion test).

Hypothesis	Female	Male	p-value	Statistically Significant?	Statistical test
N: GF use more punctuation prosody than GM	109% of expected use (n=152)	71.4% of expected use (n=30)	≤0.05	Yes	Chi-square
N: More MF use punctuation prosody than MM	27.4% of comments (n=43)	17.3% of comments (n=43)	≤0.01	Yes	Two-sample proportion test

Table 27 Results from Hypothesis N (Intersectional): Punctuation and Emoji Prosody

4.3 Intersectionality

As mentioned in the introduction, one of the objectives of this study was to investigate any differences in online language use at the intersection of gender and ethnicity. Do females with a migration background speak more like females with a non-migration background than males with a migration background speak like males with a non-migration background? If we run the statistical tests from section 4.1 on the two migration background groups separately, we see that the language between males of different backgrounds is statistically different in five tests out of 36, whereas the language between females of different backgrounds differs in one test of 36. This suggests that, in an online setting, the social power dynamic that most affects language production is gender, rather than ethnicity.

The results in sections 4.1 and 4.2 tell us which features are characteristic of each of the four subsets. For females with a migration background, there does not seem to be any one feature that defines the language by contrast to the other groups. Females with a non-migration background use many emoji and much grapheme prosody, males with a non-migration background are most adversarial and use most CMC terminology, and males with a migration background use language mixing and more profanity. All of this indicates that the language of females with a migration background is closest to the standard of the four subsets, which would suggest that this group is most affected by linguistic gatekeeping, even in the pseudo-anonymised online space of their peers.[22]

[22] A previous investigation (Cotgrove 2017; unpublished) found that longer YouTube comments (in CCUs) also tended to reflect standard language, so one might hypothesize that these females'

Having established the tendencies for each of the subsets, I have selected a 'typical' comment for each of the four groups from the YYPC:

Female with a Non-Migration Background

Julie Abt *Jaaaaaa ich liebe Zartbitter Schokolade* 😹😹😍

('Yeaaaaah I love dark chocolate 😹😹😍')

This typical comment from a young woman with a German background is two CCUs long, just below the average of 2.5, and contains grapheme and emoji prosody, as well as several different emoji. It is positive in sentiment and does not contain bad language.

Male with a Non-Migration Background

PriismVenom[OG] *Heul jz mal nicht wtf man merkt echt das du unter 13 bist..*

('Don't cry now wtf you can really tell that [sic] you are under 13')

This comment typifies the style of a male with a non-migration background. Although it is four CCUs long, above the 2.5 average, it only has three words per CCU, which is lower than the 4.8 average of a woman with the same background. It

---

standard orientation is explained by their tendency to write longer comments. Cotgrove (2017) marked up a smaller corpus with 14 linguistic features of online language that represented conceptual orality (as explained in section 2.1) to see if their use changed based on the length of a comment (in CCUs). The investigation found that the longer the comment, the fewer spoken features were used, such as emoji, prosody, and colloquial speech. This demonstrated a continuum of conceptual orality in online language. I applied the same analysis to the YYPC, and, although the results cannot be presented in full here, the YYPC data supported the findings of the previous investigation that there is a continuum of conceptual orality, decreasing with comment length. However, as the average comment length of females with a migration background is *shorter* than both non-migration gender groups', at 2.4 CCUs per comment, it is reasonable to suggest that it is gender and ethnicity that are the factors behind the standard language orientation of this subset, and not greater comment length.

is adversarial in disposition and contains a contraction and CMC terminology. Males with a non-migration background produce the fewest emoji on average.

Female with a Migration Background

Serap Dogan *Super Video, hat mich für Videoideen inspiriert* ♡

('Super video, inspired me for video ideas ♡')

This is a characteristic comment from a woman with a migration background as it has a positive force and contains very few distinguishing non-standard features compared to the other subsets, apart from an emoji. Comments from this group contain one emoji on average. This comment is two CCUs long, around the average of 2.4 CCUs and contains four words per CCU, which sits just below the 4.2 average.

Male with a Migration Background

Sahan Furkan *Tugay abi mach mal bitte wieder GTA videos* 😁😁😁

'Tugay [name of YouTuber] older brother [polite Turkish term of address for another man of roughly similar age] make GTA videos again please 😁😁😁'

This is a typical comment from a man with a migration background. It is two CCUs long, close to the average, the lowest of the four subsets, of 2.1 CCUs. It contains language mixing, here illustrated by the Turkish term *Abi*. It contains the high use of a single emoji, and is neither positive nor adversarial in nature, which is most common for this group.

5 Conclusions

This study investigated possible online language differences based on gender and migration background using a corpus of YouTube comments, a new source of

language barely studied. The study offered an important opportunity to revisit findings about online language and add new knowledge about gender, ethnicity, and online language. The investigation was guided by three questions, which are answered below:

1. *Are the language traits associated with male and female gender performativity, both online and offline, reflected in online German youth language?*

The findings, presented in section 4.1, provide a useful update to previous research, both dealing with online and offline linguistic and discursive gender performance. I do not claim that the prediction by Balka (1993) that the internet will provide an equal space for males and females to participate has been realised, but females do not seem to feel they must restrict the volume of language they produce (as Trömel-Plötz (1979) feared).[23] Although males on YouTube still wrote slightly more comments than females, the complexity of the comments by both genders was very similar, indicating that young people do not seem to be pressured to restrict what they want to communicate. Rather, it is the "how" that differs. For example, there are elements of males' language that reveal a desire to assert a stereotypical image of masculinity – it is often more profane, argumentative and lacking emotional cues, which reaffirms earlier research by Herring (2003, 1996c). Although this last feature of male language has changed thanks to emoji, emoji are by far a female linguistic feature, in both proportion of users and use per comment. There are gender differences as to which emoji are used and the variety; females use a wider range of

[23] Given the assumptions to assign gender to the commenters, per section 3.1.4.

emoji, whereas males tend to use a much smaller range, mostly the 'Face With Tears of Joy' (😂). Females appear still to be more expressive when it comes to modulating the emotion expressed in a comment by repeating punctuation marks and letters (graphemes), as in the research by Herring & Zelenkauskaite (2009) and Waseleski (2006), except when it comes to representing laughter, where the use of laughter (*lol* or *haha*, and variants) is a male feature. Males tend to converse more within a same-gender environment than in mixed settings, which corresponds to earlier research by Savicki, Lingenfelter & Kelley (1996), where they antagonised each other with linguistic one-upmanship. Nevertheless, the spoken features that have traditionally characterised male language, such as contractions and other non-prestige features, as per Labov (1972) in spoken language and Baron (2004) in online communication, do not seem to be representative of online language for German young people, although the methodology does make some assumptions regarding the demographics of the commenters.

2. *Are the findings of previous research of linguistic differences between males and females with a non-migration background reflected in online German youth language?*

The language differences between males and females with a non-migration background follow roughly the same pattern as the overall gender differences, with a few slight exceptions. The proportion of males and females who use punctuation/emoji prosody and the proportion who use emoji, as well as the average use of both features in a comment are not statistically different. Furthermore, there is no difference in the use of CMC terminology amongst genders

with a non-migration background, but compared to those with a migration background, the use of CMC terminology is statistically higher. However, the overall usage of these features is still in line with the overall findings, although the corpus was smaller, given only 56.1% of comments could be identified for ethnicity. Other studies, such as Siebenhaar (2006), have focused on online use of regionalisms and dialect, which may present its own new set of gender differences in online language within an ethnically German framework that differ from spoken expectations, such as those discussed by Clyne (1995: 98-105), but that has not been investigated here.

3. *Are the findings of previous research of linguistic differences between males and females with a migration background reflected in online German youth language?*

Previous research on the language differences between males and females with a migration background is largely supported by the YYPC data. Language mixing is far more prevalent amongst males, which confirms the findings of Auer (2003) and Tekin & Colliander (2010). The more standard nature of female language in the YYPC correlates with the findings of Bierbach & Birken-Silverman (2007), Labov (1972: 243) and Trudgill (1972: 182-83), who affirm that this linguistic phenomenon is precipitated by the pressure to assert one's social status through the use of prestige forms of language. The YYPC suggests that females with a migration background write a similar number of comments to their male counterparts, although there are far fewer, whereas there are far more comments from females with a non-migration background than males with a non-migration background. Perhaps this is due to the lack of ethnicity and gender signifiers in the usernames of

females with a migration background, as shown by the difficulty in assigning ethnicity to commenters compared to gender. This lack of overt gender or ethnicity performance may also be linked to the more standard nature of language used by this group, although more research is needed on this topic. The language differences across gender of those with a migration background are more numerous than between the non-migration background populations, but mostly still conform to the general gender differences addressed in the second research question.

Underlying all of these differences remains the central issue of how social power structures control language production: gender performativity (Butler 2006) pressures those who identify as either male or female to conform to existing linguistic behavioural patterns; structural racism and 'othering' (Said 2003) does the same to those with a migration background. Online language, although it contains its own linguistic and discursive norms, appears to remain in thrall to the wider societal expectations that frame it. However, the data from the YouTube Young Person corpus is encouraging in one sense: previous linguistic characteristics, for both gender and ethnicity, such as the avoidance of emotional markers by males, appear to be becoming less prominent. Other previously held notions of the 'language of men and women', too, such as the previous assertion that males use more contractions, have not been supported in this study. However, linguistic differences that may express societal inequality, which also applies to questions of ethnicity (L2 idiosyncrasies notwithstanding), are still present in the YYPC, particularly in discursive features, such as the adversarial disposition of males.

The results of this study shed new light on how gender and ethnicity affect language production online, particularly in an intersectional context, perhaps most strikingly in the more standard-orientation of the language of females with migration backgrounds. Quantitative investigations of online language are, thus far, a little researched area, and this study can serve as a springboard for further studies to investigate both larger and more specific online corpora, to provide further insights into online language. The ever-growing amount of publicly available online German from social media and other interactive websites mean that corpora that focus on different aspects of German society can be collated and analysed similarly to the investigations in this essay. These can then be used to highlight specific areas of gender linguistic difference, or identify relevant areas that have not been covered in this essay. As such websites remain popular and the "digital natives" become the majority demographic in society, a diachronic investigation of online language change would be useful, in order to map whether the changes in spoken or written language are reflected in online language, or if features from online language are absorbed into spoken language, and vice versa. It would also be interesting to explore whether there is a difference between corpora collated from different social media sources, such as Facebook. Research on the online language of those who belong to other intersections of non-dominant population groups, such as those from less privileged socioeconomic backgrounds (cf. Hellinger & Bussmann 2003b) or LGBT+ (cf. Hanmer 2010; Pulos 2013), would help contribute to these fledgling research areas. The focus thus far is on critical discourse in these online spaces, rather than linguistic variation that is covered in spoken language (cf. Cameron & Kulick 2003). This essay investigates just a few of the possibilities to

examine how societal power structures can affect online language but demonstrates that, even within the last decade, the online linguistic landscape has dramatically changed and will continue to do so.

6 Appendix

The size of the YYPC corpus makes it too large to include in the main document. An online version containing the comments and conversations is available as a Google Sheets document to view and download at this link: https://goo.gl/TuzXfN

7 Bibliography

Online Sources

BurdaForward. 2016. 'Mobile Effects 2016: Wie wir Smartphone und Tablet im Alltag nutzen', ≤https://www.burda-forward.de/uploads/tx_mjstudien/BF_Digitalmarkt_MobileEffects2016.pdf> [accessed 20.08.2017]

References

Albert, Georg. 2013. *Innovative Schriftlichkeit in digitalen Texten: syntaktische Variation und stilistische Differenzierung in Chat und Forum* (Berlin: Akademie Verlag)

Androutsopoulos, Jannis. 2001. 'Ultra korregd Alder! Zur medialen Stilisierung und Popularisierung von 'Türkendeutsch'', *Deutsche Sprache*, 29: 321-39

———. 2004. '*... und jetzt gehe ich chillen*: Jugend- und Szenesprachen als lexikalische Erneuerungsquellen des Standards.' in *Standardvariation: Wie viel Variation verträgt die deutsche Sprache?*, ed. by Ludwig Eichinger and Werner Kallmeyer (Berlin: De Gruyter), pp. 171-206

———. 2006. 'Multilingualism, Diaspora, and the Internet: Codes and Identities on German-based Diaspora Websites', *Journal of Sociolinguistics*, 10: 520-47

———. 2008. 'Potentials and Limitations of Discourse-centred Online Ethnography', *Language@Internet*, 5: 9

———. 2010. 'Localizing the Global on the Participatory Web.' in *The Handbook of Language and Globalization*, ed. by Nikolas Coupland (Oxford: Wiley-Blackwell), pp. 201-31

———. 2013. 'Participatory Culture and Metalinguistic Discourse: Performing and Negotiating German Dialects on YouTube.' in *Discourse 2.0: Language and New Media*, ed. by Deborah Tannen and Anne Marie Trester (Washington D.C.: Georgetown University Press), pp. 47-72

———. 2015. 'Networked Multilingualism: Some Language Practices on Facebook and their Implications', *International Journal Of Bilingualism*, 19: 185-205

Androutsopoulos, Jannis, and Alexandra Georgakopoulou. 2003. *Discourse Constructions of Youth Identities* (Philadelphia: John Benjamins)

Auer, Peter. 2003. ''Türkenslang': Ein jugendsprachlicher Ethnolekt des Deutschen und seine Transformationen.' in *Spracherwerb und Lebensalter*, ed. by Annelies Häcki Buhofer (Tübingen: Franke), pp. 255-64

Auer, Peter, and Inci Dirim. 2003. 'Socio-cultural Orientation, Urban Youth Styles and the Spontaneous Acquisition of Turkish by non-Turkish Adolescents in Germany.' in *Discourse Constructions of Youth Identities*, ed. by Jannis Androutsopoulos and Alexandra Georgakopoulou (Philadelphia: John Benjamins), pp. 223-46

Balka, Ellen. 1993. 'Women's Access to On-line Discussions about Feminism', *The Electronic Journal of Communication*, 3 ≤http://www.cios.org/EJCPUBLIC/003/1/00311.HTML> [accessed 12 June 2017]

Barbour, Stephen, and Patrick Stevenson. 1998. *Variation im Deutschen: Soziolinguistische Perspektiven* (Berlin: De Gruyter)

Baron, Naomi S. 2004. 'See you Online: Gender Issues in College Student Use of Instant Messaging', *Journal of Language and Social Psychology*, 23: 397-423

Bierbach, Christine, and Gabriele Birken-Silverman. 2007. 'Names and Identities, or: How to be a Hip Young Italian Migrant in Germany.' in *Style and Social Identities: Alternative Approaches to Linguistic Heterogeneity*, ed. by Peter Auer (Berlin: De Gruyter), pp. 121-54

Braun, Friederike, Anja Gottburgsen, Dagmar Stahlberg, and Sabine Sczesny. 1998. 'Können Geophysiker Frauen Sein? Generische Personenbezeichnungen Im Deutschen', *Zeitschrift Für Germanistische Linguistik*, 26: 265–83

Brown, Penelope. 1991. 'Sind Frauen Höflicher? Befunde Aus Einer Maya-Gemeinde', in *Von Fremden Stimmen: Weibliches Und Männliches Sprechen Im Kulturvergleich* (Suhrkamp), pp. 101–29

Bundesamt für Migration und Flüchtlinge. 2009. 'Grunddaten der Zuwandererbevölkerung in Deutschland: Working Paper 27', ≤http://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/WorkingPapers/wp27-grunddaten.html> [accessed 20.08.2017]

Butler, Judith. 2006. *Gender Trouble: Feminism and the Subversion of Identity*, 2nd Edition (New York: Routledge)

Caglar, Ayse S. 1998. 'Popular Culture, Marginality and Institutional Incorporation: German-Turkish Rap and Turkish Pop in Berlin', *Cultural Dynamics*, 10: 243-61

Cameron, Deborah. 2003. 'Gender and Language Ideologies.' in *The Handbook of Language, Gender*, ed. by Susan Ehrlich and Miriam Meyerhoff (Oxford: John Wiley & Sons), pp. 447-67

Cameron, Deborah. 2005. 'Why Is Language a Feminist Issue?', in *The Feminist Critique of Language: A Reader*, ed. by Deborah Cameron, 2nd Edition (Abingdon: Routledge), pp. 1–21

Cameron, Deborah, and Don Kulick. 2003. *Language and Sexuality* (Cambridge: Cambridge University Press)

Clammer, John. 2015. 'Performing ethnicity: performance, gender, body and belief in the construction and signalling of identity', *Ethnic and Racial Studies*, 38: 2159-66

Clyne, Michael G. 1995. *The German language in a changing Europe* Cambridge University Press)

Coates, Jennifer. 2013. *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language* (Abingdon: Routledge)

Cotgrove, Louis. 2017. 'Is computer-mediated communication 'written colloquial speech' (Kilian 2001)? A quantitative study of German-language You Tube comments.'

Crenshaw, Kimberle. 1989. 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics', *University of Chicago Legal Forum*, 1989: 139-68

Crystal, David. 2006. *Language and the internet*, 2nd (Cambridge: Cambridge University Press)

Cunha, Evandro, Gabriel Magno, Virgilio Almeida, Marcos André Gonçalves, and Fabricio Benevenuto. 2012. 'A Gender Based Study of Tagging Behavior in Twitter', *Proceedings of the 23rd ACM conference on Hypertext and social media*: 323-24

Cutler, Cecilia. 1999. 'Yorkville Crossing: White teens, hip hop and African American English', *Journal of Sociolinguistics*, 3: 428-42

Danet, Brenda. 1998. 'Text as Mask: Gender, Play, and Performance on the Internet.' in *Cybersociety 2.0: Revisiting Computer-Mediated Communication and Community*, ed. by Steven G. Jones (Los Angeles: Sage Publications), pp. 129-58

Dörnyei, Zoltran. 2007. *Research Methods in Applied Linguistics* (Oxford: Oxford University Press)

Durrell, Martin. 2003. *Using German: a guide to contemporary usage*, 2nd ed. (Cambridge: Cambridge University Press)

Dürscheid, Christa. 2006. 'Medienkommunikation und Jugendsprache.' in *Perspektiven der Jugendsprachforschung*, ed. by Christa Dürscheid and Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 117-31

Dürscheid, Christa, and Sarah Brommer. 2009. 'Getippte Dialoge in neuen Medien. Sprachkritische Aspekte und linguistische Analysen', *Linguistik Online*, 37: 9 ≤https://bop.unibe.ch/linguistik-online/article/view/511> [accessed 23 May 2016]

Eckert, Penelope. 2003. 'Language and Gender in Adolescence.' in *The handbook of language and gender*, ed. by Janet Holmes and Miriam Meyerhoff (Oxford: John Wiley & Sons), pp. 529-45

Eckert, Penelope, and Sally McConnell-Ginet. 1992. 'Think Practically and Look Locally: Language and Gender as Community-Based Practice', *Annual Review of Anthropology*, 21.1: 461–88

Fagan, Sarah M. B. 2009. *German: a linguistic introduction* (Cambridge: Cambridge University Press)

García, Ofelia. 2010. 'Languaging and ethnifying' in *The handbook of language and ethnic identity. Disciplinary and regional perspectives*, ed. by Joshua Fishman and Ofelia García (Oxford: Oxford University Press), pp. 519-534

García, Ofelia, and Li Wei. 2014. *Translanguaging: Language, Bilingualism and Education* (London: Palgrave Macmillan)

Gräßel, Ulrike. 1991. *Sprachverhalten und Geschlecht: eine empirische Studie zu geschlechtsspezifischem Sprachverhalten in Fernsehdiskussionen* (Regensburg: Centaurus-Verlagsgesellschaft)

Hanmer, Rosalind. 2010. 'Internet fandom, queer discourse, and identities', *LGBT identity and online new media (147-158). New York: Routledge*

Heller, Monica, and Erica McClure. 1988. *Codeswitching: Anthropological and Sociolinguistic Perspectives* (Berlin/Boston, DE: De Gruyter Mouton)

Hellinger, Marlis, and Hadumod Bussmann. 2003a. 'Engendering Female Visibility in German.' in *Gender Across Languages: The Linguistic Representation of Women and Men*, ed. by Marlis Hellinger and Hadumod Bussmann (Amsterdam: John Benjamins), pp. 141-74

——— (ed.)^(eds.). 2003b. *Gender across languages: the linguistic representation of women and men: Volume 1* (Amsterdam: John Benjamins)

Herring, Susan C. 1996a. *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives* (Philadelphia: John Benjamins)

———. 1996b. 'Gender and Democracy in Computer-Mediated Communication.' in *Computerization and controversy: Value conflicts and social choices*, ed. by Rob Kling (San Diego: Morgan Kaufmann), pp. 476-99

———. 1996c. 'Gender Differences in Computer-Mediated Communication: Bringing Familiar Baggage to the New Frontier.' in *CyberReader*, ed. by Victor Vitanza (Boston, MA: Allyn & Bacon), pp. 144-154

———. 2003. 'Gender and Power in On-line Communication.' in *The handbook of language and gender*, ed. by Janet Holmes and Miriam Meyerhoff (Oxford: John Wiley & Sons), pp. 202-28

Herring, Susan C., and Asta Zelenkauskaite. 2009. 'Symbolic Capital in a Virtual Heterosexual Market', *Written Communication*, 26: 5-31

Holmes, Janet. 1992. *An Introduction to Sociolinguistics* (London: Longman)

Holmes, Janet. 1984. '"Women's Language": A Functional Approach', *General Linguistics*, 24.3: 149–78

Huffaker, David A., and Sandra L. Calvert. 2005. 'Gender, Identity, and Language Use in Teenage Blogs', *Journal of Computer-Mediated Communication*, 10: 00-00

Hummel, Cornelia. 1984. '"Sie haben jetzt ja lange geredet, Frau Lieberherr": Entschuldigungen, Vorwürfe, Bitten und direkte Anreden in Fernsehdiskussionen.' in *Gewalt durch Sprache: die Vergewaltigung von Frauen in Gesprächen* ed. by Senta Trömel-Plötz (Frankfurt am Main: Fischer), pp. 258-87

Hutchby, Ian, and Robin Wooffitt. 1998. *Conversation analysis : principles, practices, and applications* (Cambridge ; Malden, Mass.: Polity Press)

Keim, Inken. 2007a. *Die "türkischen Powergirls": Lebenswelt und kommunikativer Stil einer Migrantinnengruppe in Mannheim* (Tübingen: Narr)

———. 2007b. 'Socio-cultural Identity, Communicative Style, and their Change over Time: A Case Study of a Group of German-Turkish Girls in Mannheim, Germany.' in *Style and Social Identities: Alternative Approaches to Linguistic Heterogeneity*, ed. by Peter Auer (Berlin: De Gruyter), pp. 155-86

Kendall, Lori. 1998. 'Meaning and Identity in "cyberspace": The Performance of Gender, Class, and Race Online', *Symbolic interaction*, 21: 129-53

Kilian, Jörg. 2001. 'T@stentöne. Geschriebene Umgangssprache in computervermittelter Kommunikation. Historisch-kritische Ergänzungen zu einem neuen Feld der linguistischen Forschung.' in *Chatkommunikation. Sprache, Interaktion, Sozialität & Identität in Synchroner Computervermittelter Kommunikation*, ed. by Michael Beißwenger (Stuttgart: Ibidem), pp. 55-78

———. 2010. 'T@stentöne. Geschriebene Umgangssprache in comuptervermittelter kommunikation. Historisch-kritische Ergänzungen zu einem neuen Feld der linguistischen Forschung.' in *Deutsche Gegenwartssprache: Globalisierung, neue Medien, Sprachkritik*, ed. by Iris Forster and Tobias Heinz (Stuttgart: Reclam), pp. 63-96

Klann-Delius, Gisela. 2005. *Sprache Und Geschlecht: Eine Einführung* (Berlin: Springer-Verlag)

Kleinberger Günther, Ulla, and Carmen Spiegel. 2006. 'Jugendliche schreiben im Internet: grammatische und orthographische Phänomene in normungebundenen Kontexten.' in *Perspektiven der Jugendsprachforschung*, ed. by Christa Dürscheid and Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 101-15

Koch, Peter, and Wulf Österreicher. 1985. 'Sprache der Nähe–Sprache der Distanz', *Romanistisches Jahrbuch*, 36: 15-43

Labov, William. 1966. *The Social Stratification of English in New York City* (Cambridge: Cambridge University Press)

———. 1972. *Sociolinguistic Patterns* (Philadelphia: University of Pennsylvania Press)

Lakoff, Robin Tolmach. 1975. *Language and Woman's Place* (New York: Harper & Row)

Le Page, R. B., and Andrée Tabouret-Keller. 1985. *Acts of Identity: Creole-based Approaches to Language and Ethnicity* (Cambridge: Cambridge University Press)

Levon, Erez. 2010. 'Organizing and Processing Your Data: The Nuts and Bolts of Quantitative Analyses.' in *Research Methods in Linguistics*, ed. by Lia Litosseliti (London: Continuum), pp. 68-92

Lewis, Gwyn, Bryn Jones, and Colin Baker. 2012. 'Translanguaging: origins and development from school to street and beyond', *Educational Research and Evaluation*, 18: 641-54

Lytra, Vally, Peter Martin, Taşkin Baraç, and Arvind Bhatt. 2010. 'Investigating the Intersection of Multilingualism and Multimodality in Turkish and Gujarati Literacy Classes.' in *Sites of multilingualism: Complementary schools in Britain today*, ed. by Vally Lytra and Peter Martin (Stoke-on-Trent: Trentham Books), pp. 19-31

Meyerhoff, Miriam. 2002. 'Communities of Practice.' in *The Handbook of Language Variation and Change*, ed. by J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes (Malden, MA: Blackwell), pp. 526-48

Milroy, Lesley, and Sue Margrain. 1980. 'Vernacular Language Loyalty and Social Network', *Language in Society*, 9: 43-70

Ott, R Lyman, and Micheal T Longnecker. 2015. *An introduction to statistical methods and data analysis* Nelson Education)

Prokopiou, Evangelia, and Tony Cline. 2010. 'Constructing Cultural and Academic Identities in Community Schools: A Socio-cultural and Dialogical Approach.' in *Sites of multilingualism: Complementary schools in Britain today*, ed. by Vally Lytra and Peter Martin (Stoke-on-Trent: Trentham Books), pp. 71-83

Pulos, Alexis. 2013. 'Confronting heteronormativity in online games: A critical discourse analysis of LGBTQ sexuality in World of Warcraft', *Games and Culture*, 8: 77-97

Pusch, Luise F. 1984. *Das Deutsche als Männersprache : Aufsätze und Glossen zur feministischen Linguistik*, 1. Aufl. (Frankfurt am Main: Suhrkamp)

Rampton, Ben. 2014. *Crossing: Language and Ethnicity Among Adolescents* Florence : Taylor and Francis)

Rollefson, J. Griffith. 2013. ' 'Gheddos du Monde': Sounding the Ghetto, Occupying the Nation from Berlin to Paris.' in *Hip-Hop in Europe: Cultural Identities and Transnational Flows*, ed. by Sina Nitzsche and Walter Grünzweig (Berlin: LIT Verlag), pp. 227-40

Said, Edward W. 2003. *Orientalism*, 4th Edition (London: Penguin)

Savicki, Victor, Dawn Lingenfelter, and Merle Kelley. 1996. 'Gender Language Style and Group Composition in Internet Discussion Groups', *Journal of Computer-Mediated Communication*, 2: 0-0

Schegloff, Emanuel A. 2007. *Sequence organization in interaction : a primer in conversation analysis* (Cambridge ; New York: Cambridge University Press)

Schlobinski, Peter. 1995. 'Jugendsprachen: Speech Styles of Youth subcultures.' in *The German Language and the Real World. Sociolinguistic, Cultural, and Pragmatic Perspectives on Contemporary German,* , ed. by Patrick Stevenson (Oxford: Oxford University Press), pp. 317-39

———. 2004. 'Mündlichkeit/Schriftlichkeit in den Neuen Medien.' in *Standardvariation: Wie viel Variation verträgt die deutsche Sprache*, ed. by Ludwig Eichinger and Werner Kallmeyer (Berlin: Walter de Gruyter), pp. 126-42

Schoenthal, Gisela. 1989. ‘Personenbezeichnungen im Deutschen als Gegenstand feministischer Sprachkritik’, *Zeitschrift für Germanistische Linguistik*, 17.3: 296–314

Siebenhaar, Beat. 2006. 'Regionale Variation in Deutschen, österreichischen und Schweizer Chaträumen.' in *Von *hdl* bis *cul8r*: Sprache und Kommunikation in den neuen Medien*, ed. by Peter Schlobinski (Mannheim: Dudenverlag), pp. 133-47

Siever, Torsten. 2006. 'Sprachökonomie in den "Neuen Medien".' in *Von *hdl* bis *cul8r*: Sprache und Kommunikation in den neuen Medien*, ed. by Peter Schlobinski (Mannheim: Dudenverlag), pp. 71-88

Söll, Ludwig, and Franz Josef Hausmann. 1980. *Gesprochenes und geschriebenes Französisch* E. Schmidt)

Spitzmüller, Jürgen. 2006. 'Der mediale Diskurs zu" Jugendsprache": Kontinuität und Wandel.' in *Perspektiven der Jugend-sprachforschung* ed. by Christa Dürscheid and Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 33-50

Statistik Austria. 2012. "Migration & Integration." in, edited by Kommission für Migrations- und Integrationsforschung and der Österreichischen Akademie der Wissenschaften. Wien.

Statistisches Bundesamt. 2015. "Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2015." in, edited by Statistisches Bundesamt. Wiesbaden.

———. 2016. 'Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2016', ≤https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/Migrationshintergrund.html> [accessed 20.08.2017]

Steil, Janice M., and Jennifer L. Hillman. 1993. 'The Perceived Value of Direct and Indirect Influence Strategies', *Psychology of Women Quarterly*, 17.4: 457–462

Storrer, Angelika. 2001. 'Getippte Gespräche oder dialogische Texte? Zur kommunikationstheoretischen Einordnung der Chat-Kommunikation.' in *Sprache im Alltag*, ed. by Lehr Andrea, Kammerer Matthias, Konerding Klaus-Peter, Storrer Angelika, Thimm Caja and Wolski Werner (Berlin: De Gruyter), pp. 439-65

Tagliamonte, Sali. 2016a. *Teen Talk: The Language of Adolescents* (Cambridge: Cambridge University Press)

Tagliamonte, Sali A. 2016b. 'So sick or so cool? The language of youth on the internet', *Language in Society*, 45: 1-32

Tekin, Özlem, and Peter Colliander. 2010. 'Das »Türkendeutsch«', *Zeitschrift für interkulturelle Germanistik*, 1: 49-61

Thimm, Caja. 1990. *Dominanz und Sprache: Strategisches Handeln im Alltag* (Wiesbaden: Springer)

Trömel-Plötz, Senta. 1979. *Frauensprache in unserer Welt der Männer* (Konstanz: Universitätsverlag)

———. 1982. *Frauensprache: Sprache der Veränderung* (Frankfurt am Main: Fischer)

———. 1984a. 'Gewalt durch Sprache.' in *Gewalt durch Sprache: die Vergewaltigung von Frauen in Gesprächen*, ed. by Senta Trömel-Plötz (Frankfurt am Main: Fischer), pp. 50-67

———. 1984b. *Gewalt durch Sprache: die Vergewaltigung von Frauen in Gesprächen*, Originalausg. (Frankfurt am Main: Fischer)

Trudgill, Peter. 1972. 'Sex, Covert Prestige and Linguistic Change in the Urban British English of Norwich', *Language in Society*, 1: 179-95

Waseleski, Carol. 2006. 'Gender and the Use of Exclamation Points in Computer-Mediated Communication: An Analysis of Exclamations Posted to Two Electronic Discussion Lists', *Journal of Computer-Mediated Communication*, 11: 1012-24

Watzlawick, Sonja. 2006. 'Jugendbiografien und Sprachgebrauch im Wandel der Zeit.' in *Perspektiven der Jugend-sprachforschung* ed. by Christa Dürscheid and Jürgen Spitzmüller (Frankfurt am Main: Peter Lang), pp. 281-94

West, Candace, and Don H. Zimmerman. 1977. 'Women's Place in Everyday Talk: Reflections on Parent-Child Interaction', *Social problems*, 24: 521-29

Wiese, Heike. 2009. 'Grammatical Innovation in Multiethnic Urban Europe: New Linguistic Practices among Adolescents', *Lingua*, 119: 782 - 806

Wirth, Uwe. 2005. 'Chatten. Plaudern mit anderen Mitteln.' in *Websprache.net: Sprache und Kommunikation im Internet*, ed. by Torsten Siever, Peter Schlobinski and Jens Runkehl (Berlin: De Gruyter), pp. 67-84

Witmer, Diane F., and Sandra Lee Katzman. 1997. 'On-Line Smiles: Does Gender Make a Difference in the Use of Graphic Accents?', *Journal of Computer-Mediated Communication*, 2

Wodak, Ruth. 1981. 'Women Relate, Men Report: Sex Differences in Language Behaviour in a Therapeutic Group', *Journal of Pragmatics*, 5: 261-85

Wolf, Alecia. 2000. 'Emotional Expression Online: Gender Differences in Emoticon Use', *CyberPsychology & Behavior*, 3: 827-33